

У статті описано напівавтоматичний метод розширення базової онтології для предметної області «радіаційна безпека», оснований на принципах NLP. Метод дозволяє розширити онтологію новими екземплярами і відношеннями. Для вирішення проблеми неоднозначності слів було використано словник синонімів. В роботі наводяться результати експерименту, виконаного для поповнення онтології новими екземплярами, знайденими в спеціалізованому текстовому корпусі

Ключові слова: вилучення інформації, правила розпізнавання відношень, правила розпізнавання екземплярів, розширення онтології

В статье описан полуавтоматический метод расширения базовой онтологии для предметной области «радиационная безопасность», основанный на принципах NLP. Метод позволяет расширить онтологию новыми экземплярами и отношениями. Для решения проблемы неоднозначности слов был использован словарь синонимов. В работе представлены результаты эксперимента, выполненного для расширения онтологии новыми экземплярами, обнаруженными в специализированном текстовом корпусе

Ключевые слова: извлечение информации, правила распознавания отношений, правила распознавания экземпляров, расширение онтологии

ПРИМЕНЕНИЕ МАСШТАБНЫХ ЛИНГВИСТИЧЕСКИХ РЕСУРСОВ ДЛЯ РАСШИРЕНИЯ ОНТОЛОГИИ ПРЕДМЕТНОЙ ОБЛАСТИ (НА ПРИМЕРЕ ОБЛАСТИ «РАДИАЦИОННАЯ БЕЗОПАСНОСТЬ»)

Е. А. Орбинская
Аспирант*

Университет им. Братьев Люмьер, Лион, Франция
E-mail: : olena.orobinska@univ-lyon2.fr

Н. В. Шаронова
Доктор технических наук, профессор**
E-mail: nvsharonova@mail.ru

А. Ю. Дорошенко
Аспирант**

E-mail: marykate90@mail.ru

Ж - Ю. Шоша

Заслуженный профессор университета Лион2,
ответственный за кооперацию с Харьковом*

E-mail: Jean-Hugues.Chauchat@univ-lyon2.fr

*Лаборатория ERIC

«Хранение, Представление и Обработка Знаний»

5, ул. Пьер Мендес Франс, Брон Cedex, Франция, 69676

**Кафедра интеллектуальных компьютерных систем

Национальный технический университет

”Харьковский политехнический институт”

ул. Пушкинская, 79/2, г. Харьков, Украина, 61002

1. Введение

Радиационная безопасность – это совокупность мероприятий, регулируемых Международными базовыми нормами по радиационной безопасности. Эти нормы основаны на принципах Международной комиссии по радиационной безопасности, чьей основной задачей является уменьшение риска негативных последствий для здоровья вследствие облучения радиацией или радиоактивными веществами, используемыми для промышленных (производство энергии), медицинских, сельскохозяйственных, образовательных или исследовательских целей. Проводимые исследования сфокусированы на развитии процесса полу-

автоматического получения знаний для расширения и пополнения Базовой Онтологии по Радиационной Безопасности, концептуализация и начальное моделирование которой были выполнены с помощью экспертов предметной области. На рис. 1 показан фрагмент этой онтологии.

Предлагаемая методика была использована для расширения списка концептов базовой онтологии (БО) и установления между ними семантических отношений синонимии. БО была определена с помощью экспертов предметной области, а для ее расширения было создано два параллельных корпуса (русский и французский) на основе стандартов по радиационной безопасности, доступных на разных языках.

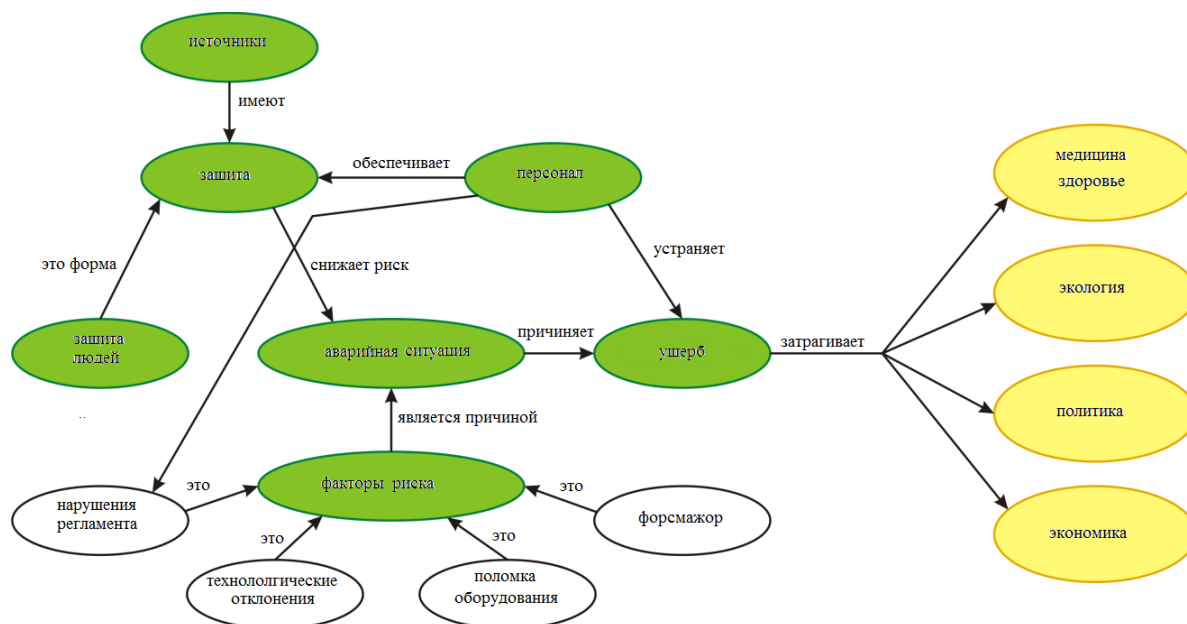


Рис. 1. Фрагмент онтологии по радиационной безопасности

Протестированы были оба корпуса. Следует отметить, что результаты, полученные на французском и русском корпусах, являются сопоставимыми, хотя для первого (французского) корпуса они оказались лучше. В данной работе представлены результаты, полученные на основе текстов на русском языке.

2. Обзор литературы и постановка проблемы

Данный обзор сосредоточен на методах, используемых для автоматического создания онтологии предметной области. Особенностью таких онтологий является то, что их основу составляет терминология предметной области. Традиционно для автоматизированного построения онтологий использовались статистические методы. Позже, благодаря возросшей эффективности синтаксических анализаторов, более широко начали применяться лингвистические методы [1].

В последние годы машиночитаемые лексические ресурсы приобретают все более важную роль. Примерами таких источников являются общие базы знаний, тезаурусы, словари синонимов, многоязычные словари и специализированные глоссарии предметных областей. Обучающие методы предназначены либо для создания онтологии «с нуля», либо для расширения существующей базовой онтологии.

Можно отметить две стратегии для построения онтологии. Первая, так называемая технология «слоеного пирога» (Layer cake) [2], где каждый последующий шаг использует результаты предыдущего этапа. Первым этапом является извлечение терминов-кандидатов и затем на их основе определение отношений и обучающих правил. Существует большое количество платформ, основанных на этом подходе [3, 4]. Другой подход, напротив, предполагает с самого начала использование лингвистической модели, позволяющей обнаружение значений концептов, исходя из заранее определенных отношений [5, 6].

Работы R. Navigli и коллег представляют расширенные свойства крупных лексических ресурсов для онтологического инжиниринга [7, 8]. В настоящее время онтологический инжиниринг сфокусирован на решении следующих проблем: расширение существующих онтологий, определение отношений, классификация отношений. Статистические методы сохраняют лидирующую роль в обнаружении терминов-кандидатов в заданном корпусе [9]. Однако этим методам недостает эффективности для того, чтобы классифицировать термины-кандидаты по типу конкретного семантического отношения, например, сгруппировать все синонимичные существительные.

Дистрибутивный анализ позволяет сгруппировать термины, связанные различными типами отношений [10]. Однако и он ограничен в возможности различить синтагматические (такие как словосочетания или глагольно-именные отношения) и парадигматические отношения (такие как отношения синонимии, гиперонии и т. д.) [11]. Использование лексико-синтаксических шаблонов обеспечивает высокую точность, однако использование таких шаблонов сильно ограничено контекстом, где они могут быть использованы. Как правило, с помощью этого метода может быть обнаружен лишь один тип отношений [12]. Проблема классификации отношений является одной из самых сложных, так как любое семантическое отношение может быть выражено лингвистически множеством способов.

Каждый из перечисленных выше подходов вносит свою лепту в развитие технологий автоматизированного построения онтологий. В действительности, множество различных техник должно быть применено в итеративном обучающем процессе для эффективного сбора знаний. Несмотря на активность исследований, ведущихся в области онтологического инжиниринга, эффективность решения таких задач, как автоматическое расширение онтологии новыми элементами, остается не совсем удовлетворительной. Этот факт объясняется многообразием лексических и синтаксических форм и способов выражения значения разных понятий

и семантических отношений между ними, что делает задачу автоматизированного анализа весьма сложной. С другой стороны, в настоящий момент уже существует достаточно большое количество профессиональных лексикографических ресурсов, где информация, необходимая для тонкого семантического анализа текста, представлена в явном виде, доступном для машинной обработки. Однако разработка эффективных методов, позволяющих использовать такие ресурсы для построения онтологий, находится в самом начале.

3. Цель и задачи исследования

Целью данного исследования является разработка метода использования готового словаря синонимов для расширения базовой онтологии предметной области, а именно обнаружения в корпусе возможных лексических вариантов (т. е. синонимов) исходных концептов, предложенных экспертами области.

Задачей данного исследования является разработка алгоритма расширения БО новыми экземплярами, обнаруженными. Для отбора терминов-кандидатов используется внешний словарь синонимов, что существенно повышает релевантность результатов отбора терминов уже на первом этапе.

4. Описание и реализация метода полуавтоматического расширения базовой онтологии ПО на основе использования внешнего лексического ресурса

Поскольку число концептов верхнего уровня, как правило, невелико [8], мы полагаем, что наилучшим способом для запуска процесса построения онтологии с использованием автоматических методов является использование исходной модели предметной области, реализованной с помощью экспертов данной области. В наиболее простой форме это может быть список основных концептов – базовый словарь. Но впоследствии различные обучающие (т. е. автоматические) методы могут быть активно использованы, чтобы пополнить и расширить базовую онтологию.

Существует несколько путей расширения исходной онтологии:

- добавить новые концепты к уже существующим;
- расширить каждый концепт новыми экземплярами – разными лексическими вхождениями одного понятия, т. е. синонимами;
- расширить виды отношений, связывающих концепты.
- Все эти способы могут быть объединены с тем, чтобы реализовать общий циклический процесс расширения базовой онтологии. Данная работа сфокусирована на втором пути.

Отношения между любыми объектами могут быть выявлены путем сравнения их свойств. Для того чтобы избежать неопределенности слов «свойство», «предикат» и «глагол» в дальнейшем, мы даем их определения. Свойство – это некий атрибут, качество или характеристика объекта. Как отмечается в [13]: «предикат обычно описывает свойство одушевленного или неодушевленного объекта, являющегося подлежащим (субъектом), или описывает ситуацию,

в которой данный субъект играет некоторую роль. В простых предложениях субъект, как правило, указывает на исполнителя действия (одушевленного или неодушевленного), тогда как предикат описывает само действие». В русском языке глаголы играют ключевую роль в описании действий и других свойств субъекта. Другими словами, глаголы явным способом выражают свойства субъекта. Поэтому слова «глагол», «предикат» и «свойство» могут рассматриваться как синонимы. Будем использовать слово «предикат» для глаголов, которые выражают значимые свойства или отношения между концептами онтологии.

Следует отметить, что в любом языке существует множество лексических способов выражения свойств объектов. Например, в двух фразах «*негативные эффекты радиационного облучения*» и «*радиационное облучение вызывает негативные эффекты*» для концепта «*радиационное облучение*» определено свойство «*негативные эффекты*». Однако в первом случае это осуществляется с помощью родительного падежа, а во втором случае с помощью глагола *вызывать* в настоящем времени. Такое разнообразие лингвистических форм существенно осложняет процесс обработки текстов естественного языка. Как отмечалось выше, для выражения свойств концептов были использованы глаголы.

Для достижения поставленной цели и расширения базовой онтологии предметной области (ПО) новыми экземплярами, разработан метод, основанный на использовании внешнего лингвистического ресурса, а именно машиночитаемого словаря синонимов (MRD). Для отбора релевантных терминов-кандидатов и во избежание проблемы неоднозначности адаптирован метод анализа формальных концептов – FCA [14].

Описание метода

Идея заключается в том, чтобы найти в текстовом корпусе все возможные лексические вхождения данного концепта с помощью внешнего лингвистического ресурса, такого как словарь синонимов, так как разные лексические вхождения одного концепта связаны семантически.

Для применения метода необходима реализация нескольких предварительных шагов:

- морфологическая разметка корпуса;
- поверхностный анализ предложений;
- проверка и подтверждение предикатов;
- подтверждение синонимов и маркировка концептов разными лексическими вариантами.

Собственно процесс пополнения онтологии является завершающим этапом. Чтобы проверить и подтвердить результаты, необходимо экспертное вмешательство. Схема метода представлена на рис. 2, и далее подробно описан каждый этап.

Морфологическая разметка

Корпус размечен с помощью синтаксического анализатора TreeTagger, с предварительно установленной моделью русского языка. На выходе, для каждого слова корпуса можно сохранить два типа данных: морфологическую информацию и лемму. TreeTagger для русского языка различает более 750 морфологических форм. Для достижения поставленной цели и упрощения анализа количество разных возвращаемых тегов было уменьшено до 10. Так, при проведении морфологического анализа различается лишь часть речи

каждого слова без дополнительного указания рода, склонения и числа.

Информация о некоторых морфологических тегах дана в табл. 1. Процессы разметки тегами и лемматизации обеспечивают начальную фильтрацию данных, уменьшают разреженность данных и дают дополнительное сглаживание [4]. В эксперименте были использованы леммы. Выходными данными этого шага являются: $W=\{w\}$ – множество всех существительных корпуса и $V=\{v\}$ – и множество всех глаголов корпуса без дубликатов.

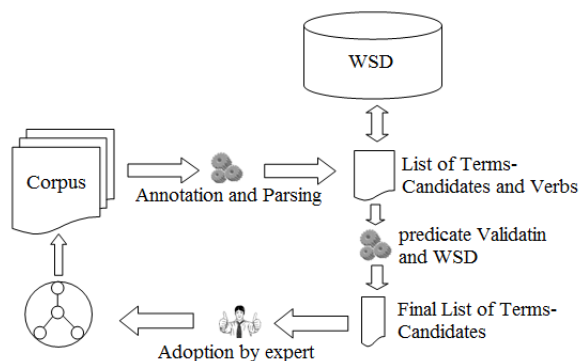


Рис. 2. Схема процесса расширения базовой онтологии ПО

Чтобы отобрать предикаты, необходимо определить критерий их отбора, для чего был рассчитан вес каждого глагола. Значение веса зависит от двух показателей: взаимной информации (1) и модифицированного коэффициента Жаккара (коэффициент производительности) (2). Первый принимает во внимание «внутреннюю зависимость» слов каждой пары по их взаимному распределению

$$MI(w_i, v_j) = p(v_j) * \log(p(v_j)) - p(w_i | v_j) * \log(p(w_i | v_j)). \tag{1}$$

Значение второго коэффициента можно объяснить следующим образом: в соответствии с идеей «семантического поля» [16], (одинаковые) слова имеют разный смысл в разных контекстах. Вследствие того факта, что общее число глаголов намного меньше, чем существительных (в среднем 1:10), большинство глаголов являются слишком абстрактными, они могут соединяться с большим количеством других слов. Поэтому, чтобы уменьшить значение общих глаголов, был рассчитан «коэффициент продуктивности» следующим образом:

$$k_{ij} = \frac{f_{ij}}{N_i + N_j}, \tag{2}$$

где f_{ij} – это число появлений в корпусе пары (w_i, v_j) ; N_i – это число разных глаголов, ассоциированных с существительным w_i и N_j – это число разных существительных, ассоциированных с глаголом v_j . В итоге, объединенный коэффициент рассчитывается как произведение величины взаимной информации и коэффициента продуктивности:

$$K_{ij} = MI(w_i, v_j) * k_{ij}. \tag{3}$$

Подтверждение синонимов и маркировка концептов

В качестве исходных данных используется список концептов верхнего уровня $L=\{l_i | i=1...n, n \in \mathbb{N}\}$, где l_i – это лексический вариант концепта в лексиконе (в списке концептов). Из внешнего словаря извлекается множество списков синонимов $DL = \{DL_i | i = 1...n, n \in \mathbb{N}\}$, найденных в нем для каждого концепта; иначе: каждый DL_i является множеством синонимов из словаря для концепта l_i . Пересечение всех списков синонимов и множества существительных корпуса дает список терминов-кандидатов:

$$CL_i = W \cup DL_i \tag{4}$$

и

$$CL_i := \{w \subseteq W | \forall w : (w, l_i) \in I_{syn}\}, \tag{5}$$

где $I_{syn} \subseteq DL \times W$ означает отношение синонимии между множеством словаря и множеством имен существительных корпуса. Для оценки синонимов был адаптирован метод анализа формальных концептов FCA.

Определение 1. Множество (G, M, I) называется *формальным контекстом*, если G и M являются непересекающимися (разъединенными) множествами, элементы которых связаны бинарным отношением $I: I \subseteq G \times M$. Элементы множества G называются **объектами**, элементы множества M называются **атрибутами**,

Таблица 1

Морфологические теги TreeTagger

Исходные теги TreeTagger	Часть речи	Упрощенный тег
Afc*, Afp*	все формы прилагательных	ADJ
Mc*	все формы числительных	NUM
Nc*	все именные формы	Noun
P.*	все местоимения	PRN
R*	наречия	ADV
Vm*	все глаголы	Verb
Vmps*, Vmpp*	частицы	Part

Примечание: символ «*» соответствует одному либо нескольким любым другим символам.

Поверхностный анализ предложений

Поверхностный анализ каждого предложения – это следующий необходимый шаг обработки. Его задачей является обнаружить именные группы (чтобы распознать возможные терминологические шаблоны [15]), глагольные группы и их взаимное расположение в предложении. Для этого были проиндексированы каждое существительное и глагол. Предположение, что в русском языке большинство предложений, написанных в академическом стиле, имеют линейную структуру с прямым порядком слов: подлежащее – сказуемое – дополнение, позволило записать отношение между каждым субъектом (подлежащим) и его предикатом (сказуемым) в виде пар (w, v) .

Проверка и подтверждение предикатов (отношений)

Из каждого предложения были сохранены пары (w_i, v_j) , где индекс существительного меньше индекса глагола ($i < j$). Таким образом, устанавливаются бинарные отношения между существительными и глаголами. Однако не все глаголы являются релевантными модели предметной области, поскольку не все они представляют какие-либо важные свойства концептов.

а элементы множества I определяют инцидентность объектов и атрибутов или, другими словами, принадлежность атрибута объекту.

Определение 2. Формальный концепт (A, B) является парой разъединенных множеств объектов $A \subseteq G$ и атрибутов $B \subseteq M$ соответственно, если справедливо следующее утверждение: каждый объект из A обладает всеми атрибутами из B и наоборот, все атрибуты из B принадлежат каждому объекту из A.

Пусть имеется формальный контекст (G, M, I). Определим A' для $A \subseteq G$ следующим образом $A' := \{m \subseteq M \mid \forall g \in A: (g, m) \in I\}$ – как множество атрибутов, общих для всех объектов A; аналогично определим B' для $B \subseteq M$ как $B' := \{g \subseteq G \mid \forall m \in B: (g, m) \in I\}$. – как множество всех объектов, обладающих всеми атрибутами B.

Проще говоря, A' – это множество атрибутов, общих для всех объектов множества A, и B' – это множество объектов, обладающих всеми атрибутами из B.

В соответствии с определением 2 пара (A, B) является формальным концептом (G; M; I) тогда и только тогда, когда $A \subseteq G, B \subseteq M, A' = B, B' = A$.

Путем взвешивания для каждого концепта были отобраны предикаты. В данном случае, формальным концептом является множество подтвержденных синонимов и их атрибутов (отобранных глаголов) для каждого понятия предметной области.

5. Описание эксперимента по расширению онтологии лексическими вариантами вхождений концептов

В этом разделе представлено описание применения предлагаемого метода на примерах, полученных для области радиационная безопасность.

Чтобы реализовать эксперимент, использовалось 9 основных понятий ПО – концептов верхнего уровня: $L = \{\text{безопасность, защита, излучение, источник, население, облучение, персонал, риск, ущерб}\}$.

Вначале был создан специализированный текстовый корпус. Он состоит из 58 отраслевых стандартов, норм и отчетов, утвержденных Международной комиссией по радиационной безопасности и Национальной комиссией по радиационной защите Украины. В настоящее время корпус состоит более чем из 600 000 слов. Дополнительно было протестировано несколько тезаурусов и словарей синонимов для русского языка, доступных on-line. Следует отметить, что их качество существенно отличается. В данном случае был использован Полный словарь синонимов русского языка [17]. Список синонимов для некоторых понятий предметной области представлен в табл. 2.

Для концепта «ущерб» в корпусе было обнаружено лишь 7 слов из 22 синонимов в словаре: $SL_{\text{ущерб}} = \{\text{вред, потеря, повреждение, авария, осложнение, ухудшение, убыль}\}$. В корпусе обнаружилось 13 разных глаголов, ассоциированных с концептом “ущерб”: $V = \{\text{включать, возмещать, вызывать, использовать, наносить, обеспечивать, ограничивать, оказаться, превышать, предотвращать, причинять, связывать, сопровождаться}\}$, однако после взвешивания лишь 9 из них были отобраны в качестве предикатов: $\{\text{включать, возмещать, вызывать, наносить, ограничивать, превышать, предотвращать,}$

$\text{причинять, сопровождаться}\}$. Результат обработки этого процесса показан на рис. 3.

Таблица 2

Список синонимов для обозначения концептов предметной области

Понятие	Синонимы из внешнего словаря
Ущерб	убыток, потери, вред, урон, изъясн, потеря, повреждение, поломка, авария, протори и убытки, подрыв, осложнение, утрата, разор, ухудшение, протори, порча, невыгода, пагуба, шкода, брешь, наклад
Излучение	изливание, излитие, источение, свет, испускание, эманация, радиация, лучеиспускание, сноп, фонирование,

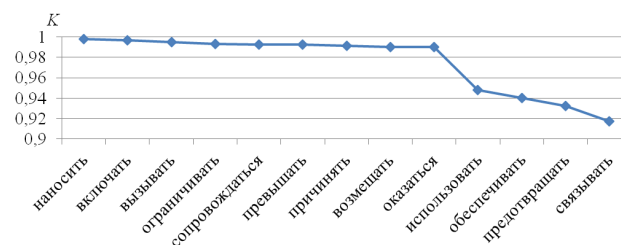


Рис. 3. Отбор предикатов (по вертикальной оси отложены значения обобщенного коэффициента взвешивания K)

Наконец, в соответствии с определением формального концепта, данным выше, лишь 4 кандидата были отобраны и 3 из них окончательно подтверждены экспертом, т. е. 75 %. Для понятия *излучение* в корпусе был найден только один синоним, и он также был подтвержден с помощью FCA и одобрен экспертом. Эти результаты представлены в табл. 3.

Таблица 3

Список подтвержденных синонимов для расширения базовой онтологии

Понятие базовой онтологии	Список синонимов
Ущерб	вред, повреждение, авария, осложнение
Излучение	радиация

Релевантность синонимов, подобранных для других концептов, составила от 50 до 100 %.

Аналогичный эксперимент был проделан также для параллельного французского корпуса с использованием электронного словаря синонимов CRISCO. В результате эксперимента оказалось, что, что качество полученных результатов для французской версии онтологии оказалось выше, что можно объяснить недостаточной полнотой русскоязычного ресурса и более сложной структурой предложений статей русскоязычного корпуса, что требует дополнительного, более тонкого анализа.

6. Выводы

В статье представлен метод полуавтоматического расширения онтологии, основанный на использовании специализированного текстового корпуса и внешнего лексического ресурса, такого как машиночитаемый словарь синонимов. Одним из важных преимуществ

метода является прямое использование установленного вручную списка базовых понятий предметной области.

Использование внешнего словаря позволяет обнаружить в корпусе различные лексические вхождения одного понятия с высокой точностью (лучшая полученная точность в нашем эксперименте составила 100 %).

Узким местом метода является его зависимость от полноты используемых ресурсов, корпуса и словаря. Было протестировано несколько русскоязычных тезаурусов. Тезаурус RussNet имеет подходящую структуру, однако его мощность недостаточна для того, чтобы соответствовать требованиям по полноте для выбран-

ной предметной области. Другой тезаурус – РуТез, концептуальный лингвистический ресурс, содержит около 158 тысяч слов и выражений, а также более 210 тысяч отношений [18]. Однако его структура не позволяет непосредственно применить его для предлагаемого алгоритма.

В среднем, все экспериментальные результаты показали достаточно хорошую точность: более 70 %. Этот факт позволяет продолжить эксперимент в том же направлении: в дальнейшем предполагается использовать словарь синонимов для создания классов глаголов, описывающих значимые свойства концептов, чтобы улучшить отклик (в нашем случае он достигает 50 %).

Литература

1. Schmid, H. Probabilistic Part-of-Speech tagging using decision trees [Text] / H. Schmid // In Proceedings of the International Conference on New Methods in Language Processing, Manchester, UK, 1994. – P. 44–49.
2. Buitelaar, P. Ontology Learning from An Overvie [Text] / P. Buitelaar, P. Cimiano, B. Magnini // Methods, Evaluation and Applications. Eds. IOS Press, 2008.
3. Maedche, A. Mining Ontologies from Text. [Text] / A. Maedche, S. Staab // In Proceedings of the 12th International Conference on Knowledge Engineering and Knowledge Management: EKAW'2000, French, 2000. – P. 189–202.
4. Cimiano, P. Learning concept hierarchies from text corpora using formal concept analysis [Text] / P. Cimiano, A. Hotho, S. Staab // In Journal of Artificial Intelligence Research. – 2005. Vol. 24.– P. 305–339.
5. Aussenac-Gilles, N. Revisiting Ontology Design: a methodology based on corpus analysis [Text] / N. Aussenac-Gilles, N. Biébow B. and S. Szulman // In Proceedings of the 12th International Conference on Knowledge Engineering and Knowledge Management: EKAW'2000, French, 2000. – P. 172–188.
6. Mondary, T. Construction d'ontologies à partir de textes : la phase de conceptualisation [Text] / T. Mondary, S. Després, A. Nazarenko, and S. Szulman // In Proceeding of International Conference IC 2008 'Ingénierie des Connaissances', Franch, 2008. – P. 87–98.
7. Missikoff, M. The Usable Ontology: An Environment for Building and Assessing a Domain Ontology [Text] / M. Missikoff, R. Navigli and P. Velardi // In Proceedings of the International Semantic Web Conference (ISWC). – USA, 2002. – P. 39–53. doi: 10.1162/COLI_a_00146
8. Velardi, P. OntoLearn Reloaded: A Graph-Based Algorithm for Taxonomy Induction [Text] / P. Velardi, F. Stefano, R. Navigli // In Computational Linguistics. – 2012. – Vol. 39, Issue 3. – P. 665–707. doi: 10.1162/COLI_a_00146
9. De Nicola, A. A software engineering approach to ontology building [Text] / A. De Nicola, M. Missikoff, R. Navigli // Inf. Syst. – 2009. – Vol. 34, Issue 2. – P. 258–275.
10. Morlane-Hondère, F. Étude des manifestations de la relation de méronymie dans une ressource distributionnelle [Text] / F. Morlane-Hondère, C. Fabre // In Proceeding of International Conference TALN'2012, France, 2012. – P. 169–182.
11. Bourigault, D. Pour une terminologie textuelle [Text] / D. Bourigault, M. Slodzian // Terminologies Nouvelles. – 1999. – Vol. 19. – P. 29–32.
12. Périne, A. Hybrid acquisition of semantic relations based on context normalization in distributional analysis [Text] / A. Périne, T. Hamon // In Proceeding of International Conference Terminology and Artificial Intelligence (TIA), France, 2013. – P. 113–121.
13. Huddleston, R. A Student's Introduction to English Grammar [Text] / R. Huddleston, G. Pullum. – Cambridge Univ. Press, 2005.
14. Ganter, B. Formal Concept Analysis: Mathematical Foundations [Text] / B. Ganter, R. Wille. – New York :Springer-Verlag, Inc., Secaucus, 1999
15. Orobinska, O. Enrichissement d'une ontologie de domaine par extension des relations taxonomiques à partir de corpus spécialisé [Text] / O. Orobinska, J-H. Chauchat, N. Charonova // In Proceeding of 10th International Conference In Proceeding of International Conference Terminology and Artificial Intelligence (TIA), France, 2013. – P. 129–137.
16. Apresjan, Ju. Analyse distributionnelle des significations et champs sémantiques structurés [Text] / Ju. Apresjan // Langages, 1e année, n°1, Recherches sémantiques, 1966. – P. 44–74.
17. Сайт словаря «Полный словарь синонимов русского языка» [Электронный ресурс] / Режим доступа: <http://словарь-синонимов.рф>
18. Лукашевич, Н. В. Тезаурусы в задачах информационного поиска [Текст] / Н. В. Лукашевич. – М. : Издательство Московского университета, 2011. – 512 с.