



ИНФОКОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ INFOCOMMUNICATION TECHNOLOGIES

УДК 004.048

DOI 10.52575/2687-0932-2024-51-2-466-478

Об обнаружении ошибочного употребления термина в русскоязычном тексте на основе онтологии предметной области

Кудинов В.А., Бородина М.Е.

Курский государственный университет

Россия, 305000, Российская Федерация, г. Курск, ул. Радищева, д. 33

E-mail: kudinov@kursksu.ru, borodina_me@kursksu.ru

Аннотация. Семантический анализ текста активно исследуется в качестве важного направления компьютерной лингвистики. Достижения в его развитии можно проследить в интеллектуальных системах, осуществляющих обработку текста на естественном языке. Один из видов таких систем направлен на поиск и исправление ошибок в тексте. Однако подобные системы неспособны работать со специальной лексикой научных текстов, следовательно, задача поиска лексико-семантических ошибок, связанных с неправильным употреблением термина, не может быть переложена на интеллектуальную систему, так как для этого отсутствует соответствующее теоретическое и программное решение. В связи с чем целью исследования является формализация задачи обнаружения ошибочного использования термина предметной области в русскоязычном тексте. Постановка задачи следует из её практического применения и предметной области. Предложенная математическая модель задачи описана с помощью понятийного аппарата теории множеств. Осуществлен переход к математической модели онтологии и её логическому описанию, позволяющему в дальнейшем разработать онтологию предметной области для обнаружения ошибочного использования термина в тексте на языке описания онтологий OWL 2. В ходе формализации задачи, также указывающей на использование спроектированной онтологии, предложен механизм сравнения контекстов термина из обрабатываемого текста и онтологии. Произведена оценка ошибок первого и второго рода, сформирован алгоритм принятия решения о наличии или отсутствия связи анализируемого термина и предметной области онтологии. Таким образом, результатом исследования является формализация подхода к обнаружению ошибочного использования термина в русскоязычном научном тексте на основе онтологии предметной области.

Ключевые слова: семантический анализ текста, онтология предметной области, автоматическая обработка текста

Для цитирования: Кудинов В.А., Бородина М.Е. 2024. Об обнаружении ошибочного употребления термина в русскоязычном тексте на основе онтологии предметной области. Экономика. Информатика. 51(2): 466–478. DOI 10.52575/2687-0932-2024-51-2-466-478

About Detecting Incorrect Use of Terminology in Russian Texts Based on Domain Ontology

Vitaly A. Kudinov, Maria E. Borodina

Kursk State University

33 Radishcheva St., Kursk, Russian Federation, 305000, Russia

E-mail: kudinov@kursksu.ru, borodina_me@kursksu.ru

Abstract. Semantic analysis of text is actively explored as an important direction in computational linguistics. Advancements in its development can be traced in intelligent systems that process natural language text. One type

of such systems is aimed at finding and correcting errors in the text. However, these systems are incapable of working with the specialized lexicon of scientific texts. Therefore, the task of detecting lexico-semantic errors related to the incorrect use of terminology cannot be delegated to an intelligent system, as there is a lack of corresponding theoretical and programmatic solutions. Consequently, the research aims to formalize the task of detecting incorrect usage of terminology in the Russian text within a specific subject area. The problem formulation arises from its practical application and the subject area. The proposed mathematical model of the task is described using the conceptual apparatus of set theory. The transition to the mathematical model of ontology and its logical description is carried out, allowing for the further development of the subject area ontology to detect incorrect terminology usage in text using the OWL 2 ontology language. During the formalization of the task, which also implies the use of the designed ontology, a mechanism for comparing the contexts of the term from the processed text and the ontology is suggested. Errors I and II types were assessed, and an algorithm for making a decision about the presence or absence of a connection between the analyzed term and the ontology subject area was formed. Thus, the result of the research is the formalization of an approach to detecting the incorrect use of terminology in Russian scientific texts based on the ontology of the subject area.

Keywords: semantic analysis of text, domain ontology, automatic text processing

For citation: Kudinov V.A., Borodina M.E. 2024. About Detecting Incorrect Use of Terminology in Russian Texts Based on Domain Ontology. Economics. Information technologies, 51(2): 466–478 (in Russian). DOI 10.52575/2687-0932-2024-51-2-466-478

Введение

На сегодняшний день существует множество онлайн-сервисов, обнаруживающих лексические ошибки в русскоязычном тексте [Пикалёв, Вовнянко, Денищенко, 2018]. Работа подобных систем основана на словарях сочетаемости. Более продвинутые сервисы предлагают исправления найденных ошибок. Эти системы объединены популярностью среди редакторов и SEO-специалистов, однако они не позволяют пользователю настроить предметную область, стиль речи для эффективной обработки текста с определенной лексикой.

В учебном пособии «Прикладное программное обеспечение: системы автоматической обработки текстов» Михаилом Георгиевичем Мальковским была рассмотрена одна из наиболее интересных и перспективных задач прикладного программирования – задача автоматической обработки тестов на естественном языке.

Система ЛИНАР (Литературно-Научный Редактор) является интеллектуальной системой комплексного контроля качества и редактирования русскоязычных текстов. В ее формализации была описана следующая проблема: «При написании отдельных фрагментов текста разными авторами для обозначения одной и той же сущности могут быть использованы различные термины, что усложняет понимание текста. Автоматическое обнаружение подобных конфликтов требует привлечения глубоких знаний о понятийном и терминологическом аппарате предметной области, и в ЛИНАР не реализуется» [Мальковский, Грацианова, Полякова, 2000].

1. Постановка задачи

Рассмотрим процесс проверки редактором научного текста. Научный стиль речи характеризуется ограниченным словарем, поэтому в сравнении с художественным текстом научный демонстрирует лексическую бедность. Лексические единицы, формирующие научный текст, можно разделить на общие и специальные. Общие лексические единицы организуют общеупотребительную лексику языка, тогда как специальные являются терминами, «логизированной, подвергнутой социальной обработке и отбору частью специальной лексики» [Митрофанова, 1972].

Если редактор недостаточно компетентен в предметной области, то доступные онлайн-сервисы исправления ошибок в тексте не помогут обнаружить возможное семантическое противоречие, возникшее при употреблении термина в неверном значении, они смогут

произвести работу только над общеупотребительной лексикой [Юргель, 2019]. Для решения этой проблемы необходимо использовать понятийный и терминологический аппарат предметной области [Мальковский, Грацианова, Полякова, 2000], которым является онтология предметной области [Jacquette, 2014].

В качестве неправильного употребления термина в тексте будем рассматривать лексико-семантическую ошибку, когда термин предметной области используется в данном контексте вместо нужного слова из-за незнания или недостаточного знания их лексических значений.

Обнаружение ошибочного употребления термина в тексте сведем к сопоставлению фактического контекста термина с ожидаемым. Для решения задачи потребуется формализация знаний предметной области, представляющая собой спецификацию имен терминов и их значений, следовательно, необходимо разработать онтологию предметной области [Мальковский, Грацианова, Полякова, 2000].

Ограничимся обработкой терминов не в произвольном контексте, а в фиксированном – в контексте их определений, то есть в рамках терминологического словаря. Определения выбранной предметной области не будут содержать формулы.

2. Проектирование онтологии

Множество слов обозначим W (Word), множество терминов – T (Term).

Введем понятие контекста в виде множества C (Context):

$$C = \{c \mid \exists w \in W (w \in c)\}.$$

Таким образом, каждый контекст состоит из слов, одно слово может принадлежать одному и более контексту. Одному контексту соответствует один термин. Получаем взаимно однозначную функцию отображения множества контекстов на множество терминов $g : C \rightarrow T$, при этом отображение множества слов на множество контекстов является сюръективным $f : W \rightarrow C$ [Болтянский, Савин, 2002].

Один термин может быть использован в определении другого термина. Например, термин t_2 входит в состав термина t_1 , следовательно, множество T отображается на себя: $d : T \rightarrow T$.

Проиллюстрируем отображения:

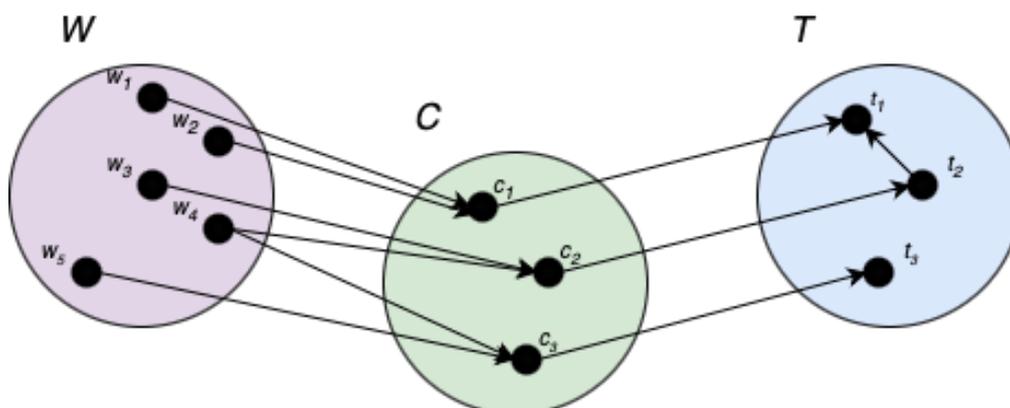


Рис. 1. Отображения g, f, d

Fig. 1. Mappings g, f, d

Исходя из $f : W \rightarrow C$ и $g : C \rightarrow T$ следует, что для каждого элемента $w \in W$ определен его образ $c = f(w), c \in C$. В свою очередь, для этого элемента c определен образ $t = g(c), t \in T$. То есть для элемента w мы можем поставить в соответствие t , минуя c . Получаем отображение множества W на множество T , называемое композицией отображений f и g [Warner, 2009]:

$$(gf)(w) = g(f(w)).$$

Проиллюстрируем пунктирной линией полученную композицию отображений:

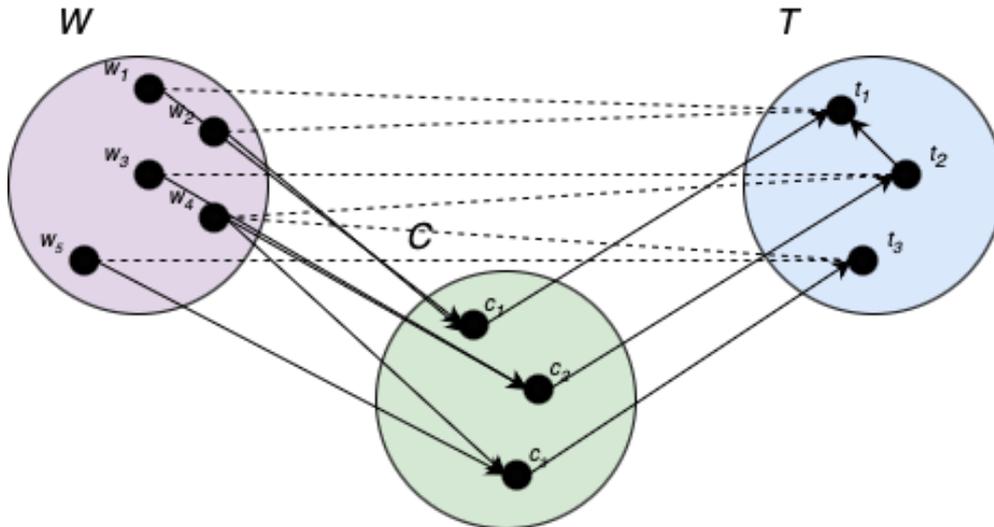


Рис. 2. Отображения g, f композицией g и f (пунктирная линия)
 Fig. 2. Mappings g, f with composition g and f (dotted line)

Математическое определение онтологии [Гаврилова, Хорошевский, 2000]:

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

где:

X – конечное множество концептов (понятий, терминов) предметной области, которую представляет онтология O ;

R – конечное множество отношений между концептами (понятиями, терминами) заданной предметной области;

F – конечное множество функций интерпретации (аксиоматизации).

Осуществим переход к онтологической модели [Thanapalasingam, Osborne, Virukou, Motta, 2018]. Конечное множество концептов будет содержать множества слов, терминов и контекстов: $X = W, C, T$.

Конечное множество отношений между концептами содержит функции отображения g и f : $R = g, f$.

Конечное множество функций интерпретации представляет собой композицию функций отображения g, f и функцию отображения d : $F = (gf)(w), d$.

В дальнейшем потребуется обработка спроектированной онтологии. В данной работе предлагается представление онтологии в виде логической модели, поэтому воспользуемся описательной логикой для её формализации.

Описательные логики – семейство формализмов, используемых для представления знаний [Baader, 2003]. Каждая описательная логика характеризуется своим набором конструкторов и индуктивным правилом, с помощью которого из атомарных концептов и ролей строятся составные концепты и роли соответственно.

В нашем случае для описания онтологии достаточно работать с EL-логикой, допускающей только конструкторы конъюнкции (\wedge), квантора существования (\exists) и концепта «вещь» (T) [Changsheng, 2007]. Добавление конструкторов и, соответственно, переход к другой описательной логике будут повышать вычислительную сложность логической модели [Капустина, Пальчунов, 2019].

С помощью EL-логики опишем множества R и F математической модели онтологии, рассмотренной выше. Набор концептов указан в множестве X .

Множеству R соответствует TBox (терминология) онтологии, опишем её:

$Term \equiv Context$

$Context \equiv Word$

Соответствие термина определенному контексту введем с помощью роли *hasTerm* – контекст1 «имеет термин» термин1, а для ситуации вложенности терминов *includedIn* – термин1 «входит в состав» термин2 [Peng, Tang, Kulmanov, Niu, Hoehndorf, 2022].

Накладываем количественное ограничение для роли *hasTerm*, так как контекст относится только к одному термину [Sikos, 2017]:

$$Context \sqsubseteq \exists hasTerm.T.$$

В таком случае рассматриваемое слово на уровне математической модели будет рассматриваться как два разных элемента w_i и t_i . На уровне онтологии подобного рода ситуации будут отражены в отношениях между понятиями [Salatino, Thanapalasingam, Mannocci, Virukou, Osborne, Motta, 2020].

Для каждого концепта будет определен набор индивидов, представителей данного концепта – АВох (данные) онтологии [Грибова, Шалфеева, 2019]. Таким образом, получим онтологию предметной области, которую обычно представляют в виде графа [Schneider, Schopf, Vladika, Galkin, Simperl, Matthes, 2022]. На рисунке 3 приведен общий пример полученного графа:

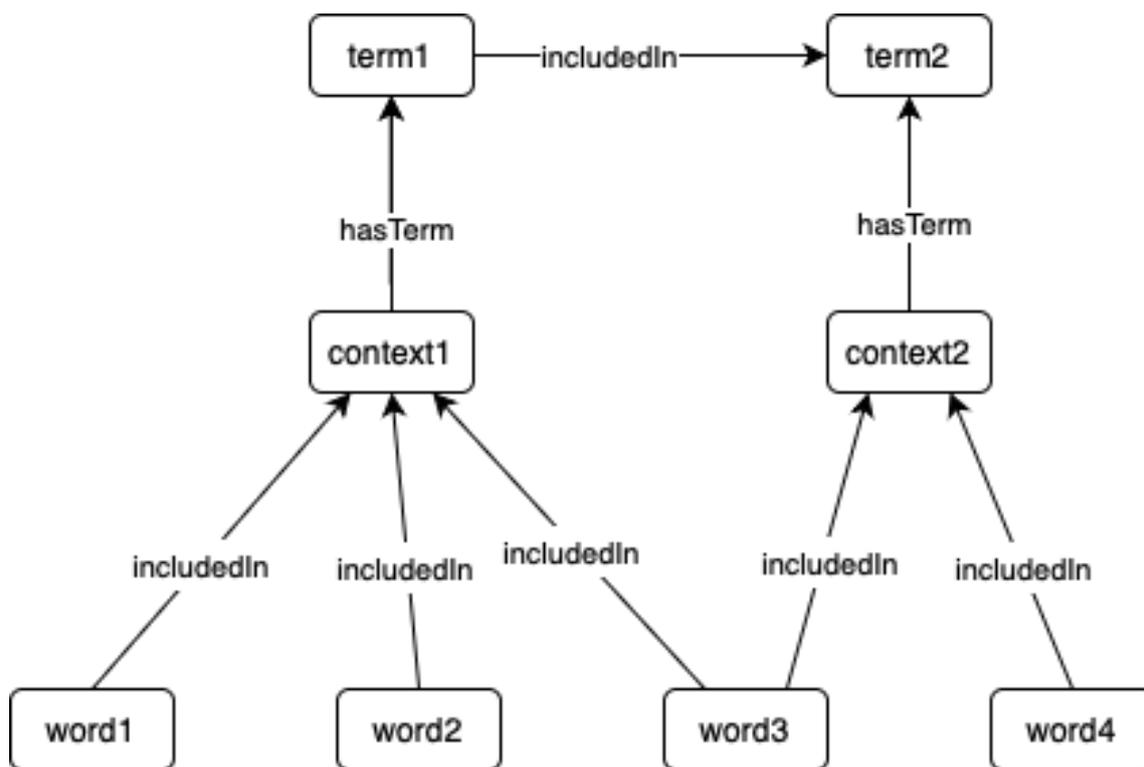


Рис. 3. Пример онтологии предметной области
Fig. 3. An example of a domain ontology

Дальнейшая разработка и заполнение онтологии предлагается производить на языке описания онтологий OWL 2, поддерживающем EL-логику [Баранова, Гоглев, Мигалин, Муштак, 2022].

3. Формализация задачи

В ходе постановки задачи проверка ошибочного употребления термина была сведена к сопоставлению фактического контекста термина с ожидаемым контекстом. Ожидаемым

контекстом являются связанные с термином слова в онтологии предметной области (см п. 2 Проектирование онтологии). Фактический термин будет извлекаться из обрабатываемого текста. Для этого введем понятие знаниевой структуры и рассмотрим процесс преобразования входных данных от цифрового представления научного текста до набора знаниевых структур.

Знаниевая структура представляет собой термин предметной области в качестве главного слова с зависимыми словами, формирующими контекст. Таким образом, она содержит синтаксис и семантику, выделенные из текста.

На рисунке 4 представлены этапы процесса обнаружения ошибочного употребления термина в тексте:

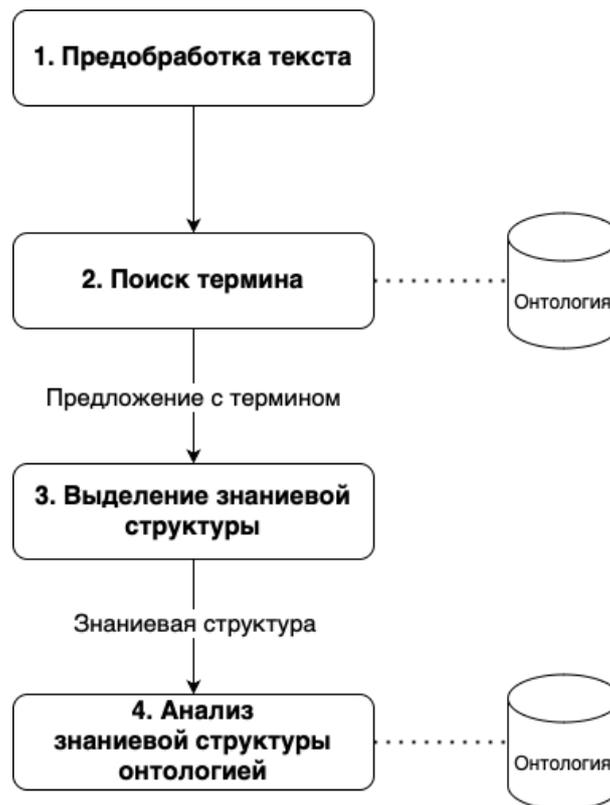


Рис. 4. Процесс обнаружения ошибочного употребления термина предметной области в тексте
Fig. 4. Process of detecting erroneous use of a domain term in a text

Рассмотрим каждый этап.

1. Предобработка текста

Производится токенизация (текст представляется в виде списка предложений со списками слов) и лемматизация (слова заменяются на их начальные формы) текста.

2. Поиск термина

Поиск в тексте термина предметной области со словарем определяет предложения, которые будут обрабатываться в дальнейшем. В качестве словаря выступает онтология предметной области [Wei, Wang, Zhang, Bhatia, Arnold, 2021].

3. Выделение знаниевой структуры

Предложения, содержащие термины предметной области, подвергаются синтаксическому анализу. Из них извлекается знаниевая структура – термин предметной области с зависимыми словами. Результатом синтаксического анализа текста является дерево синтаксического подчинения слов, однако это ресурсозатратный процесс [Chuang, Gupta, Manning, Neer, 2013]. В рамках рассматриваемого исследования достаточно производить частичный синтаксический анализ для поиска зависимых слов обнаруженного термина. Для

извлечения знаниевой структуры можно использовать синтаксический анализатор [Wang, Liu, Desai, Danilevsky, Han, 2014].

4. Анализ знаниевой структуры онтологией

Полученный из текста термин с фактическим контекстом сопоставляется с ожидаемым контекстом из онтологии. Графически обе эти структуры представлены на рисунке 5 [Папуша, 2020], однако обработка фактического контекста будет осуществляться с помощью логического анализа онтологии предметной области, а не графического.

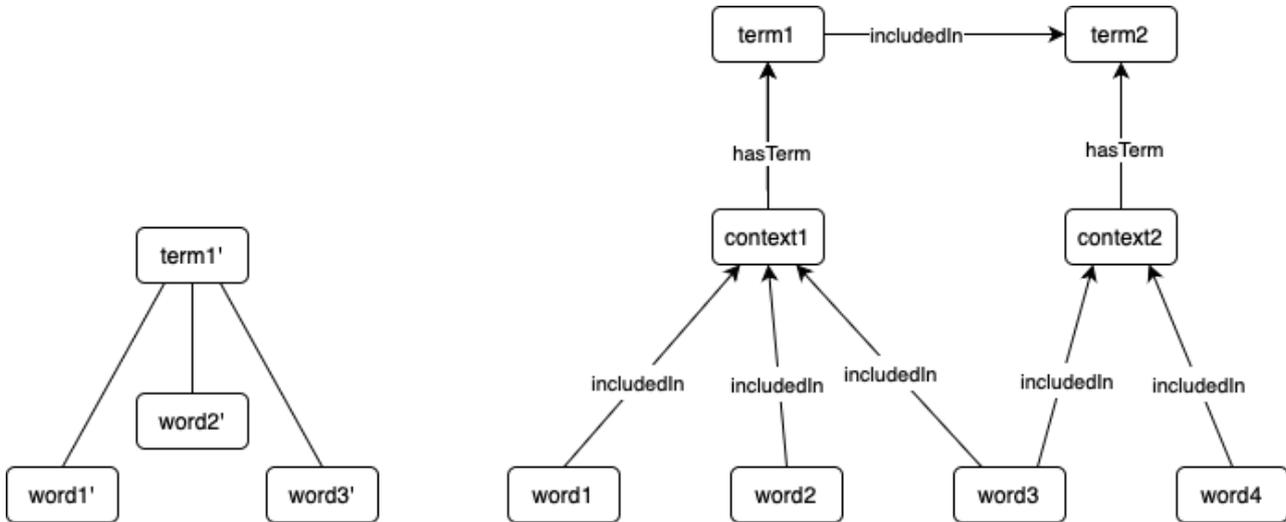


Рис. 5. Знаниевая структура (слева), извлеченная из текста, и онтология предметной области (справа)

Fig. 5. Knowledge structure (left) extracted from text and domain ontology (right)

Предлагается анализировать слова знаниевой структуры и находить их в онтологии предметной области, затем сравнивать термин онтологии, с которым совпало большинство зависимых слов, со словами знаниевой структуры с термином знаниевой структуры. Если термины не совпадают, система в качестве рекомендации для исправления ошибки предлагает ожидаемый термин, то есть термин из онтологии.

Уточним количество слов знаниевой структуры, которое должно совпасть со словами контекста онтологии, чтобы можно было говорить о наличии связи между фактическим и ожидаемым контекстом.

Пусть n – количество слов знаниевой структуры, а m_i – количество слов знаниевой структуры, совпадающее с контекстом i -ого термина онтологии. Случайная дискретная величина X принимает значение 1, если определение анализируемого термина содержится в онтологии, то есть контекст из знаниевой структуры совпал с контекстом термина онтологии, и значение 0, если определение в онтологии отсутствует с вероятностями p_1 и p_2 соответственно.

Получаем $p_1 = \frac{m_i}{n}$, $p_2 = \frac{n - m_i}{n}$. Закон распределения случайной дискретной величины X :

Закон распределения случайной дискретной величины X
 The probability distribution of a discrete random variable X

x_i	0	1
p_i	$1 - p_1$	p_1

Тогда информационная энтропия случайной величины X будет равна [Болтянский, Савин, 2002]:

$$H(X) = -\sum_{i=1}^n p_i \log_2 p_i = -p_1 \log_2 p_1 - (1-p_1) \log_2 (1-p_1).$$

График зависимости значения $H(X)$ от p_1 приведен на рисунке 6.

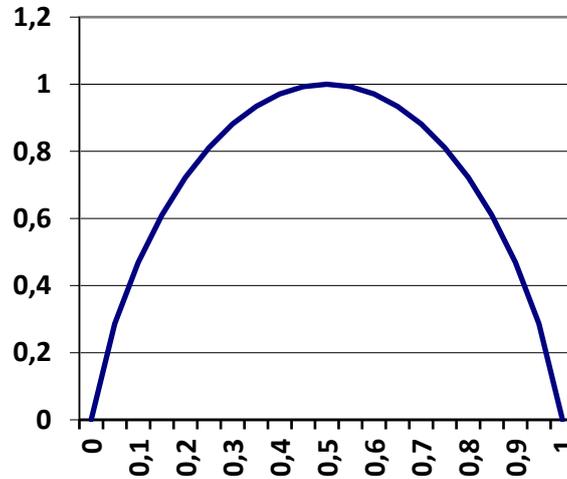


Рис. 6. График зависимости энтропии от вероятности p_1
 Fig. 6. Plot of the relationship between entropy and probability p_1

Максимальное значение энтропии достигается при $p_1 = 0.5$, когда количество слов знамиевой структуры, совпадающих со словами одного из контекстов онтологии, m_i равно половине количества всех слов знамиевой структуры n .

Таким образом, мы можем говорить о наличии связи контекста анализируемого термина с контекстом термина из онтологии, если $p_1 = \frac{m_i}{n} > 0,5$, и утверждать об отсутствии связи, если $p_1 = \frac{m_i}{n} < 0,5$. Однако, в случае $p_1 = \frac{m_i}{n} = 0,5$ получаем максимальное значение неопределенности, появляется необходимость разрешить эту неопределенность.

Нулевая гипотеза (H_0): анализируемый термин принадлежит предметной области онтологии.

Альтернативная гипотеза (H_1): анализируемый термин не принадлежит предметной области онтологии.

Так как мы ограничились обработкой терминов в контексте их определений (см. п. 1 Постановка задачи), альтернативная гипотеза верна, если обнаруженный термин является неоднозначным словом, то есть слово употреблено не в значении, указанном в онтологии [Клышинский, Бунтякова, Карпик, 2021].

Если обнаруженный термин, совпадающий с термином онтологии предметной области, не подвергается анализу, то возможен пропуск ошибочного употребления термина в тексте. Была принята альтернативная гипотеза при истинной нулевой гипотезе, допущена ошибка I рода.

Если производится анализ контекста термина, совпадающего с термином онтологии предметной области, но обладающим другим значением, то результат анализа будет неверным. В этом случае мы приняли нулевую ложную гипотезу, допустив ошибку II рода.

Вычислим вероятности ошибок I и II рода, для этого введём события:

A – термин принадлежит предметной области онтологии;

B – термин не принадлежит предметной области онтологии, являясь омонимом;

C – по произведенной оценке контекста (значение p_1) термин принадлежит предметной области онтологии;

D – по произведенной оценке контекста (значение p_2) термин не принадлежит предметной области онтологии.

Вероятность события B есть ни что иное, как вероятность встретить многозначное слово в тексте, а вероятность события A – вероятность противоположного события.

В «Языке научно-технической литературы» О.Д. Митрофанова [Митрофанова, 1972] утверждает, что пополнение научного словаря происходит в значительной мере не за счет возникновения новых слов, а за счет привлечения слов общего языка. Поэтому обратимся к труду, в котором произведена количественная оценка грамматической неоднозначности некоторых европейских языков [Клышинский, Логачева, Карпик, Бондаренко, 2020].

Для русского языка вероятность встретить в тексте однозначное слово 0,493, то есть: $P(A) = 0,493$, $P(B) = 0,507$.

$P(C) = p_1 = \frac{m_i}{n}$, $P(D) = p_2 = \frac{n - m_i}{n}$ по закону распределения случайной дискретной величины X (см. табл. 1).

Вероятность ошибки первого рода называется уровнем значимости [Михайличенко, 2022]. Она представляет собой произведение вероятностей независимых событий A и D :

$$\alpha = P(A)P(D) = 0,493p_2 = 0,493 \frac{n - m_i}{n}.$$

Вероятность ошибки второго рода:

$$\alpha = P(B)P(C) = 0,507p_1 = 0,507 \frac{m_i}{n}.$$

Уменьшая уровень значимости, увеличиваем вероятность ошибки второго рода. Для разрабатываемого обнаружения ошибочного употребления термина в русскоязычном научном тексте отдаём приоритет уменьшению вероятности ошибки первого рода, и рекомендуем в качестве снижения вероятности ошибки второго рода производить предобработку текста с помощью библиотек синтаксического анализа, содержащих модули снятия омонимии в тексте [Клышинский, Бунтякова, Карпик, 2021].

Возвращаясь к разрешению неопределенности при $p_1 = 0,5$, значение 0,5 отношения m_i и n согласно уменьшению уровня значимости включаем в показатель наличия связи между анализируемым термином и термином онтологии: если $p_1 = \frac{m_i}{n} \geq 0,5$, связь присутствует, иначе – связь отсутствует, дальнейший анализ не производится.

Если в тексте найдено более одного термина предметной области, то производится последовательная обработка каждого предложения.

Возможной ситуацией также является совпадение слов знаниевой структуры со словами более чем одного термина. Назовем подобную ситуацию неоднозначной и приведем в виде блок-схемы на рисунке 7 алгоритм принятия решения о наличии или отсутствии связи анализируемого термина с предметной областью онтологии на этапе процесса обнаружения ошибочного употребления термина, анализа знаниевой структуры онтологией (см. рис. 4).

Пусть t^i – анализируемый термин знаниевой структуры,

k_i, k_j – два максимальных значения отношений $\frac{m_i}{n}$ и $\frac{m_j}{n}$ по убыванию соответственно,

где m_i, m_j количество совпадающих слов контекста с словами i -ого, j -ого терминов онтологии предметной области,

t_i, t_j – i -ый, j -ый термины онтологии предметной области.

Если один термин входит в определение другого термина (находятся в отношении *includen*), то при совпадении хотя бы одного слова знаниевой структуры с контекстом термина, находящегося выше в иерархии, сравниваем проверяемый термин с этим термином.

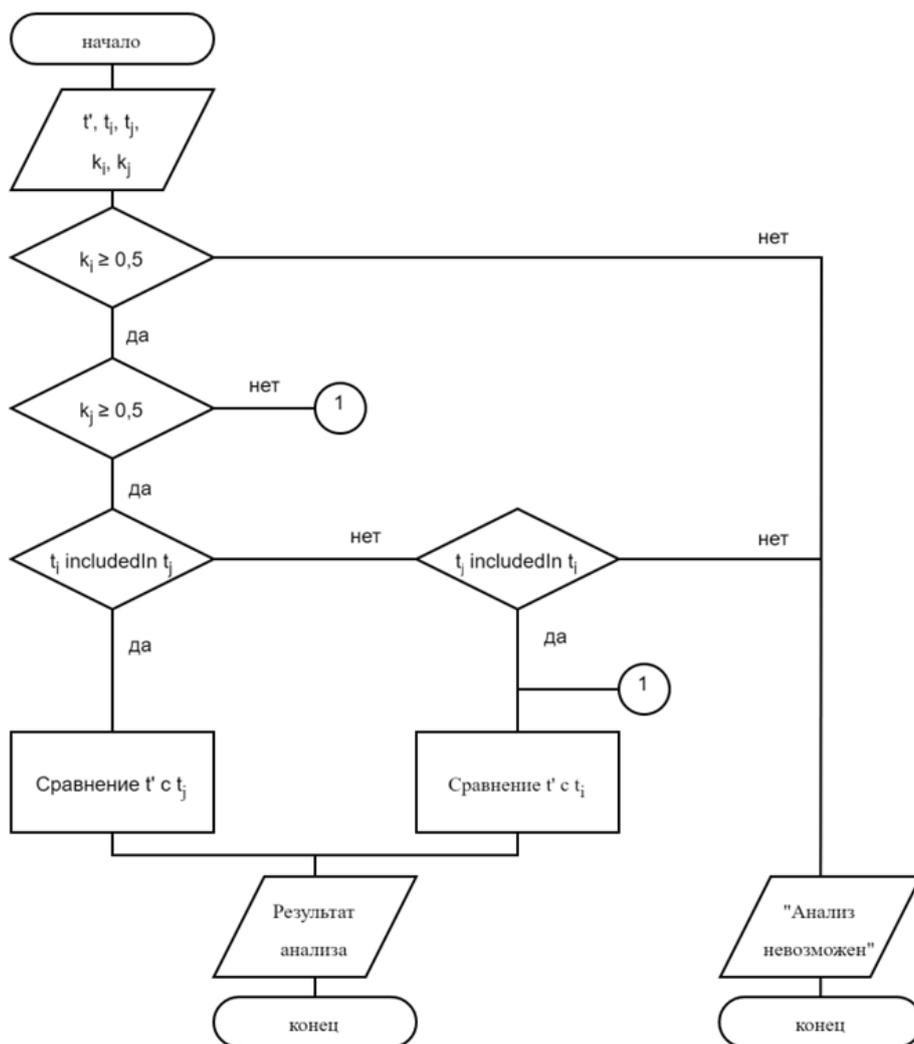


Рис. 7. Алгоритм принятия решения о возможности проведения анализа
 Fig. 7. Algorithm for deciding on the possibility of conducting analysis

Заключение

В работе приведена формализация задачи обнаружения ошибочного употребления термина предметной области в русскоязычном научном тексте. Предложен механизм анализа контекста проверяемого слова, извлекаемого из знаниевой структуры обрабатываемого текста, с помощью онтологии предметной области. Проведена оценка ошибок первого и второго рода и сформирован алгоритм решения неоднозначных ситуаций, когда анализируемый термин не связан с предметной областью онтологии.

В качестве перспективы развития исследования предлагается переход к контекстно-семантическому анализу [Горячкина, 2023], поддерживаемому онтологией предметной области. Таким образом система может обрабатывать термин в произвольном контексте, а не только в контексте определения.

Список источников

Мальковский М.Г., Грацианова Т.Ю., Полякова И.Н. 2000. Прикладное программное обеспечение: системы автоматической обработки текстов. Электронная книга. URL: https://royallib.com/read/malkovskiy_mihail/prikladnoe_programmnoe_obespechenie_sistemi_avtomaticheskoy_obrabotki_tekstov.html (дата обращения: 19 декабря 2023).

Список литературы

- Баранова О.В., Гоглев Н.Н., Мигалин С.А., Муштак О.И. 2022. Подходы к построению и использованию онтологии предметной области интеллектуальной системы управления рисками. *International Journal of Open Information Technologies*, 10(9): 41–52.
- Болтянский В.Г., Савин А.П. 2002. Беседы о математике. Книга 1. М., ФИМА, МЦНМО, 368 с.
- Гаврилова Т.А., Хорошевский В.Ф. 2000. Базы знаний интеллектуальных систем. СПб., Питер, 384 с.
- Горячкина С.Ю. 2023. Методика контекстно-семантического анализа: возможности применения на примере исследования термина ЗОЖ. *Бюллетень Национального научно-исследовательского института общественного здоровья имени Н. А. Семашко*, 2: 69–75.
- Грибова В.В., Шалфеева Е.А. 2019. Онтология диагностики процессов. *Онтология проектирования*, 4(34): 449–461. DOI: 10.18287/2223-9537-2019-9-4-449-461.
- Капустина А.И., Пальчунов Д.Е. 2019. Разработка методов интеграции автоматических средств логического вывода для порождения знаний в онтологической модели. *Вестник НГУ. Серия: Информационные технологии*, 3: 29–42. DOI: 10.25205/1818-7900-2019-17-3-29-42.
- Клышинский Э.С., Бунтякова В.А., Карпик О.В. 2021. Исследование грамматической неоднозначности наиболее частотных слов русского языка. *Препринты ИПМ им. М.В.Келдыша*, 58: с. 22.
- Клышинский Э.С., Логачева В.К., Карпик О.В., Бондаренко А.В. 2020. Количественная оценка грамматической неоднозначности некоторых европейских языков. *Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация*, 1(18): 5–21. DOI: 10.25205/1818-7935-2020-18-1-5-21
- Митрофанова О.Л. 1972. Язык научно-технической литературы. М., МГУ, 147 с.
- Михайличенко А.А. 2022. Аналитический обзор методов оценки качества алгоритмов классификации в задачах машинного обучения. *Ежеквартальный рецензируемый, реферируемый научный журнал «Вестник АГУ»*, 4(311): 52–59. DOI: 10.53598/2410-3225-2022-4-311-52-59
- Папуша С.И. 2020. Онтология и графовые базы данных. *Проблемы экономики и юридической практики*, 3: 268–272.
- Пикалёв Я.С., Вовнянко А.С., Денищенко И.Я. 2018. Анализ автоматических систем проверки правописания русского языка. *Международный рецензируемый научно-теоретический журнал «Проблемы искусственного интеллекта»*, 2(9): 60–67.
- Юргель В.Ю. 2019. Сложности моделирования естественного языка. *Вестник науки и образования*, 23(77): 12–14.
- Baader F. 2007. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge, UK 564 p.
- Changsheng H. 2007. *Non-standard Inference for Explaining Subsumption in the Description Logic EL with General Concept Inclusions and Complex Role Inclusions*. Dresden, Germany, Technische universitat dresden 57 p.
- Chuang J., Gupta S., Manning C., Heer J. 2013. Topic model diagnostics: Assessing domain relevance via topical alignment. *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, 28: 612–620.
- Jacquette D. 2014. *Ontology*. NY, USA, Routledge 368 p.
- Peng X., Tang Z., Kulmanov M., Niu K., Hoehndorf R. 2022. Description Logic EL++Embeddings with Intersectional Closure. <https://doi.org/10.48550/arXiv.2202.14018>
- Salatino A.A., Thanapalasingam T., Mannocci A., Birukou A., Osborne F., E. Motta. 2020. The computer science ontology: A comprehensive automatically-generated taxonomy of research areas. *Data Intelligence*, 2: 379-416. DOI: 10.1162/dint_a_00055.
- Schneider P., Schopf T., Vladika J., Galkin M., Simperl, Matthes F. 2022. A decade of knowledge graphs in natural language processing: A survey. In He, Y., Ji, H., Li, S., Liu, Y., and Chang, C.-H., editors. *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing*, 1: 601–614.
- Sikos L.F. 2017. *Description Logics in Multimedia Reasoning*. Adelaide, Australia, Springer International Publishing AG 205 p. DOI 10.1007/978-3-319-54066-5.
- Thanapalasingam, T., Osborne, F., Birukou, A., Motta, E. 2018. Ontology-Based Recommendation of Editorial Products. *The Semantic Web – ISWC*, 2: 8–12. DOI:10.1007/978-3-030-00668-6_21.
- Wang C., Liu J., Desai N., Danilevsky M., Han J. 2014. Constructing topical hierarchies in heterogeneous information networks. *Knowledge and Information Systems*, 3: 529–558.

- Warner S. 2009. *Set Theory for Beginners: A Rigorous Introduction to Sets, Relations, Partitions, Functions, Induction, Ordinals, Cardinals, Martin's Axiom, and Stationary Sets – Softcover*. Columbia, MD, U.S.A., GreatBookPrices 208 p.
- Wei X., Wang S., Zhang D., Bhatia P., Arnold A.O. 2021. Knowledge enhanced pretrained language models: A comprehensive survey. ArXiv, abs/2110.08455.

References

- Baranova O.V., Goglev N.N., Migalin S.A., Mushtak O.I. 2022. Podhody k postroeniju i ispol'zovaniju ontologii predmetnoj oblasti intellektual'noj sistemy upravlenija riskami [Approaches to construction and use domain ontologies intelligent control system risks]. *International Journal of Open Information Technologies*, 10(9): 41–52.
- Boltjanskij V.G., Savin A.P. 2002. *Besedy o matematike. Kniga 1 [Conversations about mathematics. Book 1]*. M., FIMA, MCNMO, 368 p.
- Gavrilova T.A., Horoshevskij V.F. 2000. *Bazy znaniy intellektual'nyh sistem [Knowledge bases of intelligent systems]*. SPb., Piter, 384 p.
- Goryachkina S.Yu. 2023. The procedure of contextual semantic analysis: the possibilities of application on the example of the study of the term «healthy lifestyle». *Bulletin of Semashko National Research Institute of Public Health*, 2: 69–75. (in Russian)
- Gribova V.V., Shalfeeva E.A. 2019. Ontologija diagnostiki processov [Process diagnostics ontology]. *Ontologija proektirovanija*, 4(34): 449–461. DOI: 10.18287/2223-9537-2019-9-4-449-461.
- Kapustina A.I., Palchunov D.E. 2019. Development of Methods for Integrating Automatic Logical Inference Tools to Generate Knowledge in the Ontological Model. *Vestnik NGU. Serija: Informacionnye tehnologii*, 3: 29–42. (in Russian) DOI: 10.25205/1818-7900-2019-17-3-29-42.
- Klyshinskij Je.S., Buntjakova V.A., Karpik O.V. 2021. Issledovanie grammaticheskoy neodnoznachnosti naibolee chastotnyh slov russkogo jazyka [Study of grammatical ambiguity of the most frequent words in the Russian language]. *Preprinty IPM im. M.V. Keldysha*, 58: p. 22.
- Klyshinskij Je.S., Logacheva V.K., Karpik O.V., Bondarenko A.V. 2020. Quantitative Estimation of Grammatical Ambiguity: Case of European Languages. *Vestnik NGU. Serija: Lingvistika i mezhkul'turnaja kommunikacija*, 1(18): 5–21. (in Russian) DOI: 10.25205/1818-7935-2020-18-1-5-21
- Mitrofanova O.L. 1972. *Jazyk nauchno-tehnicheskoy literatury [Language of scientific and technical literature]*. M., MGU, 147 p.
- Mihajlichenko A.A. 2022. Analytical review of methods for assessing the quality of classification algorithms. *Ezhekvartal'nyj recenziruemyj, referiruemyj nauchnyj zhurnal «Vestnik AGU»*, 4(311): 52–59. (in Russian) DOI: 10.53598/2410-3225-2022-4-311-52-59
- Papusha S.I. 2020. Ontology and graph databases. *Problemy jekonomiki i juridicheskoy praktiki*, 3: 268–272. (in Russian)
- Pikalyov Ya.S., Vovnyanko A.S., Denishenko I.Ya. 2018. Analysis of automatic verification systems spelling of the russian language. *International Peer-Reviewed Scientific Journal «Problems Of Artificial Intelligence» (PAI)*, 2(9): 60–67. (in Russian)
- Jurgel V.Yu. 2019. Complexities of natural language modeling. *Vestnik nauki i obrazovanija*, 23(77): 12–14. (in Russian)
- Baader F. 2007. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge, UK 564 p.
- Changsheng H. 2007. Non-standard Inference for Explaining Subsumption in the Description Logic EL with General Concept Inclusions and Complex Role Inclusions. *Dresden, Germany, Technische universitat dresden* 57 p.
- Chuang J., Gupta S., Manning C., Heer J. 2013. Topic model diagnostics: Assessing domain relevance via topical alignment. *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, 28: 612–620.
- Jacquette D. 2014. *Ontology*. NY, USA, Routledge 368 p.
- Peng X., Tang Z., Kulmanov M., Niu K., Hoehndorf R. 2022. Description Logic EL++Embeddings with Intersectional Closure.
- Salatino A.A., Thanapalasingam T., Mannocci A., Birukou A., Osborne F., E. Motta. 2020. The computer science ontology: A comprehensive automatically-generated taxonomy of research areas. *Data Intelligence*, 2: 379-416. DOI: 10.1162/dint_a_00055.



- Schneider P., Schopf T., Vladika J., Galkin M., Simperl, Matthes F. 2022. A decade of knowledge graphs in natural language processing: A survey. In He, Y., Ji, H., Li, S., Liu, Y., and Chang, C.-H., editors. Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing, 1: 601–614.
- Sikos L.F. 2017. Description Logics in Multimedia Reasoning. Adelaide, Australia, Springer International Publishing AG 205 p. DOI 10.1007/978-3-319-54066-5.
- Thanapalasingam T., Osborne F., Birukou A., Motta E. 2018. Ontology-Based Recommendation of Editorial Products. The Semantic Web – ISWC, 2: 8–12. DOI:10.1007/978-3-030-00668-6_21.
- Wang C., Liu J., Desai N., Danilevsky M., Han J. 2014. Constructing topical hierarchies in heterogeneous information networks. Knowledge and Information Systems, 3: 529–558.
- Warner S. 2009. Set Theory for Beginners: A Rigorous Introduction to Sets, Relations, Partitions, Functions, Induction, Ordinals, Cardinals, Martin's Axiom, and Stationary Sets – Softcover. Columbia, MD, U.S.A., GreatBookPrices 208 p.
- Wei X., Wang S., Zhang D., Bhatia P., Arnold A.O. 2021. Knowledge enhanced pretrained language models: A comprehensive survey. ArXiv, abs/2110.08455.

Конфликт интересов: о потенциальном конфликте интересов не сообщалось.

Conflict of interest: о potential conflict of interest related to this article was reported.

Поступила в редакцию 02.02.2024

Поступила после рецензирования 28.04.2024

Принята к публикации 05.06.2024

Received February 02, 2024

Revised April 28, 2024

Accepted June 05, 2024

ИНФОРМАЦИЯ ОБ АВТОРАХ

Кудинов Виталий Алексеевич, доктор педагогических наук, профессор, профессор кафедры программного обеспечения и администрирования информационных систем, Курский государственный университет, г. Курск, Россия

Бородина Мария Евгеньевна, ассистент кафедры программного обеспечения и администрирования информационных систем, Курский государственный университет, г. Курск, Россия

INFORMATION ABOUT THE AUTHORS

Vitaly A. Kudinov, Doctor of Technical Sciences, Professor, Professor of the Department of Software Maintenance and Administration of Information Systems, Kursk State University, Kursk, Russian Federation

Maria E. Borodina, assistant of the Department of Software Maintenance and Administration of Information Systems, Kursk State University, Kursk, Russian Federation