Разработка автоматизированных методов представления знаний о действиях и ситуациях

Е. О. Ненашева ¹, Д. Е. Пальчунов ^{1, 2}

¹ Новосибирский государственный университет Новосибирск, Россия
² Институт математики им. С. Л. Соболева СО РАН Новосибирск, Россия

Аннотация

Статья посвящена разработке автоматизированных методов интеграции знаний, извлеченных из текстов естественного языка. Для решения этой задачи используются методы преобразования предложений естественного языка во фрагменты атомарных диаграмм. Знания, извлеченные из текстов, формализуются при помощи атомарных предложений сигнатуры, состоящей из символов констант, двухместных предикатов и дополнительных констант-ситуаций. Разработаны методы интеграции знаний, содержащихся в нескольких предложениях естественного языка, позволяющие учитывать их семантические контексты.

Ключевые слова

извлечение знаний, представление знаний, интеграция знаний, анализ текстов естественного языка, теория «Смысл ⇔ текст», формализованное представление ситуаций, атомарная диаграмма

Благодарности

Исследование выполнено при частичной финансовой поддержке Президиума СО РАН (проект «Инженерия интенсиональных онтологий в дедуктивных и информационных системах» Комплексной программы ФНИ СО РАН II.1)

Для цитирования

Ненашева Е. О., Пальчунов Д. Е. Разработка автоматизированных методов представления знаний о действиях и ситуациях // Вестник НГУ. Серия: Информационные технологии. 2019. Т. 17, № 3. С. 61–72. DOI 10.25205/1818-7900-2019-17-3-61-72

Semi-Automated Methods for Representing Knowledge of Actions and Situations

E. O. Nenasheva ¹, D. E. Palchunov ^{1, 2}

 Novosibirsk State University Novosibirsk, Russian Federation
 Sobolev Institute of Mathematics SB RAS Novosibirsk, Russian Federation

Abstract

The article is devoted to the development of semi-automated methods for integrating knowledge extracted from natural language texts. To solve this problem, we use methods for converting natural language sentences into fragments of atomic diagram of algebraic systems. The knowledge extracted from texts is formalized with atomic sentences of signature consisting of constant symbols, binary predicates and additional situation-constants. We developed methods for integrating knowledge contained in several sentences of natural language, allowing to take their semantic contexts into account.

Keywords

knowledge extraction, representation of knowledge, knowledge integration, analysis of natural language texts, theory "Meaning \Leftrightarrow text", formalized representation of situations, atomic diagram

© Е. О. Ненашева, Д. Е. Пальчунов, 2019

Acknowledgements

The study was conducted with the partial financial support of the SB RAS Presidium (project "Engineering of intensional ontologies in deductive and information systems" of the Integrated program of fundamental research SB RAS II.1)

For citation

Nenasheva E. O., Palchunov D. E. Semi-Automated Methods for Representing Knowledge of Actions and Situations. *Vestnik NSU. Series: Information Technologies*, 2019, vol. 17, no. 3, p. 61–72. (in Russ.) DOI 10.25205/1818-7900-2019-17-3-61-72

Введение

Существует огромное количество текстовых документов, представленных в цифровом виде. С каждым днем их объем стремительно растет. Когда перед человеком встает задача обработки информации, заключенной в этих документах, справиться вручную уже практически невозможно. Таким образом, возникает потребность в программном обеспечении, способном обрабатывать тексты естественного языка, извлекать из них необходимую информацию и объединять полученные знания. При работе с документами, представленными на естественном языке, крайне важно учитывать смысл текста, определения и смысл входящих в него понятий, иначе говоря, семантику текста.

Проблема формального представления семантики текста исследовалась многими специалистами. Существенный вклад в решение этой проблемы внес И. А. Мельчук, разработав теорию «Смысл ⇔ Текст» [1; 2]. В рамках разработки этой теории был создан толковокомбинаторный словарь, позволяющий учитывать возможность слов вступать в семантические и синтаксические связи с остальными словами предложения. Кроме того, И. А. Мельчук предложил рассматривать глаголы в качестве многоместных предикатов.

Теория «Смысл ⇔ Текст» была применена в [3] при разработке теоретико-модельного подхода к извлечению знаний из текстов естественного языка. Подход был реализован в программной системе LogicText [4], позволяющей строить атомарные предложения логики предикатов по предложениям русского языка. Для построения фрагментов атомарных диаграмм моделей используются словари номинализаций и валентностей глаголов.

Дальнейшее развитие теоретико-модельного подхода предложено в [5], где описан подход к формализации знаний при помощи двухместных предикатов и констант-ситуаций, а также методы заполнения «пустых» валентностей предикатов-глаголов для выявления недостающих в тексте знаний и для пополнения этих знаний.

При анализе и обработке текстов естественного языка крайне важно учитывать контекст. Для корректного понимания семантики необходимо обрабатывать несколько предложений текста одновременно, поэтому целью данного исследования стала разработка методов интеграции знаний, заключенных в разных предложениях естественного языка. Особое внимание в данной работе уделяется выявлению связей между предложениями текста.

Преобразование текста естественного языка в атомарные предложения логики предикатов

Четырехуровневая модель представления знаний

При обработке, формализации и интеграции извлеченных из текста знаний мы используем четырехуровневую модель представления знаний, предложенную ранее [6; 7]. Данная модель включает в себя:

- 1) онтологию;
- 2) общие знания;
- 3) эмпирические (прецедентные) знания;
- 4) оценочные и вероятностные знания.

ISSN 1818-7900 (Print). ISSN 2410-0420 (Online) Вестник НГУ. Серия: Информационные технологии. 2019. Том 17, № 3 Vestnik NSU. Series: Information Technologies, 2019, vol. 17, no. 3 Онтология – это спецификация смысла, определения (полные или неполные) ключевых понятий, которыми описывается данная предметная область.

Общие знания — это утверждения, принципы и закономерности, являющиеся универсальными для всей предметной области. Они считаются истинными в каждом конкретном экземпляре, прецеденте данной предметной области (на текущий момент).

Эмпирические знания (иначе говоря, прецеденты) – это описания конкретных ситуаций, экземпляров предметной области.

Оценочные и вероятностные знания – это знания, истинность которых не является точной, это приближенные, нечеткие, гипотетические или субъективные знания.

В рамках четырехуровневой онтологической модели оценочные и вероятностные знания могут порождаться на основе онтологических, общих и эмпирических знаний о предметной области.

Одной из задач данного исследования является разработка методов формализации и интеграции знаний, соответствующих разным уровням описанной модели.

Представление знаний при помощи двухместных предикатов и констант-ситуаций

Введем некоторые определения. В работе рассматриваются модели вида

$$\mathfrak{A} = \langle A; \sigma \rangle = \langle A; P_1, ..., P_n, c_1, ..., c_l \rangle$$

сигнатуры $\sigma = \langle P_1, ..., P_n, c_1, ..., c_l \rangle$, где A — универсум модели $\mathfrak{A}, P_1, ..., P_n$ — предикатные символы, а $c_1, ..., c_l$ — константные символы. Множество предложений сигнатуры σ обозначается как $S(\sigma)$.

Предложение ф называется атомарным, если

$$\varphi = (c_1 = c_2)$$
 или $\varphi = P(c_1, ..., c_n)$, где $P, c_1, ..., c_n \epsilon \sigma$.

Множество предложений

атомарное предложение или отрицание атомарного предложения} мы называем атомарной диаграммой модели A.

Подмножество атомарной диаграммы $AD(\mathfrak{A})$ модели \mathfrak{A} мы называем фрагментом атомарной диаграммы модели \mathfrak{A} , конечное подмножество – конечным фрагментом. В данной работе рассматриваются только конечные фрагменты атомарных диаграмм. Сведения по теоретикомодельному подходу к извлечению и представлению знаний можно найти в [3; 5; 8].

И. А. Мельчук предложил рассматривать глаголы в качестве многоместных предикатов. Основанием этому послужила способность глаголов вступать в связи с другими словами предложения. Это свойство было названо валентностью. Подход И. А. Мельчука был применен в [3] для извлечения и формализации знаний, содержащихся в предложениях естественного языка. В [3] разрабатывался теоретико-модельный подход к извлечению и формализации знаний: они представляются в виде наборов атомарных предложений логики предикатов. Глаголы, причастия и деепричастия представляются в виде *п*-местных предикатов. Разработанные методы были реализованы в программе Logic Text [4]. Она осуществляет построение фрагментов атомарных диаграмм моделей по предложениям естественного языка, используя словари номинализаций и валентностей глаголов. Пример работы программы представлен на рис. 1.

Одной из главных задач данного исследования является формализация знаний, извлеченных из текстов естественного языка, в виде двухместных предикатов. В первую очередь, подобный подход решает проблемы изменения валентности предикатов. *п*-местные предикаты, предложенные в [3] имеют фиксированный набор аргументов и не всегда полностью соответствуют составу слов в предложении, особенно когда речь идет об однородных членах предложения. С помощью двухместных предикатов мы можем обрабатывать слова, отвечающие на один и тот же вопрос. Кроме того, выбор двухместных предикатов в качестве ба-

зовой конструкции связан с тем, что для дальнейшей обработки знаний целесообразно использовать технологии семантической паутины (Semantic Web) [9–13].

```
Вирусы меняют[асt, оbj, что, на что] поведение программ, внедряют[асt, оbj, что, куда] себя[оbj] в их[оbj] исполняемый[асt, оbj] код
внедрять(внедрять_0, вирус, вирус, код) исполнять (исполнять_1, код) менять (менять_0, вирус, поведение, на_что_менять_0) их(программа) себя(вирус)
```

 $Puc.\ 1$. Пример работы программы Logic Text $Fig.\ 1$. The example of working with the program Logic Text

Одним из основных инструментов Semantic Web является язык описания онтологий OWL [14]. Он, в свою очередь, основан на модели представления данных RDF. RDF-утверждения имеют вид триплетов «субъект – предикат – объект».

Двухместные предикаты подходят под структуру триплетов, а значит, могут транслироваться в RDF-утверждения. Благодаря этому мы можем в дальнейшем использовать технологии Semantic Web, в частности применять ризонеры (автоматические средства логического вывода) [13; 15] ¹ к наборам двухместных предикатов, построенных по текстам естественного языка. Это позволит нам выявлять противоречия, а также обрабатывать полученные знания и порождать новые.

Другим важным аспектом работы стало использование констант-ситуаций в качестве первого аргумента двухместного предиката. Как было сказано ранее, в данном исследовании используются различные уровни представления знаний. В текстах естественного языка объекты и сущности могут иметь различные смысл и свойства в зависимости от контекста и происходящей ситуации. Для корректной обработки текстов естественного языка необходимо учитывать, о каком объекте и о каких свойствах объекта в данном предложении идет речь. Таким образом, константы-ситуации могут быть применены для решения проблемы неполноты и неточности извлекаемой информации. Например, рассмотрим предложения «Вчера Вася купил хлеб за 30 рублей. Сегодня Вася купил хлеб за 35 рублей». Несмотря на внешнюю схожесть объектов (один и тот же глагол купил, объект хлеб), понятно, что эти предложения относятся к разным ситуациям, при этом Вася в обоих предложениях один, а хлеб разный. Такие моменты необходимо учитывать при обработке текстов естественного языка.

В данной работе мы продолжаем исследование проблемы формализации и интеграции знаний, извлеченных из текстов естественного языка, с применением двухместных предикатов и констант-ситуаций, начатое в [5].

Преобразование текста естественного языка в двухместные предикаты

Кратко изложим те аспекты подхода, предложенного в [5], которые потребуются нам в данной работе. Как уже было отмечено, в качестве базовой конструкции для построения модели знаний по тексту естественного языка нами выбраны двухместные предикаты и константы-ситуации. Для обработки текста на естественном языке и извлечения из него знаний на первом этапе производится преобразование текста с помощью программной системы LogicText. В результате мы получаем набор фрагментов атомарных диаграмм моделей, иначе говоря, набор многоместных предикатов, построенных по тексту.

ISSN 1818-7900 (Print). ISSN 2410-0420 (Online) Вестник НГУ. Серия: Информационные технологии. 2019. Том 17, № 3 Vestnik NSU. Series: Information Technologies, 2019, vol. 17, no. 3

¹ См. также: OWL 2 Web Ontology Language Document Overview (Second Edition). W3C OWL Working Group. URL: http://www.w3.org/TR/owl2-overview.

Например, предложение «Рецензент предоставляет в НГУ рецензию до 31 мая» программной системой LogicText будет преобразовано в предикат: *Предоставить* (предоставить_0, рецензент, НГУ, рецензия, до 31 мая). Предоставить_0 является константой-действием

Задача следующего этапа — преобразование n-местных предикатов в двухместные. В рамках Semantic Web известен подход, когда n-местный предикат преобразуется в набор из n двухместных предикатов [16]. Мы же осуществляем преобразование из каждого n-местного предиката в набор из n+1 двухместного предиката. Перед преобразованием необходимо привести полученные на первом этапе наборы атомарных диаграмм к специальному виду, а затем заменить вспомогательные константы-действия на константы-ситуации. Более подробно преобразование предикатов описано в [5]. В результате мы имеем набор из n+1двухместного предиката.

Так, приведенный выше многоместный предикат преобразуется в следующий набор:

Что сделать (s, предоставить)

Кто (s, рецензент)

Куда (s, НГУ)

Что (s, рецензия)

Когда (s, до 31 мая)

3десь s — новая константа-ситуация, введенная специально для формального представления данного предложения.

На третьем этапе необходимо определить связи между предложениями, точнее, между константами-ситуациями. Этот шаг является ключевым для понимания общего контекста. Важно определить, являются ли ситуации тождественными, т. е. идет ли в них речь об одном и том же действии. Кроме того, в одних случаях одна константа относится к одним и тем же объектам, а в других — к совершенно разным. Таким образом, на третьем этапе выявляются эквивалентные ситуации, а также тождественные объекты.

После прохождения всех трех этапов работы мы имеем набор двухместных предикатов, построенных по фрагменту текста естественного языка, а также знаем об эквивалентности некоторых констант-ситуаций. Теперь основная задача — объединить извлеченные знания. Рассмотрим данный этап подробнее.

Интеграция знаний, содержащихся в нескольких предложениях текста естественного языка

Предположим, у нас есть ситуации $s_1, ..., s_5$, которые являются эквивалентными. Эквивалентность ситуаций s_i и s_j задается с помощью отношения **тождественности** $Id(s_i, s_j)$. Для каждой из этих ситуаций описаны фрагменты атомарных диаграмм. В таком случае для интеграции знаний создается новая ситуация s_6 . К ней добавляются все знания, которые соответствуют ситуациям $s_1, ..., s_5$, при этом сами фрагменты диаграмм $s_1, ..., s_5$ не изменяются. Такой подход позволяет нам объединять знания и при этом сохранять исходные ситуации и данные для дальнейшего отслеживания изменений и обработки ризонерами [12; 17]. В результате мы получаем новую ситуацию s_6 , включающую в себя знания из нескольких предложений текста естественного языка.

Помимо отношения тождественности, между ситуациями может возникать связь **включения** $Include(S_i, S_j)$. Такое отношение мы используем, когда ситуация s_i является частью более общей ситуации s_i . В этом случае при объединении знаний идет уточнение информации.

Например, мы имеем два предложения: «Студент находится в новом корпусе НГУ» и «Новый корпус НГУ находится в Академгородке». Первое предложение имеет меньшую продолжительность по времени и общность, чем второе предложение. Значит, ситуация, связанная с первым предложением, включается в ситуацию, связанную со вторым предложением. В результате объединения знаний и логического вывода мы получим предложение «Студент находится в Академгородке».

Формально представить описанные выше связи между ситуациями можно следующим образом:

```
Аксиомы для связи Identity
Id(s_1, s_2): ситуация s_1 тождественна ситуации s_2.
Аксиомы отношения эквивалентности:
Id(s,s),
(Id(s_1,s_2) \rightarrow Id(s_2,s_1)),
((Id(s_1, s_2) \& Id(s_2, s_3)) \to Id(s_1, s_3)).
Аксиома конгруэнции:
((Id(s_1, s_2) \& P(s_1, x)) \to P(s_2, x)).
Аксиомы для связи Include
Include(s_1, s_2): s_1 \subseteq s_2, s_1 – частное, s_2 – общее.
Аксиомы частичного порядка:
Include(s,s),
(Include(s_1, s_2) \& Include(s_2, s_1)) \rightarrow Id(s_1, s_2),
(Include(s_1, s_2) \& Include(s_2, s_3)) \rightarrow Include(s_1, s_3).
Аксиома наследования:
\Big( \big( Include(s_1, s_2) \ \& \ P(s_1, x) \big) \ \to P(s_2, x) \Big).
```

В ходе исследований было выявлено, что введение констант-ситуаций, а также отношений тождественности и включения между ними недостаточно для полноценной и семантически верной интеграции знаний. Объекты могут менять свои свойства, в том числе и с течением времени. Нам важно учитывать, что какие-то события происходят одновременно, какие-то последовательно или параллельно. Во избежание противоречий нужно понимать, какое событие произошло раньше, а также какой «временной объем» оно имело (сколько длилось, с какими событиями пересеклось). Для каждого такого события необходимо описывать набор ситуаций. По этой причине нам необходимо ввести модель, в которую, помимо прочего, будут введены временные отношения между ситуациями.

Такая модель должна включать в себя: двухместные предикаты, первым аргументом которых является определенная ситуация, набор используемых ситуаций и отношения между ситуациями (например, «раньше», «позже», «часть по времени», «часть по протяженности» и т. д.)

Вероятно, для определения временных отношений между ситуациями потребуется обработка дат каким-либо внешним источником (например, чтобы выявить, какая ситуация произошла раньше). Для этого можно использовать оракулы – веб-сервисы, объекты из внешнего мира.

Отдельно стоит отметить, что при интеграции знаний мы рассматриваем множество тех ситуаций, которые относятся к определенному событию, обстоятельству. Объем (или протяженность) каждой ситуации включает протяженность рассматриваемого события.

Итак, для того чтобы учитывать связи между ситуациями относительно времени, введены дополнительные отношения.

Для определения связей между ситуациями, которые выполняются последовательно, мы используем двухместное отношение Pahbme (или before). $Before(s_1, s_2)$ — отношение между ситуациями, когда ситуация s_1 происходит раньше ситуации s_2 . Мы можем понять, что ситуации происходят последовательно, если в предложении четко указано время или присутствуют специальные слова, указывающие на это, например: neped (этим), do (этого), pahbme, guesarrappa и подобные. Обратный случай можно обозначить двухместным отношением guesarrappa и guesarrappa — отношение между ситуациями, когда ситуация guesarrappa — происходит позже ситуации guesarrappa .

Таким образом, $Before(s_1, s_2) \Leftrightarrow After(s_2, s_1)$.

Рассмотрим данную связь между ситуациями на примере. Возьмем предложения «Комплект документов собирается на факультете» и «После комплект документов хранится в деле Обучающегося». Обратим внимание, что сущности «комплект документов» и «дело Обучающегося» будут рассматриваться как целые константы, а не как отдельные слова. После выполнения первого шага описанного выше алгоритма и приведения многоместных предикатов к специальному виду получим следующее:

- 1) Собираться (s_1 , комплект документов, на факультете)
- 2) Xраниться (s_2) , комплект документов, в деле Обучающегося, после)

Этому соответствуют следующие наборы двухместных предикатов:

1) Что $(s_1, комплект документов)$

 $Что делает(s_1, coбирается)$

 $\Gamma \partial e \left(s_{_1} \right)$, на факультете)

2) Что (s_2 , комплект документов)

Что делает(s, хранится)

 Γ де (s_2 , в деле Обучающегося)

Когда $(s_2, nосле)$

Схематично это можно представить следующим образом (рис. 2).



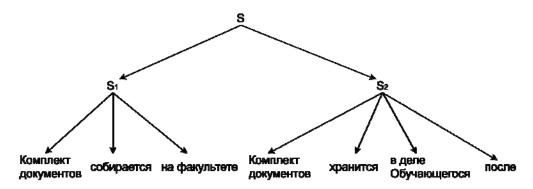
Puc. 2. Схема разбора предложения *Fig. 2.* The scheme of analysis of the sentence

Две полученных ситуации мы можем объединить согласно общему алгоритму интеграции. Его описание представлено ниже.

- 1. На первом шаге мы создаем общую ситуацию s, включающую в себя подситуации s_1 и s_2 (подситуаций может быть неограниченное количество). Другими словами, имеем отношения $Include(s_1,s)$ и $Include(s_2,s)$.
- 2. Проверяем тождественность ситуаций $Id(s_1, s_2)$ (для трех и более подситуаций делаем это последовательно, поскольку для отношения Id соблюдается транзитивность). Если удается выявить объекты, которые совпадают по смыслу, выносим их в общую ситуацию s (т. е. для этих объектов меняем текущую ситуацию на общую ситуацию).
- 3. Остальные объекты остаются в своих подситуациях, подситуации связываем соответствующим отношением.
- 4. Объединяем все полученные предикаты и отношения, строим по ним окончательное объединенное предложение.

Применим алгоритм к нашему примеру.

1. Создаем общую ситуацию s, включающую в себя ситуации s_1 и s_2 , получаем связи между ситуациями $Include(s_1,s)$ и $Include(s_2,s)$. Схематично это можно представить следующим образом (рис. 3).



Puc. 3. Добавление общей ситуации *Fig. 3.* The addition of the general situation

2. Проверяем тождественность s_1 и s_2 и выносим эквивалентные объекты в общую ситуацию s. В нашем случае такими объектами являются константы «комплект документов» и, соответственно, предикаты $\mathit{Чтo}(s_1, \mathit{комплект} \mathit{дoкументов})$ и $\mathit{Чтo}(s_2, \mathit{комплект} \mathit{дoкументов})$. Выносим тождественные объекты в общую ситуацию s путем замены первого аргумента в соответствующих двухместных предикатах, получаем предикат $\mathit{Чтo}(s, \mathit{комплект} \mathit{дoкументов})$. Остальные предикаты остаются прежними:

 $Что делает(s_1, coбирается)$

 $\Gamma \partial e \left(s_1 \right)$, на факультете)

 $Что делает(s_2, хранится)$

 Γ де (s_2 , в деле Обучающегося)

Когда $(s_2, nосле)$

Стоит отметить следующий момент: чтобы отнести тождественные объекты к общей ситуации, нужно их объединить в один. Если слова одинаковые, проблем при объединении не возникает. Но бывают случаи, когда встречаются слова, синонимичные, эквивалентные по смыслу или находящиеся в отношении «общее — частное». Например, если в одном предложении используется слово *студент*, а в другом — *бакалавр*. Сейчас для простоты будем приводить объекты к тому виду, в котором они встретились нам впервые. Если сначала мы рассмотрели предложение, где объектом был «студент», потом предложение, где объектом был «бакалавр», то, объединив эти слова, получим «студент» и вынесем его в общую ситуацию. В дальнейшем такие случаи нужно обрабатывать отдельно, возможно, в полуавтоматическом режиме (привлекая пользователя или эксперта).

Результат второго шага схематично представлен на рис. 4.

- 3. Далее необходимо связать подситуации s_1 и s_2 каким-либо отношением. Поскольку в предложении встречается ключевое слово *после*, будем связывать их временным отношением $After(s_2, s_1)$.
 - 4. Выписываем все полученные предикаты и отношения:

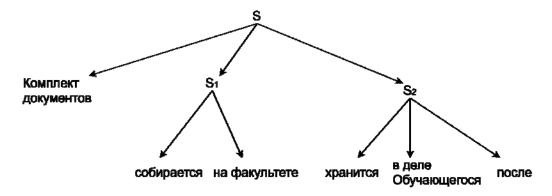
 $Include(s_1,s)$

 $Include(s_2, s)$

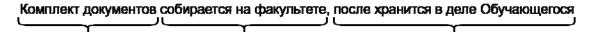
ISSN 1818-7900 (Print). ISSN 2410-0420 (Online) Вестник НГУ. Серия: Информационные технологии. 2019. Том 17, № 3 Vestnik NSU. Series: Information Technologies, 2019, vol. 17, no. 3

```
Что(s, комплект документов) Что делает(s_1, cобирается) Где (s_1, на факультете) Что делает(s_2, xранится) Где (s_2, в деле Обучающегося) Когда(s_2, после) Аfter(s_2, s_1).
```

Для составления конечного предложения сначала выписываются общие объекты (из общей ситуации), потом к ним добавляются объекты из подситуаций. Поскольку подситуации связаны временным отношением, выписываем их в соответствующем порядке (рис. 5).



Puc. 4. Вынесение эквивалентных объектов в общую ситуацию *Fig. 4.* Moving of the identical objects in the general situation



Puc. 5. Полученное предложение *Fig. 5.* The resulting sentence

Мы рассмотрели пример интеграции знаний для ситуаций, которые происходят последовательно.

Для определения связей между ситуациями, которые происходят одновременно, предлагается использовать двухместное отношение *Одновременно* (или *SameTime*). В данном случае ключевыми словами могут быть «в то же время», «одновременно» и подобные. Если в предложениях явно не указано время действия или отсутствуют ключевые слова, применяется принцип пресуппозиции. Другими словами, сначала выявляется пресуппозиция, что две ситуации s_1 и s_2 связаны отношением $sameTime(s_1,s_2)$, затем задается вопрос пользователю о верности данной пресуппозиции. Если возможности общения с пользователем нет, формируется условное утверждение: фрагмент атомарной диаграммы верен, если верна данная пресуппозиция. В противном случае формируется другой фрагмент атомарной диаграммы, поскольку нам важно рассмотреть все возможные варианты интеграции знаний.

Для ситуаций, связанных отношением *SameTime*, объединение знаний происходит по описанному выше алгоритму. Так, из предложений «Для поступления поступающий подает заявление о приеме» и «Для поступления поступающий подает документ, удостоверяющий об-

разование соответствующего уровня» мы получим «Для поступления поступающий подает заявление о приеме и документ, удостоверяющий образование соответствующего уровня».

Добавление моментов времени во фрагменты атомарных диаграмм

Некоторые ситуации могут иметь условия, меняющиеся во времени, т. е. в момент времени t_1 утверждение является истинным, а в момент времени t_2 , отличный от t_1 , это же утверждение ложно. Поэтому кроме временных отношений между ситуациями следует вводить аргумент t – время. Если в утверждение добавлять момент времени, то оно становится «абсолютно» истинным или ложным, т. е. не зависящим от ситуации и времени.

В рамках данного подхода предложена конструкция, позволяющая добавить время в многоместный предикат и преобразовать его в набор двухместных.

Сначала многоместные предикаты вида $P(t,c_1,...,c_n)$, как это описано выше, преобразуются в набор атомарных предложений $\{R(s,P),\ Q_0(s,t),Q_1(s,c_1),...,Q_n(s,c_n)\}$, где s — соответствующая константа-ситуация. Выше были рассмотрены примеры предикатов $Q_i(s,c_i)$: $Kmo(s,c_i),\ \Gamma\partial e(s,c_i),\ Kor\partial a(s,c_i)$. Эти двуместные предикаты $(Kmo,\ \Gamma\partial e,\ Kor\partial a)$ — типы аргументов предиката $P(t,c_1,...,c_n)$, они соответствуют вопросам к действию (ситуации), задаваемым предикатом $P(t,c_1,...,c_n)$.

Таким образом, при преобразовании многоместного предиката в набор двухместных происходит:

- добавление символов предикатов в качестве новых элементов и
- добавление констант *s* в качестве новых элементов.

Для того чтобы не добавлять имена предикатов в качестве новых элементов модели, мы рассматриваем также следующий способ преобразования многоместных предикатов в двухместные. Многоместный предикат вида $P(t,c_1,...,c_n)$ преобразуется в набор атомарных предложений $\{Q_0^P(s,t),Q_1^P(s,c_1),...,Q_n^P(s,c_n)\}$. Пример предиката $Q_i^P(s,c_i)$: $Kynun_F\partial e(s,c_i)$. При таком способе преобразования многоместных предикатов в двухместные, происходит добавление только констант s в качестве новых элементов модели (и, соответственно, новых сигнатурных символов). Символы предикатов в качестве новых элементов модели не добавляются. В элементарную теорию полученной модели добавляются универсальные предложения: $\forall x_1 \forall x_n (P(t,x_1,...,x_n) \rightarrow (Q_0^P(s,t) \& Q_1^P(s,x_1) \& \& Q_n^P(s,x_n)))$.

Заключение

В работе предложен подход к формализации и интеграции знаний, извлеченных из текстов естественного языка. В его основу легли теория «Смысл \Leftrightarrow Текст» И. А. Мельчука, а также теоретико-модельный подход к извлечению и представлению знаний, разработанный авторами ранее.

Разработаны методы интеграции знаний, содержащихся в нескольких предложениях естественного языка, основанные на использовании двухместных предикатов и констант-ситуаций. Предложенные методы могут быть применены для решения проблемы неполноты и неточности знаний. Также они дают возможность интегрировать как знания об объектах с неизменными свойствами, так и знания об изменяющихся во времени свойствах объектов.

В дальнейшем предложенный подход может быть дополнен использованием технологий Semantic Web, в том числе автоматических средств логического вывода, для выявления противоречий в документах естественного языка и для порождения новых знаний.

Список литературы

- 1. **Мельчук И. А.** Об одной лингвистической модели типа «Смысл Текст»: уровни представления языковых высказываний // Серия литературы и языка. 1976. Т. 33, вып. 5. С. 5–33.
- 2. **Мельчук И. А.** Опыт теории лингвистических моделей «Смысл ⇔ Текст». М., 1999.

- 3. **Махасоева О. Г., Пальчунов** Д. Е. Автоматизированные методы построения атомарной диаграммы модели по тексту естественного языка // Вестник НГУ. Серия: Информационные технологии. 2014. Т. 12, № 2. С. 64–73.
- 4. **Махасоева О. Г., Пальчунов Д. Е.** Программная система построения атомарной диаграммы модели по тексту естественного языка. Св-во о гос. регистрации программы для ЭВМ № 2014619198, зарегистрировано 10.09.2014.
- 5. **Ненашева Е. О., Пальчунов** Д. **Е.** Разработка автоматизированных методов преобразования предложений естественного языка в бескванторные формулы логики предикатов // Вестник НГУ. Серия: Информационные технологии. 2017. Т. 15, № 3. С. 49–63.
- 6. **Найданов Ч. А., Пальчунов Д. Е., Сазонова П. А.** Теоретико-модельные методы интеграции знаний, извлеченных из медицинских документов // Вестник НГУ. Серия: Информационные технологии. 2015. Т. 13, № 3. С. 29–41.
- Naydanov Ch., Palchunov D., Sazonova P. Development of automated methods for the prevention of risks of critical conditions, based on the analysis of the knowledge extracted from the medical histories // Сибирский научный медицинский журнал. 2016. Т. 36, вып. 1. С. 105–113.
- 8. **Пальчунов** Д. Е. Моделирование мышления и формализация рефлексии. Ч. 2: Онтологии и формализация понятий // Философия науки. 2008. № 2 (37). С. 62–99.
- 9. **Palchunov D. E.** Modeling of reasoning and formalization of reflection. I: Model theoretical formalization of ontology and reflection. *Philosophy of Science*, 2006, no. 4 (31), p. 86–114.
- 10. **Allemang D., Hendler J.** Semantic Web for the Working Ontologist. Morgan Kaufmann, 2008, 352 p.
- 11. **Parreiras F. S.** Semantic Web and Model-Driven Engineering. Wiley-IEEE Press, 2012, 264 p.
- 12. **Капустина А. И., Пальчунов** Д. **Е.** Разработка онтологической модели тарифов и услуг сотовой связи, основанной на логически полных определениях понятий // Вестник НГУ. Серия: Информационные технологии. 2017. Т. 15, № 2. С. 34–46.
- 13. **Корсун И. А., Пальчунов** Д. Е. Разработка интеллектуальной системы обработки и интеграции знаний на основе технологий семантической паутины // Вестник НГУ. Серия: Информационные технологии. 2018. Т. 16, № 3. С. 113–125.
- 14. **Szeredi P., Lukácsy G., Benkő T.** The Semantic Web Explained: The Technology and Mathematics behind Web 3.0. Cambridge University Press, 2014, 478 p.
- 15. **Meilicke C., Stuckenschmidt H.** A Reasoning-Based Support Tool for Ontology Mapping Evaluation. University of Mannheim, 2008.
- 16. **Gutiererz F., Dou D., Fickas S., Griffiths G.** Online Reasoning for Ontology-Based Error Detection in Text. Computer and Information Science Department University of Oregon, 2014.
- 17. **Noy N., Rector A.** Defining N-ary Relations on the Semantic Web. In: W3C Working Group Note (12 April 2006). URL: https://www.w3.org/TR/swbp-n-aryRelations/.

References

- 1. Melchuk I. A. About one linguistic model of the «Meaning ⇔ Text» type: levels of utterance representation. *Series of literature and language*, 1976, vol. 33, no. 5, p. 5–33. (in Russ.)
- 2. **Melchuk I. A.** Experience of the theory of the linguistic models "Meaning ⇔ Text". Moscow, 1999. (in Russ.)
- 3. **Makhasoeva O. G., Palchunov D. E.** Semi-automatic methods of a construction of the atomic diagrams from natural language texts. *Vestnik NSU. Series: Information Technologies*, 2014, vol. 12, no. 2, p. 64–73. (in Russ.)
- 4. **Makhasoeva O. G., Palchunov D. E.** Program system for the construction of the atomic diagram of a model from natural language texts. Certificate of the State Registration of the Computer Program No. 2014619198, registered 10.09.2014.
- 5. **Nenasheva E. O., Palchunov D. E.** Semi-automated methods of transforming sentences from natural language into quantifier-free formulas of predicate logic. *Vestnik NSU. Series: Information Technologies*, 2017, vol. 15, no. 3, p. 49–63. (in Russ.)

- 6. Naydanov Ch. A., Palchunov D. E., Sazonova P. A. Model-theoretic methods of integration of knowledge extracted from medical documents. *Vestnik NSU. Series: Information Technologies*, 2015, vol. 13, no. 3, p. 29–41. (in Russ.)
- 7. **Naydanov Ch., Palchunov D., Sazonova P.** Development of automated methods for the prevention of risks of critical conditions, based on the analysis of the knowledge extracted from the medical histories. *The Siberian Scientific Medical Journal*, 2016, vol. 36, iss. 1, p. 105–113.
- 8. **Palchunov D. E.** Modeling of reasoning and formalization of reflection. Part 2: Ontologies and formalization of concepts. *Philosophy of Science*, 2008, no. 2 (37), p. 62–99. (in Russ.)
- 9. **Palchunov D. E.** Modeling of reasoning and formalization of reflection. I: Model theoretical formalization of ontology and reflection. *Philosophy of Science*, 2006, no. 4 (31), p. 86–114.
- 10. **Allemang D., Hendler J.** Semantic Web for the Working Ontologist. Morgan Kaufmann, 2008, 352 p.
- 11. **Parreiras F. S.** Semantic Web and Model-Driven Engineering. Wiley-IEEE Press, 2012, 264 p.
- 12. **Kapustina A. I., Palchunov D. E.** The development of ontological model of tariffs and services of mobile operator, based on logically complete definitions of concepts. *Vestnik NSU. Series: Information Technologies*, 2017, vol. 15, no. 2, p. 34–46. (in Russ.)
- 13. **Korsun I. A., Palchunov D. E.** An intellectual system for processing and integrating knowledge based on Semantic Web technologies. *Vestnik NSU. Series: Information Technologies*, 2018, vol. 16, no. 3, p. 113–125. (in Russ.)
- 14. **Szeredi P., Lukácsy G., Benkő T.** The Semantic Web Explained: The Technology and Mathematics behind Web 3.0. Cambridge University Press, 2014, 478 p.
- 15. **Meilicke C., Stuckenschmidt H.** A Reasoning-Based Support Tool for Ontology Mapping Evaluation. University of Mannheim, 2008.
- 16. **Gutiererz F., Dou D., Fickas S., Griffiths G.** Online Reasoning for Ontology-Based Error Detection in Text. Computer and Information Science Department University of Oregon, 2014.
- 17. **Noy N., Rector A.** Defining N-ary Relations on the Semantic Web. In: W3C Working Group Note (12 April 2006). URL: https://www.w3.org/TR/swbp-n-aryRelations/.

Материал поступил в редколлегию Received 24.06.2019

Сведения об авторах / Information about the Authors

- **Ненашева Евгения Олеговна**, студент, 2 курс магистратуры, факультет информационных технологий, Новосибирский государственный университет (ул. Пирогова, 2, Новосибирск, 630090, Россия)
- **Evgeniya O. Nenasheva**, Student, 2 year master course, Department of Information Technology, Novosibirsk State University (2 Pirogov Str., Novosibirsk, 630090, Russian Federation) nenasheva.zhenya@gmail.com
- Пальчунов Дмитрий Евгеньевич, доктор физико-математических наук, заведующий кафедрой ОИ ФИТ, Новосибирский государственный университет (ул. Пирогова, 2, Новосибирск, 630090, Россия); ведущий научный сотрудник, Институт математики СО РАН (пр. Академика Коптюга, 4, Новосибирск, 630090, Россия)
- **Dmitry E. Palchunov**, Doctor of Physical and Mathematical Sciences, Head of the Department of General Informatics of the Faculty of Information Technologies, Novosibirsk State University (2 Pirogov Str., Novosibirsk, 630090, Russian Federation); Leading Researcher, Sobolev Institute of Mathematics SB RAS (4 Academician Koptyug Ave., Novosibirsk, 630090, Russian Federation)

palch@math.nsc.ru