

16. Salenbien R., Wack Y., Baelmans M., Blommaert M. Geographically informed automated non-linear topology optimization of district heating networks, *Energy*, 2023, Vol. 283, 128898.
17. Rodgers Waymond, Cardenas Jesus A., Gemoets Leopoldo A., Sarfi Robert J. A smart grids knowledge transfer paradigm supported by experts' throughput modeling artificial intelligence algorithmic processes, *Technological Forecasting and Social Change*, 2023, Vol. 190, 122373.
18. Jean Paul A. Yaacoub, Javier Hernandez Fernandez, Hassan N. Noura, Ali Chehab. Security of Power Line Communication systems: Issues, limitations and existing solutions, *Computer Science Review*, 2021, Vol. 39, 100331.
19. Dennis Amelia, Weston Dale, Amlôt i chard, Arnold Andreas, Carbon Danielle, Carter Holly. The role of pre-incident information and responder communication in effective management of casualties, including members of vulnerable groups, during a decontamination field exercise, *International Journal of Disaster Risk Reduction*, 2023, Vol. 94, 103806.
20. Belyakov S., Belyakova M., Bozhenyuk A., Rozenberg I. The features of generations of solutions by intellectual information systems, In: *Advances in Intelligent Systems and Computing*. Springer International Publishing Switzerland, 2016, Vol. 451, pp. 221-229.

Статью рекомендовал к опубликованию д.т.н. А.В. Боженюк.

Беляков Станислав Леонидович – Южный федеральный университет; e-mail: sbelyakov@sfedu.ru, beliacov@yandex.ru; г. Таганрог, Россия; тел.: +79185109607; д.т.н.; профессор кафедры информационно-аналитических систем безопасности имени профессора Берштейна Леонида Самойловича.

Исаев Александр Викторович – e-mail: alis@sfedu.ru, shura.isaev.99@mail.ru; тел.: +79518352784; аспирант кафедры информационно-аналитических систем безопасности имени профессора Берштейна Леонида Самойловича.

Belyakov Stanislav Leonidovich – Southern Federal University; e-mail: sbelyakov@sfedu.ru, beliacov@yandex.ru; Taganrog, Russia; phone: +79185109607; dr of eng. sc.; professor of the department of information and analytical security systems named after professor Leonid Samoilovich Bershtein.

Isaev Alexander Viktorovich – Southern Federal University; e-mail: alis@sfedu.ru, shura.isaev.99@mail.ru; phone: +79518352784; postgraduate student of the department of information and analytical security systems named after professor Leonid Samoilovich Bershtein.

УДК 004.89

DOI 10.18522/2311-3103-2024-2-38-50

Д.Ю. Кравченко

МОДЕЛЬ ОНТОЛОГИИ ЗНАНИЙ ДЛЯ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ ОБРАБОТКИ И АНАЛИЗА ТЕКСТОВ*

Статья посвящена решению научной проблемы создания верхнеуровневого описания модели онтологии знаний для интеллектуальных систем обработки и анализа текстов на естественном языке, построенной на основе оригинальной компонентной архитектуры, обеспечивающей необходимый уровень детализации спецификаций анализируемой текстовой информации. Актуальность данной задачи обусловлена необходимостью развития теоретических основ построения информационных моделей семантических зависимостей внутри текстов на естественном языке. Автором даны определения основным терминам исследуемой предметной области. Представлена формализованная постановка решаемой задачи. Проблема «информационного взрыва», причиной возникновения которой стал экспоненциальный рост объемов цифровой информации, привела к ситуации, когда до 95% информационного потока содержит неструктурированные данные. В подобных условиях, крайне актуальной становится задача создания эффективных интеллектуальных систем поиска и приобретения знаний, в том числе, интеллектуальных систем обработки и анализа текстов на естественном языке. Научным направлением решения этой частной задачи является Text Mining (ТМ) – раскопка знаний в текстовой информации. В качестве примера прикладной задачи использования приобретенных знаний, в данном исследовании, рассматривается

* Исследование выполнено за счет гранта Российского научного фонда № 22-71-10121, <https://rscf.ru/project/22-71-10121/> в Южном федеральном университете.

значимая проблема информационной поддержки процессов предупреждения и/или ликвидации последствий чрезвычайных ситуаций. В данной задаче исходными данными являются потоки текстовых сообщений (новостной информации, отчетов о техническом состоянии техногенных объектов, информации о природных явлениях и т.п.), поступающих в центры принятия решений, а на выходе формируются прогностические оценки и/или конкретные инструкции относительно оценки ситуации и предпринимаемых действий определенными специалистами. Одной из причин, сдерживающих развитие интеллектуальных систем обработки и анализа текста для решения задач поиска, приобретения и использования знаний, является недостаточно высокий уровень эффективности моделей и алгоритмов, обеспечивающих комплексное решение описанных выше задач искусственного интеллекта с учетом особенностей семантики и контекста.

Онтология знаний; обработка и анализ текстов; семантика; информационная поддержка; чрезвычайные ситуации; поддержка принятия решений; структурирование информации.

D.Yu. Kravchenko

KNOWLEDGE ONTOLOGY MODEL FOR INTELLIGENT TEXT PROCESSING AND ANALYSIS SYSTEMS

The article is devoted to solving the scientific problem of creating a top-level description of a knowledge ontology model for intelligent systems for processing and analyzing texts in natural language, built on the basis of an original component architecture that provides the necessary level of detail in the specifications of the analyzed text information. The relevance of this task is due to the need to develop the theoretical foundations for constructing information models of semantic dependencies within texts in natural language. The author gives definitions to the main terms of the subject area under study. A formalized definition of the problem being solved is presented. The problem of the "information explosion," which was caused by the exponential growth in the volume of digital information, has led to a situation where up to 95% of the information flow contains unstructured data. In such conditions, the task of creating effective intelligent systems for searching and acquiring knowledge, including intelligent systems for processing and analyzing texts in natural language, becomes extremely urgent. The scientific direction for solving this particular problem is Text Mining (TM) – the excavation of knowledge in text information. As an example of the applied task of using acquired knowledge, this study examines the significant problem of information support for the processes of preventing and/or eliminating the consequences of emergency situations. In this task, the initial data are streams of text messages (news information, reports on the technical condition of man-made objects, information about natural phenomena, etc.) arriving at decision-making centers, and the output is formed by predictive assessments and/or specific instructions regarding the assessment situations and actions taken by certain specialists. One of the reasons hindering the development of intelligent text processing and analysis systems for solving problems of searching, acquiring and using knowledge is the insufficiently high level of models and algorithms efficiency that provide a comprehensive solution to the above-described problems of artificial intelligence, taking into account the peculiarities of semantics and context.

Knowledge ontology; text processing and analysis; semantics; information support; emergency situations; decision support; information structuring.

Введение. Повышенный интерес исследователей к проблеме поиска и приобретения знаний при обработке и анализе текстовой информации привёл к появлению значительного числа определений основных терминов в данной предметной области. Следствием такой терминологической рассогласованности стала проблема неопределенности и нечеткости при описании базовых понятий: информация, данные и знания [1]. В контексте Text Mining (TM) под априорной информацией будем понимать входные текстовые документы, включающие в себя наборы неструктурированных или слабо структурированных данных. Текстовый документ является основной лингвистической единицей естественного языка. Построение модели представления текста позволяет закодировать семантические характеристики информации в виде вектора, который в дальнейшем применяется для решения различных прикладных задач [2, 3].

Задача TM – поиск и приобретение, а затем использование знаний. Знания (англ. *knowledge*) – это сложная иерархия элементов ценной информации с выявленными зависимостями и закономерностями между фактами, событиями, явлениями и процессами [4]. Ценность информации определяется на основе расчета вероятностных оценок достижения

цели решаемой прикладной задачи до и после получения определенной информации [5]. *Поиск знаний* (англ. *knowledge retrieval*) – информационный процесс возврата информации к структурированной форме. Под *приобретением знаний* (англ. *knowledge acquisition*) для текстовой информации, после приведения ее к некоторой структуре на этапе поиска, будем понимать систематизацию полученных структурированных знаний [2–5].

Результат последовательных процессов поиска и приобретения знаний определяется достоверностью, надежностью, релевантностью, интерпретируемостью нового знания. Достижимость данных характеристик знаний напрямую связана с проблемой снижения уровня *информационной неопределенности* – нехватки информации для решения поставленных прикладных задач [6].

В качестве примера прикладной задачи использования приобретенных знаний, в данном исследовании, рассматривается значимая проблема информационной поддержки процессов предупреждения и/или ликвидации последствий чрезвычайных ситуаций. В данной задаче исходными данными являются потоки текстовых сообщений (новостной информации, отчетов о техническом состоянии техногенных объектов, информации о природных явлениях и т.п.), поступающих в центры принятия решений, а на выходе формируются прогностические оценки и/или конкретные инструкции относительно оценки ситуации и предпринимаемых действий определенными специалистами [7].

Технологии анализа и обработки текстов на естественном языке (NLP) предъявляют ряд требований к разработке интеллектуальных систем, обеспечивая тем самым конфигурируемость данных инструментов. Особенности каждой конфигурации интеллектуальной системы анализа и обработки текста для решения задач поиска, приобретения и использования знаний зависят от имеющихся пересечений определенных разделов лингвистики [8], а именно: технологии построения и анализа слов (морфология); технологии построения и анализа предложений (синтаксис); технологии содержательного анализа и смысловой оценки (семантика).

1. Задачи поиска, приобретения и использования знаний при анализе и обработке текстов. Представим указанные в названии пункта задачи поиска, приобретения и использования знаний в виде процесса извлечения знаний (*knowledge extraction*). Данный процесс в самой общей форме представляет собой последовательность следующих этапов, показанных на рис. 1: поступление корпуса текста на вход интеллектуальной системы; решение задач предобработки текста; применение оператора преобразования полученной после предобработки входных данных текстовой информации.



Рис. 1. Процесс извлечения знаний в общем виде

Формально данное отображение запишем в следующем виде:

$$F_{\text{extraction}}: [Doc_x, x = 1 \dots T] \rightarrow \text{knowledge}, \quad (1)$$

где $F_{\text{extraction}}$ – оператор преобразования полученной после предобработки входных данных текстовой информации; Doc_x – множество текстовых документов; knowledge – знания, необходимые для решения последующих, иерархически вышестоящих задач, например, для решения задачи информационной поддержки процессов предупреждения и/или ликвидации последствий чрезвычайных ситуаций.

Одним из *способов представления* полученных знаний является построение модели онтологии, описывающей предметную область в виде множества концептов (понятий) (K_i , $i = 1 \dots N$, где N – количество концептов) с заданной на нём системой связей (R_q , $q = 1 \dots M$, где M – количество связей), являющихся по своей сути несимметричными семантическими отношениями. Таким образом, элементом подобной модели является кортеж длины три: $\langle k_i, r_q, k_j \rangle$, $i \neq j$. Система понятий, подходящая для реализации машинного вывода в контексте решения прикладных задач использования приобретенных знаний, подразумевает применение таких категорий семантических отношений, как, например, «эквивалентность», «класс – подкласс», «часть – целое» и другие [9, 10].

Отметим, что тексты на естественном языке в явном или неявном виде содержат в себе компоненты смысла, имеющие в своей основе структуру описанных выше кортежей, имитирующих знания человека и позволяющих строить информационные модели с практически неограниченными возможностями масштабирования. Именно процессы построения и обработки подобных информационных моделей (онтологий, графов знаний, моделей мира) лежат в основе методов поиска, приобретения и использования знаний при анализе и обработке текста.

Описанный выше способ представления данных делит процессы поиска и приобретения знаний на этапы *извлечения концептов* (concept extraction) и *извлечения отношений* между ними (relation extraction). В интеллектуальных информационных системах происходит комбинирование данных этапов. Способ комбинирования определяет тип конкретной интеллектуальной информационной системы (ИИС). Так, например, *открытые* ИИС извлекают все возможные отношения между концептами в корпусе текста, а *закрытые* – только отношения из ранее заданной выборки концептов.

Задача *поиска знаний* в интеллектуальных системах обработки и анализа текстов на естественном языке в большей степени связана с качественным построением синтаксической схемы текста на основе применения парсера. С этой точки зрения наиболее важными будем считать процедуры структурирования и фильтрации текстовой информации, которые позволят определить основные смысловые элементы текста, такие как, например, *ключевые слова* (key words).

2. Постановка задачи. Извлечение ключевых слов является фундаментальной задачей в *обработке естественного языка* (Natural Language Processing, NLP) и включает в себя выявление и извлечение наиболее релевантных и значимых слов или фраз из заданного текста. Парсеры играют ключевую роль в этом процессе, анализируя синтаксическую структуру текста и помогая выявлять ключевые компоненты, представляющие важные понятия или темы.

1. *Синтаксический анализ структуры для извлечения ключевых слов.* Для выполнения извлечения ключевых слов с использованием синтаксического анализа структуры, можно сосредоточиться на конкретных синтаксических единицах, таких как именные фразы (NPs) [11] и глагольные фразы (VPs). Эти фразы часто являются хорошими кандидатами на роль ключевых слов, так как обычно содержат важную информацию о субъекте, объекте или действии в предложении. Например, рассмотрим предложение: "The swift fox jumps over the lazy dog." С помощью синтаксического анализа структуры будут выделены следующие фразы:

- ◆ именные фразы (NPs): "The swift fox", "the lazy dog";
- ◆ глагольная фраза (VP): "jumps over".

Из извлеченных фраз можно выделить следующие ключевые слова: "swift fox" (быстрая лиса); "lazy dog" (ленивая собака); "jumps over" (перепрыгивает через), как существенные компоненты данного предложения.

2. *Синтаксический анализ зависимостей для извлечения ключевых слов.* Синтаксический анализ зависимостей помогает определить отношения между субъектом, глаголом и объектом, а также другие существенные синтаксические зависимости, которые вносят вклад в общий смысл предложения. Ключевые слова могут быть извлечены из этих зависимостей, причем предпочтение отдается словам, несущим значительный семантический вес и играющим важные роли в структуре предложения [12].

В контексте извлечения ключевых слов текст преобразуется в граф, где вершины представляют собой возможные ключевые слова, а ребра – их отношения. Взаимосвязь между ключевыми фразами-кандидатами может быть определена по тому, как часто они встречаются вместе или насколько семантически близки.

Предположим, что строится ориентированный граф $G = (V, E)$, где V – множество вершин, а E – множество ребер. Оценка или важность вершины определяется как [11, 12]:

$$S(V_i) = (1 - d) + d * \sum_{j \in In(V_i)} \frac{1}{|Out(V_j)|} S(V_j), \quad (2)$$

где $In(V_i)$ – это набор вершин, которые указывают на V_i , а $Out(V_i)$ – это набор вершин, на которые указывает V_i . При этом d – это коэффициент затухания, который устанавливается в диапазоне от 0 до 1.

В рамках решения задачи поиска знаний в интеллектуальных системах обработки и анализа текстов на естественном языке данное исследование предполагает создание дополнительного фильтра на выходе парсера, что позволит извлечь смысловую часть предложения из полученной синтаксической схемы текстовой информации для дальнейшего использования в процессах приобретения знаний.

При решении задач приобретения знаний необходимо сформулировать постановку и условия задачи, равно как и само понятие решения. Пусть существует определенный подход S , позволяющий на множестве K анализируемых объектов знаний из множества P разных предметных областей атрибутам (характеристикам) X ставить в соответствие значения (величины) Y , которые приводят к оптимальной обработке разнородных распределенных информационных ресурсов R на основе применения методов M . При этом обработка информационных ресурсов должна проводиться с учетом контекста H , что является необходимым условием повышения эффективности информационного процесса приобретения знаний (функции) F . Таким образом, Y является достоверным выводом, а S – решающим подходом, построенным на основе комплекса методов и правдоподобных рассуждений. Очевидно, что в данном случае эффективность вывода напрямую связана с интеллектуальностью метода анализа и обработки имеющихся информационных ресурсов. Полученный вывод представляет собой систему проанализированных отношений и соответствий, приводящих к обнаружению неявных зависимостей и закономерностей между атрибутами X на множестве объектов знаний K . Идентифицированные новые зависимости и закономерности между объектами знаний являются основой для приобретения нового знания [4].

Реализация моделей и методов приобретения знаний на основе использования интеллектуальных систем существенно изменило аппарат формальных рассуждений. Основной особенностью задач построения рассуждений в контексте данного исследования является наличие информационной неопределенности и большой размерности, что требует получения и анализа значительного числа альтернативных вариантов структуры системно значимых отношений между объектами знаний исследуемых предметных областей.

Задача *приобретения знаний* сводится к оценке семантической близости понятий (концептов), позволяющей распределить их по классам для дальнейшего построения онтологии предметной области.

3. Построение модели онтологии знаний для интеллектуальных систем обработки и анализа текстов. Построение верхнеуровневой модели онтологии знаний для интеллектуальных систем обработки и анализа текстов проведено с учетом результатов анализа проблем поиска, приобретения и использования знаний, представленных в данной работе ранее.

Несмотря на тот факт, что подход к анализу и обработке текстов на основе извлечения ключевых слов не является новым, в определенных условиях он может дать значительный эффект при определении смысла предложения. Приходя на вход, корпус текста после предварительной обработки приобретает вид входного набора данных, содержащего множество предложений. Для повышения эффективности процесса извлечения ключевых слов онтология имеет список множества имеющихся ключевых слов (*Key Words, KWs*), что

позволяет в дальнейшем на основе процедур нормализации и определения их синонимии находить прецеденты, ускоряющие работу интеллектуальной системы обработки и анализа текстов (рис. 2).



Рис. 2. Компонент онтологии для извлечения и обработки ключевых слов

Таким образом, представленный на рис. 2 компонент верхнеуровневого описания онтологии знаний включает в себя множество предложений, поступивших на вход интеллектуальной системы (3), и множество ключевых слов [13–17], извлеченных и сохраненных в онтологии (4):

$$Sent = \{sent_1, \dots, sent_n\}, \quad (3)$$

$$KW_s = \{kw_1, \dots, kw_m\}. \quad (4)$$

На данных множествах построено бинарное отношение инцидентности (5) между *Sent* и *KW_s*:

$$Rel_{SentKW_s} \subseteq Sent \times KW_s. \quad (5)$$

В онтологии обязательно должно храниться предложение, из которого было выбрано ключевое слово (*KW*). Также необходимо хранить информацию о том, является *KW* объектом или субъектом, так как это сильно влияет на векторную репрезентацию *KW*. Вектор *KW* так же хранится в онтологии. Каждое вновь извлеченное ключевое слово попадает в онтологию либо новым элементом множества *KW_s*, либо как синоним уже существующего *KW*. При этом, необходимо учитывать лингвистические особенности нового элемента онтологической модели (часть речи, род, число, лицо и т.п.), так как эта информация значительно влияет на смысл предложения (рис. 3) [18–21]. Например, в следующих предложениях: «Я системный администратор»; «Я выполняю функции системного администратора»; «Я отвечаю за системное администрирование»; «Он хотел бы стать системным администратором», – при одинаковых (или почти одинаковых) ключевых словах определяется разный смысл. Нужен каталог *KW_s* для каждой доменной (предметной) области, так как *KW_s* нужны до начала разбора предложения. Все слова в онтологии надо делить на категории «значимые» (*Meaningful Words, MWs*) и «уточняющие» (*Clarifying Words, CWs*). Для «значимых» слов должны быть сформированы списки синонимов, а также учтены лингвистические особенности.

Онтология имеет эффективный механизм управления ключевыми словами, что позволяет организовать поиск синонимов (их «склеивание») и, при необходимости, проводить модификацию онтологии (изменение системы внутренних связей между понятиями (концептами)) [20, 21]. Для этого построена иерархия ключевых слов, в которой на верхнем уровне находится поисковое ключевое слово (*Search Key Word, SKW*), включенное в запрос пользователя на входе интеллектуальной системы, например, «транспортное предприятие». Затем, после обработки предложений, на следующем уровне иерархии появляется набор актуальных ключевых слов (*Actual Key Word, AKW*), например, более широкое понятие – «транспортно-логистический кластер» или «транспортный холдинг» и т.п. Данная группа синонимичных *AKW* на следующем уровне иерархии объединяется в единый класс, характерным понятием которого является нормализованное ключевое слово (*Normalized Key Word, NKW*) – часто встречающееся ключевое слово в данном контексте, с наибольшим значением семантической близости, полученной на основе оценки косинусной меры сходства с вектором признаков данного класса (рис. 3).

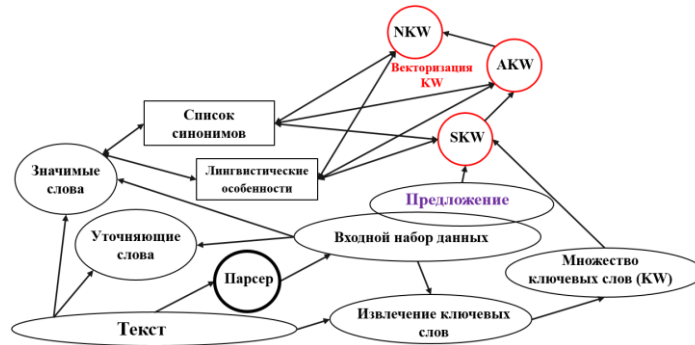


Рис. 3. Расширенный компонент онтологии для извлечения и обработки ключевых слов

Таким образом, представленный на рис. 3 расширенный компонент верхнеуровневого описания онтологии знаний включает в себя множество «значимых» слов с точки зрения извлекаемого смысла MWs (6), множество «уточняющих» слов CWs (7), множество поисковых ключевых слов $SKWs$ (включает в себя извлекаемое из предложений множество KWs , а также ключевые слова, поступающие в интеллектуальную систему через запросы пользователей) (8), множество актуальных ключевых слов $AKWs$ (9), множество нормализованных ключевых слов $NKWs$ (10):

$$MWs = \{mw_1, \dots, mw_e\}, \quad (6)$$

$$CWs = \{cw_1, \dots, cw_j\}, \quad (7)$$

$$SKWs = \{skw_1, \dots, skw_x\}, \quad (8)$$

$$AKWs = \{akw_1, \dots, akw_y\}, \quad (9)$$

$$NKWs = \{nkw_1, \dots, nkw_z\}, \quad (10)$$

На данных множествах построены следующие бинарные отношения инцидентности (11) и (12):

$$Rel_{SentMWs} \subseteq Sent \times MWs, \quad (11)$$

$$Rel_{SentCWs} \subseteq Sent \times CWs, \quad (12)$$

а также следующие бинарные отношения инцидентности (13)-(17) между синонимами:

$$Syn_{SKWsAKWs} \subseteq SKWs \times AKWs, \quad (13)$$

$$Syn_{SKWsNKWs} \subseteq SKWs \times NKWs, \quad (14)$$

$$Syn_{AKWsNKWs} \subseteq AKWs \times NKWs, \quad (15)$$

$$Syn_{AKWsAKWs} \subseteq AKWs \times AKWs, \quad (16)$$

$$Syn_{MWsMWs} \subseteq MWs \times MWs, \quad (17)$$

Глаголы (*Verbs*) тоже хранятся в онтологии, а также имеют нормализованный вид (*NVerbs*) и взаимосвязи с определенными KWs на основе выявленных закономерностей и зависимостей между ними (иногда один глагол может быть связан с несколькими объектами). В онтологии описаны критерии принадлежности к определенной группе контекстов для кластеризации нормализованных смыслов. Предложения с определенными в них KWs распределены по контекстам (предметным областям). Предусмотрена фильтрация слов (наиболее часто встречающихся), определяющих конкретный контекст. Сам контекст определяется на основе «реального смысла» *Real Meaning* (триплета «*субъект – глагол – объект*» («*Sbj – Verb – Obj*»)), построенного на основе слов взятых непосредственно из предложения) (рис. 4). Наиболее проблемной задачей является построение кластеров определенных контекстов.

Таким образом, полное верхнеуровневое описание онтологии знаний для интеллектуальных систем обработки и анализа текстов (рис. 4) включает в себя множества глаголов ($Verbs$) (18) и нормализованных глаголов ($NVerbs$) (19):

$$Verbs = \{verb_1, \dots, verb_s\}, \quad (18)$$

$$NVerbs = \{nverb_1, \dots, nverb_h\}. \quad (19)$$

Отношения синонимии между элементами данных множеств представлено следующим бинарным отношением инцидентности (20):

$$Syn_{VerbsNVerbs} \subseteq Verbs \times NVerbs. \quad (20)$$

В построенной онтологии определены ограничения и правила. Например, можно установить правило, что каждое предложение должно содержать как минимум одно существительное и один глагол. Это поможет проверять полноту и состоятельность онтологической структуры при решении задач извлечения знаний. Необходимо предусмотреть возможность масштабируемости онтологии, что позволит добавлять и изменять ее компоненты.

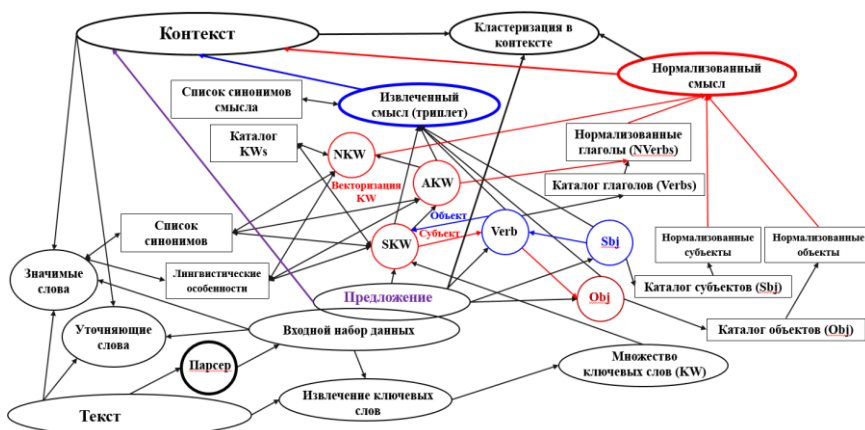


Рис. 4. Верхнеуровневое описание модели онтологии знаний для интеллектуальных систем обработки и анализа текстов

Имеющиеся в онтологии субъекты (Sbj) и объекты (Obj) наполняют множество концептов (понятий) ($K_i, i = 1 \dots N$, где N – количество концептов) с заданной на нём системой связей (бинарных отношений) ($R_q, q = 1 \dots M$, где M – количество связей), являющихся по своей сути несимметричными семантическими отношениями, как это было показано в главе 1.

Конечное множество лексических меток (словарь онтологии) представлено множеством L (21):

$$L = \{L_1, \dots, L_u\}. \quad (21)$$

Выражением (22) представлено антисимметричное, транзитивное, нереклексивное бинарное отношение, являющееся отношением частичного порядка на множестве понятий (концептов) K :

$$Rel_K \subseteq K \times K, Rel_K \in K. \quad (22)$$

Бинарное отношение инцидентности между множествами L и K задано следующим выражением (23):

$$Rel_{LK} \subseteq L \times K. \quad (23)$$

Бинарное отношение инцидентности между множествами L и R задано следующим выражением (24):

$$Rel_{LR} \subseteq L \times R. \quad (24)$$

Тогда под формальной моделью онтологии верхнего уровня описания будем понимать следующий кортеж (25):

$$O = \langle K, R, L, I, Sent, KW, Word, Verb, Rel, Syn, A \rangle, \quad (25)$$

где I – множество экземпляров понятий; $KW = \{KW_s, SKW_s, AKW_s, NKW_s\}$ – интегрированное множество ключевых слов; $Word = \{MW_s, CW_s\}$ – интегрированное множество слов в онтологии; $Verb = \{Verbs, NVerbs\}$ – интегрированное множество глаголов в онтологии; $Rel = \{Rel_k, Rel_{LK}, Rel_{LR}, Rel_{SentKW_s}, Rel_{SentMW_s}, Rel_{SentCW_s}\}$ – интегрированное множество бинарных отношений между элементами онтологии; $Syn = \{Syn_{SKW_sAKW_s}, Syn_{SKW_sNKW_s}, Syn_{AKW_sNKW_s}, Syn_{AKW_sAKW_s}, Syn_{MW_sMW_s}, Syn_{VerbsNVerbs}\}$ – интегрированное множество бинарных отношений синонимии между элементами онтологии; A – аксиомы онтологии.

Таким образом, в данном пункте работы построена верхнеуровневая модель онтологии знаний, которая отличается от известных аналогов применением оригинальной компонентной архитектуры, позволяющей обеспечить высокий уровень детализации спецификаций анализируемой текстовой информации. Представленная модель даёт понимание термина «онтология» в зависимости от контекста текстовой информации и целей его анализа и обработки. В этом заключается гибкость интеллектуальных систем обработки и анализа текстов на естественном языке, повышающая эффективность процессов извлечения знаний.

4. Результаты вычислительного эксперимента. Предложенная верхнеуровневая модель онтологии протестирована при решении задачи структурирования знаний, основным этапом которой является извлечение ключевых слов. Онтология построена в специально созданном программном приложении на основе декомпозиции информации о различных вариантах возникновения предпосылок чрезвычайных ситуаций, а также действий по ликвидации их возможных последствий.

Исследователь задает произвольное количество уровней декомпозиции. Такая организация информационной модели позволяет выделить уровни детализации анализируемой информации. Каждый узел онтологии имеет привязку к определенному элементу информации.

Первому уровню онтологии принадлежит только один понятийный узел. Он является корнем модели и обобщает все ее содержимое, являясь метаинформацией (контекстом). Каждый новый уровень раскрывает содержание концептов (понятий) предыдущего уровня, тем самым раскрывая для исследователя всю предметную область и обнаруживая наличие междисциплинарных связей с другими узлами информационной модели.

Онтология включает в себя множество предметных областей $F = \{F_1, F_2, F_3 \dots F_z\}$, где $z = \overline{1, m}$, m – integer constant (количество предметных областей), причем, $F_z = \{O_{z1}, O_{z2}, O_{z3}, \dots, O_{zk_z}\}$, где k_z – integer constant (количество элементов (концептов) знаний в F_z), причем, в случае задания отношений между концептами одной предметной области получим, что $\forall i, j \rightarrow i = \overline{1, k_z}; j = \overline{1, k_z}; i \neq j \exists C_z[i][j] = n$, где n – количество связей между O_{zi} и O_{zj} , z – integer constant (номер исследуемой предметной области). В случае задания всех, в т.ч. междисциплинарных, отношений между концептами предметных областей получим, что $\forall a, b \rightarrow a = \overline{1, \sum_{z=1}^m k_z}; b = \overline{1, \sum_{z=1}^m k_z}; a \neq b \exists C_{interdisc}[a][b] = p$, где p – количество всех, в т.ч. междисциплинарных, связей между концептами.

Алгоритм структурирования знаний на основе извлечения ключевых слов, необходимый для анализа и обработки предложенной информационной модели, описан в работе [7]. В результате проведенных исследований получена следующая зависимость временной сложности работы алгоритма от числа вершин онтологии (рис. 5).

Временная сложность алгоритма в представленном примере составила $O(n^2)$, где n – количество анализируемых алгоритмом входных данных. По мнению автора в большинстве реальных случаев информационной поддержки процессов предупреждения и/или ликвидации последствий чрезвычайных ситуаций не понадобится онтология большой размерности. Более вероятным диапазоном используемого количества вершин в он-

тологии будет $n \in [5000: 50000]$, в этом случае ВСА алгоритма структурирования информации при поддержке принятия решений по предупреждению и ликвидации последствий чрезвычайных ситуаций составит $O(n)$.



Рис. 5. Зависимость времени работы метода от числа вершин семантической сети

Заключение. В данной статье представлена разработка верхнеуровневого описания модели онтологии знаний для интеллектуальных систем обработки и анализа текстов на естественном языке, построенной на основе оригинальной компонентной архитектуры, обеспечивающей необходимый уровень детализации спецификаций анализируемой текстовой информации. Актуальность данной задачи обусловлена необходимостью развития теоретических основ построения информационных моделей семантических зависимостей внутри текстов на естественном языке.

В статье представлен аналитический обзор задач поиска, приобретения и использования знаний при анализе и обработке текстов на естественном языке. В качестве примера прикладной задачи использования приобретенных знаний, в данном исследовании, рассматривается значимая проблема информационной поддержки процессов предупреждения и/или ликвидации последствий чрезвычайных ситуаций.

Основными результатами проведенного исследования являются следующие:

1. Даны определения основным терминам исследуемой предметной области.
2. Представлена формализованная постановка решаемой задачи.
3. Построена верхнеуровневая модель онтологии знаний, которая отличается от известных аналогов применением оригинальной компонентной архитектуры, позволяющей обеспечить высокий уровень детализации спецификаций анализируемой текстовой информации. Представленная модель даёт понимание термина «онтология» в зависимости от контекста текстовой информации и целей его анализа и обработки. В этом заключается гибкость интеллектуальных систем обработки и анализа текстов на естественном языке, повышающая эффективность процессов извлечения знаний.

Для оценки эффективности применения данной информационной модели разработано программное приложение и проведен вычислительный эксперимент. Полученные результаты проведенных экспериментальных исследований подтверждают эффективность предложенной верхнеуровневой модели онтологии знаний для интеллектуальных систем обработки и анализа текстов на естественном языке.

В развитие данного исследования автор планирует в целях детализации системы отношений между элементами предложенной онтологической структуры провести разработку низкоуровневой модели онтологии знаний, отличающейся применением оригинальной структуры связей между концептами, которая позволит строить множество смысловых паттернов и проводить оценку их семантической близости.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Барсегян А., Куприянов М., Степаненко В., Холод И. Технологии анализа данных: Data Mining, Text Mining, Visual Mining, OLAP. – 2-е изд. – СПб.: БХВ-Петербург, 2008.
2. Кравченко Д.Ю., Кравченко Ю.А., Курейчик В.В., Марков В.В. Математическое описание процесса поддержки принятия решений при оценке семантической близости знаний в конкретизированной модели онтологии // Современные компьютерные технологии: Матер. II научно-методической конференции НПП. – Таганрог: Изд-во ЮФУ, 2021. – С. 25-28.

3. *Хорошевский, В.Ф.* Семантические технологии: ожидания и тренды // Открытые семантические технологии проектирования интеллектуальных систем. – 2012. – № 2. – С. 143-158.
4. *Кравченко Ю.А.* Концептуальные основы рефлексивно-адаптивного подхода к построению интеллектуальных информационных систем // Известия ЮФУ. Технические науки. – 2011. – № 7 (120). – С. 167-171.
5. *Корогодина В.И., Корогодина В.Л.* Информация как основа жизни. – Дубна: Издательский центр «Феникс», 2000. – 208 с.
6. *Кравченко Ю.А.* Информационные модели приобретения знаний и методы их классификации, структурирования, интеграции и семантического поиска: дис. ... д-ра техн. наук: 05.13.17. – Таганрог: 2021. – 314 с.
7. *Кравченко Д.Ю., Герасименко Е.М., Кравченко Ю.А., Кулиев Э.В.* Поддержка принятия решений по предупреждению и ликвидации последствий чрезвычайных ситуаций на основе нечеткого метода структурирования информации // Известия ЮФУ. Технические науки. – 2023. – № 2 (232). – С. 201-212.
8. *Agarwal M.* An Overview of Natural Language Processing // International Journal for Research in Applied Science and Engineering Technology. – 2019. – No. 7. – P. 2811-2813.
9. *Наумов В.Н.* Анализ данных и машинное обучение. Методы и инструментальные средства. – СПб: ИПЦ СЗИУ РАНХиГС, 2020. – 260 с.
10. *Пимешков В.К., Шишаев М.Г.* Методы извлечения знаний из естественно-языковых текстов // Тр. Кольского научного центра РАН. Серия: Технические науки. – 2022. – Т. 13, № 2. – С. 31-45.
11. *Mao X., Huang S., Li R., Shen L.* Automatic keywords extraction based on co-occurrence and semantic relationships between words // IEEE Access. – 2020. – Vol. 8. – P. 117528-117538.
12. *Kravchenko Yury, Mansour Ali, Mohammad Juman, Silega Nemury, Kravchenko Daniil.* Harnessing key phrases in constructing a concept-based semantic representation of text using clustering techniques // Lecture Notes in Computer Science. – Vol. 14335. – P. 190-201.
13. *Kravchenko D.Y., Kravchenko Y.A., Kursitys I.O.* Architecture and Method of Integrating Information and Knowledge on the Basis of the Ontological Structure // Advances in Intelligent Systems and Computing. – 2018. – Vol. 658. – P. 93-103.
14. *Kravchenko D.Y., Kravchenko Yu.A., Kursitys I.O.* Ontological Approach for Designing a Multi-agent Behavior Model in the Internet Environment // Journal of Physics: Conference Series. – 2019. – Vol. 1333. – No. paper 032043.
15. *Ефремова О.А., Павлов С.В.* Онтологическая модель интеграции разнородных по структуре и тематике пространственных баз данных в единую региональную базу данных // Онтология проектирования. – 2017. – Т. 7, № 3 (25). – С. 323-333.
16. *Kozierkiewicz-Hetmanska A., Pietranik M.* The Knowledge Increase Estimation Framework for Ontology Integration on the Concept Level // Journal of Intelligent and Fuzzy Systems. – 2017. – Vol. 32. – P. 1161-1172.
17. *Кравченко Ю.А., Бранченко Д.С.* Исследование преимуществ онтологического подхода в семантическом анализе данных // Информационные технологии, системный анализ и управление. – 2016. – Т. 1. – С. 146-151.
18. *Brandt S., Kontchakov R., Ryzhikov V. et al.* Ontology-Based Data Access with a Horn Fragment of Metric Temporal Logic // 31st Conference on Artificial Intelligence. – 2017. – Vol. 31. – P. 1-17.
19. *Eiter T., Parreira J.X., Schneider P.* Spatial Ontology Mediated Query Answering over Mobility Streams // The Semantic Web - 14th International Conference. – 2017. – Part I. – Vol. 10249. – P. 219-237.
20. *Кравченко Д.Ю., Кравченко Ю.А., Мансур А.М., Мохаммад Ж.Х.* Метод автоматического извлечения ключевых слов // Тр. международного научно-технического конгресса «Интеллектуальные системы и информационные технологии – 2022» («ИС & ИТ-2022», «IS&IT'22»). Т. 1. Научное издание. – Таганрог: Изд-во Ступина С.А., 2022. – С. 90-97.
21. *Кравченко Д.Ю., Кравченко Ю.А., Мансур А.М., Мохаммад Ж.Х.* Модифицированный метод построения семантического представления текста на основе методов кластеризации и взвешивания терминов // Тр. XII международной научно-технической конференции «Технологии разработки информационных систем (ТРИС-2022)». – Таганрог: 2022. – С. 94-100.

REFERENCES

1. *Barsegyan A., Kupriyanoa M., Stepanenko V., Kholod I.* Tekhnologii analiza dannykh: Data Mining, Text Mining, Visual Mining, OLAP [Data analysis technologies: Data Mining, Text Mining, Visual Mining, OLAP]. 2nd ed. St. Petersburg: BKhV-Peterburg, 2008.

2. Kravchenko D.Yu., Kravchenko Yu.A., Kureychik V.V., Markov V.V. Matematicheskoe opisanie protsessa podderzhki prinyatiya resheniy pri otsenke semanticheskoy blizosti znaniy v konkretizirovannoy modeli ontologii [Mathematical description of the decision support process when assessing the semantic proximity of knowledge in a specified ontology model], *Sovremennye komp'yuternye tekhnologii: Mater. II nauchno-metodicheskoy konferentsii NPR* [Modern computer technologies: Proceedings of the II scientific and methodological conference of scientific and scientific research]. Taganrog: Izd-vo YuFU, 2021, pp. 25-28.
3. Khoroshevskiy, V.F. Semanticheskie tekhnologii: ozhidaniya i trendy [Semantic technologies: expectations and trends], *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh sistem* [Open semantic technologies for designing intelligent systems], 2012, No. 2, pp. 143-158.
4. Kravchenko Yu.A. Kontseptual'nye osnovy refleksivno-adaptivnogo podkhoda k postroeniyu intellektual'nykh informatsionnykh sistem [Conceptual foundations of a reflexive-adaptive approach to the construction of intelligent information systems], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2011, No. 7 (120), pp. 167-171.
5. Korogodin V.I., Korogodina V.L. Informatsiya kak osnova zhizni [Information as the basis of life]. Dubna: Izdatel'skiy tsentr «Feniks», 2000, 208 s.
6. Kravchenko Yu.A. Informatsionnye modeli priobreteniya znaniy i metody ikh klassifikatsii, strukturirovaniya, integratsii i semanticheskogo poiska: dis. ... d-ra tekhn. nauk [Information models of knowledge acquisition and methods of their classification, structuring, integration and semantic search: dr. of eng. sc. diss.]: 05.13.17. Taganrog: 2021, 314 p.
7. Kravchenko D.Yu., Gerasimenko E.M., Kravchenko Yu.A., Kuliev E.V. Podderzhka prinyatiya resheniy po preduprezhdeniyu i likvidatsii posledstviy chrezvychaynykh situatsiy na osnove nechetkogo metoda strukturirovaniya informatsii [Support for decision-making to prevent and eliminate the consequences of emergency situations based on the fuzzy method of information structuring], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2023, No. 2 (232), pp. 201-212.
8. Agarwal M. An Overview of Natural Language Processing, *International Journal for Research in Applied Science and Engineering Technology*, 2019, No. 7, pp. 2811-2813.
9. Naumov V.N. Analiz dannykh i mashinnoe obuchenie. Metody i instrumental'nye sredstva [Data analysis and machine learning. Methods and tools]. St. Petersburg: IPTS SZIU RANKHiGS, 2020, 260 p.
10. Pimeshkov V.K., Shishaev M.G. Metody izvlecheniya znaniy iz estestvenno-yazykovykh tekstov [Methods for extracting knowledge from natural language texts], *Tr. Kol'skogo nauchnogo tsentra RAN. Seriya: Tekhnicheskie nauki* [Proceedings of the Kola Scientific Center of the Russian Academy of Sciences. Series: Technical Sciences], 2022, Vol. 13, No. 2, pp. 31-45.
11. Mao X., Huang S., Li R., Shen L. Automatic keywords extraction based on co-occurrence and semantic relationships between words, *IEEE Access*, 2020, Vol. 8, pp. 117528-117538.
12. Kravchenko Yury, Mansour Ali, Mohammad Juman, Silega Nemury, Kravchenko Daniil. Harnessing key phrases in constructing a concept-based semantic representation of text using clustering techniques, *Lecture Notes in Computer Science*, Vol. 14335, pp. 190-201.
13. Kravchenko D.Y., Kravchenko Y.A., Kursitys I.O. Architecture and Method of Integrating Information and Knowledge on the Basis of the Ontological Structure, *Advances in Intelligent Systems and Computing*, 2018, Vol. 658, pp. 93-103.
14. Kravchenko D.Y., Kravchenko Yu.A., Kursitys I.O. Ontological Approach for Designing a Multi-agent Behavior Model in the Internet Environment, *Journal of Physics: Conference Series*, 2019, Vol. 1333, No. paper 032043.
15. Efremova O.A., Pavlov S.V. Ontologicheskaya model' integratsii raznorodnykh po strukture i tematike prostranstvennykh baz dannykh v edinuyu regional'nuyu bazu dannykh [Ontological model for integrating spatial databases of heterogeneous structure and subject matter into a single regional database], *Ontologiya proektirovaniya* [Design Ontology], 2017, Vol. 7, No. 3 (25), pp. 323-333.
16. Kozierkiewicz-Hetmanska A., Pietranik M. The Knowledge Increase Estimation Framework for Ontology Integration on the Concept Level, *Journal of Intelligent and Fuzzy Systems*, 2017, Vol. 32, pp. 1161-1172.
17. Kravchenko Yu.A., Branchenko D.S. Issledovanie preimushchestv ontologicheskogo podkhoda v semanticheskoy analize dannykh [Study of the advantages of the ontological approach in semantic data analysis], *Informatsionnye tekhnologii, sistemnyy analiz i upravlenie* [Information technologies, system analysis and management], 2016, Vol. 1, pp. 146-151.
18. Brandt S., Kontchakov R., Ryzhikov V. et al. Ontology-Based Data Access with a Horn Fragment of Metric Temporal Logic, *31st Conference on Artificial Intelligence*, 2017, Vol. 31, pp. 1-17.
19. Eiter T., Parreira J.X., Schneider P. Spatial Ontology Mediated Query Answering over Mobility Streams, *The Semantic Web - 14th International Conference*, 2017, Part I, Vol. 10249, pp. 219-237.

20. Kravchenko D.Yu., Kravchenko Yu.A., Mansur A.M., Mokhammad Zh.Kh. Metod avtomaticheskogo izvlecheniya klyuchevykh slov [Method for automatic extraction of keywords], *Tr. mezhdunarodnogo nauchno-tekhnicheskogo kongressa «Intellectual'nye sistemy i informatsionnye tekhnologii – 2022» («IS & IT-2022», «IS&IT'22»)* [Proceedings of the international scientific and technical congress “Intelligent systems and information technologies - 2022” (“IS & IT-2022”, “IS&IT'22”)]. Vol. 1. Scientific publication. Taganrog: Izd-vo Stupina S.A., 2022, pp. 90-97.
21. Kravchenko D.Yu., Kravchenko Yu.A., Mansur A.M., Mokhammad Zh.Kh. Modifitsirovannyi metod postroeniya semanticheskogo predstavleniya teksta na osnove metodov klasterizatsii i vzhivaniya terminov [A modified method for constructing a semantic representation of text based on clustering and term weighting methods], *Tr. XII mezhdunarodnoy nauchno-tekhnicheskoy konferentsii «Tekhnologii razrabotki informatsionnykh sistem (TIS -2022)»* [Proceedings of the XII International Scientific and Technical Conference “Information Systems Development Technologies (TRIS-2022)”]. Taganrog: 2022, pp. 94-100.

Статью рекомендовал к опубликованию к.т.н. С.Г. Буланов.

Кравченко Даниил Юрьевич – Южный федеральный университет; e-mail: dkravchenko@sfedu.ru; г. Таганрог, Россия; тел.: 88634371651; кафедра систем автоматизированного проектирования; аспирант.

Kravchenko Daniil Yurievich – Southern Federal University; e-mail: dkravchenko@sfedu.ru; Taganrog, Russia; phone: +78634371651; the department of computer aided design, postgraduate.

УДК 621.337.11:004.942:519.876.5

DOI 10.18522/2311-3103-2024-2-50-59

П.П. Чернусь, П.П. Чернусь, А.А. Яковлев, Р.В. Сахабудинов, А.С. Голосий

СИНТЕЗ СИСТЕМЫ УПРАВЛЕНИЯ КОМПЛЕКСА ПОДЪЕМА ПОДВОДНОГО ГРУЗА

Для транспортировки подводного груза применяется судно-носитель (СН). Судно носитель оснащается комплексом подъема подводного груза, предназначенным для захвата груза, его подъема, закрепления на СН и транспортировки к пункту базирования. Комплекс подъема подводного груза включает: спускаемый модуль, механизм подъема, механизм фиксации, механизм демпфирования и систему управления. В статье представлены результаты разработки математической модели подводного груза, который описан на основании теорем об изменении количества движения и кинетического момента механической системы. В уравнениях линейного и углового перемещений груза присутствует присоединенная к нему масса жидкости. С учётом относительно малых линейных и угловых скоростей, а также малых углов вращения выполнена линеаризация уравнения динамики в скалярном виде. Модель груза реализована в стандартных блоках системы имитационного моделирования. Для синтеза системы управления груз представлен передаточной функцией в виде апериодического звена второго порядка. В работе синтезирован трехконтурный подчиненный регулятор груза с обратными связями по положению, по скорости и по току. Получены выражения для расчёта обобщённых динамических характеристик замкнутой систем второго порядка в зависимости от относительного коэффициента усиления регулятора, рассчитаны параметры системы. Исследования, проведенные на математических моделях системы, позволили получить начальные сведения о линейном и угловом перемещении спускаемого модуля в установившемся режиме, перемещении точек внешнего и внутреннего подвесов, величине силы на тросах, о моменте и скорости, развиваемых электродвигателями лебедок. Моделирование режимов спуска, стабилизации и подъема позволили скорректировать параметры оборудования и добиться удовлетворительных результатов функционирования комплекса.

Моделирование; подводный груз; масса; инерция; кинетический момент; качка; линеаризация; система управления; передаточная функция; замкнутый контур; Пи-регулятор.