

RU

Терминологическая фреймово-сетевая модель концептуально-семантической организации понятий компьютерной лингвистики

Прохорова О. Н., Польщикова О. Н.

Аннотация. Объект исследования – концептуально-семантическая организация терминологического аппарата, предмет – логико-понятийная структура понятий, обозначаемых терминами компьютерной лингвистики. Цель исследования – выявление особенностей системной организации понятий компьютерной лингвистики на основе применения терминологической фреймово-сетевой модели. В статье осуществлен краткий анализ применения лингвистами фреймового подхода для моделирования структуры терминологических систем. Предложена терминологическая фреймово-сетевая модель предметной области «Компьютерная лингвистика». В качестве примеров в виде схем представлены фрагменты терминологической сети фреймов предметной области «Синонимия терминов компьютерной лингвистики». Научная новизна работы состоит в том, что дано оригинальное определение терминологической сети фреймов и впервые сформулированы правила ее построения. Полученные результаты исследования показали возможность применения терминологической фреймово-сетевой модели для анализа логико-понятийной структуры предметной области «Компьютерная лингвистика» и организации семантических терминообразовательных процессов и лексико-смысловых отношений в изучаемой терминологии.

EN

A terminological frame-network model of the conceptual and semantic organization of notions of computational linguistics

Prokhorova O. N., Polshchikova O. N.

Abstract. The research object is the conceptual and semantic organization of the terminological apparatus, the research subject is the logical and conceptual structure of notions denoted by terms of computational linguistics. The study aims to identify the features of the systemic organization of notions of computational linguistics based on the use of a terminological frame-network model. The paper briefly analyzes the use of the frame approach to modeling the structure of terminological systems by linguists. A terminological frame-network model of the subject area “Computational linguistics” is proposed. The study is illustrated with fragments of the terminological network of frames of the subject area “Synonymy of terms of computational linguistics” presented in the form of diagrams. The work is novel in that it provides an original definition of a terminological network of frames and is the first to formulate the rules for its construction. The research findings showed the possibility of using a terminological frame-network model to analyze the logical and conceptual structure of the subject area “Computational linguistics” and the organization of semantic term-formation processes and lexicosemantic relations in the terminology under study.

Введение

В лингвистических исследованиях широко представлены аспекты формирования и функционирования различных терминологий. В то же время некоторые терминосистемы до сих пор недостаточно изучены, требуют основательной проработки и систематизации для дальнейшего упорядочения. Одним из таких сравнительно молодых пластов специальной лексики является терминология компьютерной лингвистики, которая интенсивно развивается благодаря широкому практическому распространению информационных технологий обработки естественного языка и проникновению их в различные общественные сферы (Агузумцян,

Великанова, Польщиков и др., 2021; Польщиков, Лазарев, Константинов и др., 2020; Польщиков, Польщикова, Игитян и др., 2019; Khurana, Koli, Khatter et al., 2023; Velikanova, Polshchikov, Likhosherstov et al., 2021).

Компьютерная лингвистика динамично совершенствуется и является сегодня одной из важнейших предметных областей, изучение которой открывает эффективный технологический инструментарий для решения многих прикладных задач аналитического характера. Для получения и дополнения знаний в этой сфере необходима разработка упорядоченных структур (баз данных), отражающих системные взаимосвязи между специальными понятиями, обозначаемыми с помощью соответствующих терминов.

Актуальность исследования обусловлена потребностью выявления системной организации терминов компьютерной лингвистики на основе глубокого научного анализа концептуально-семантических связей, установившихся между понятиями соответствующей профессиональной сферы. Адекватное логико-понятийное структурирование специальных названий компьютерной лингвистики будет способствовать упорядочению терминологии данной предметной области, обеспечению ее всестороннего изучения. Систематизация терминов компьютерной лингвистики поможет структурировать систему знаний, относящихся к изучаемой предметной области. Разработка, упорядочение и корректировка терминологической базы позволит создать условия для профессиональной коммуникации.

Для достижения поставленной цели исследования в статье решены следующие задачи: 1) определить понятие терминологической фреймовой сети; 2) обозначить основные правила ее построения; 3) проиллюстрировать применение терминологической фреймово-сетевой модели для анализа логико-понятийной структуры предметной области «Компьютерная лингвистика» и организации семантических терминообразовательных процессов и лексико-смысловых отношений в исследуемой терминологии.

Исследование концептуально-семантической организации терминологии компьютерной лингвистики выполнено на основе применения фреймового моделирования для представления иерархической упорядоченности связей между понятиями, обозначаемыми с помощью терминов анализируемой предметной области. Для изучения семантических особенностей специальных названий был использован метод дефинитивного анализа фактического материала.

Материалом исследования послужили 518 терминов, отобранные методом направленной выборки из текстов по тематике компьютерной лингвистики и автоматической обработки естественного языка (Большакова Е. И., Воронцов К. В., Ефремова Н. Э., Клышинский Э. С., Лукашевич Н. В., Сапин А. С. Автоматическая обработка текстов на естественном языке и анализ данных. М.: Изд-во НИУ ВШЭ, 2017; Хобсон Л., Ханнес Х., Коул Х. Обработка естественного языка в действии. СПб.: Питер, 2020), 16 сборников докладов ежегодной международной конференции по компьютерной лингвистике и информационным технологиям «Диалог» (<https://www.dialog-21.ru/>), Русско-английского тезауруса по компьютерной лингвистике (РАТКЛ. <https://uniserv.iis.nsk.su/thes/index.php?ent=1>). Отбор терминов производился на основе анализа их дефиниций.

Теоретической базой послужили труды, посвященные изучению теории фреймов (Минский, 1979), использованию фреймовых шаблонов при построении систем автоматической обработки естественного языка (Луканин, 2011), применению фреймового моделирования для систематизации терминологического аппарата (Евсюкова, Карнась, 2021; Ли Ян, 2020; Туманян, 2022).

Практическая значимость исследования состоит в возможности применения его материалов и результатов при разработке компьютерных терминологических тезаурусов и баз данных. Полученные результаты рекомендуются также использовать в процессе преподавания терминоведческих дисциплин, лексикологии, учебных курсов по профилю компьютерной лингвистики.

Обсуждение и результаты

Терминология компьютерной лингвистики представляет собой сравнительно молодой и интенсивно развивающийся пласт специальной лексики. Уникальные возможности для исследования семантических связей, имеющихся между понятиями профессиональной сферы, предоставляет теория фреймов, предложенная Марвином Минским (1979). В когнитологической науке известна гипотеза, состоящая в том, что мышление человека оперирует фреймовыми структурами организации знаний (Луканин, 2011, с. 9). Фреймовый подход используется лингвистами для анализа терминологических систем (Евсюкова, Карнась, 2021; Ли Ян, 2020; Туманян, 2022). При этом для описания структуры семантических связей между понятиями применяются не только фреймы и наполняемые их слоты, но и дополнительные элементы, например субфреймы, субслоты и даже микрослоты (Трофимова, Щитова, 2022). Данные элементы используются как вспомогательные средства детализации отношений внутри терминологических массивов, однако строгому определению их сущности и функционала в литературных источниках уделяется недостаточно внимания.

На наш взгляд, для моделирования логико-понятийной структуры предметной области «Компьютерная лингвистика» целесообразно использовать терминологическую фреймовую сеть. Под терминологической сетью фреймов предметной области будем понимать модель концептуально-семантической организации понятий предметной области, представленную в виде иерархически связанных фреймов-терминов. Построение терминологической фреймово-сетевой модели предлагается осуществлять на основе указанных ниже правил:

1. Каждый фрейм включает следующие элементы:

- *имя фрейма*, обозначающее определенное понятие предметной области, представленное с помощью соответствующего термина;
- *слоты*, обозначающие признаки этого понятия.

2. Каждый слот включает:

- *имя слота*, обозначающее некоторый признак понятия предметной области;
- набор *значений*, именуемых относящихся к этому признаку понятия и представленных с помощью соответствующих терминов.

3. Каждый термин, с помощью которого именуется значение слота, выносится на нижестоящий уровень модели в виде фрейма с соответствующим именем.

Другими словами, фрейм представляет собой некоторую структуру (рамку), наполняемую множеством слотов (ячеек), элементами которых выступает множество значений, которые, в свою очередь, являются фреймами нижестоящего уровня, что показано на Рисунке 1.

Фрейм вышестоящего уровня – это, по сути, сетевая иерархическая структура, связывающая между собой фреймы нижестоящих уровней. Терминологическая фреймово-сетевая модель всей предметной области – это фрейм, который расположен в модели на самом верхнем уровне, именуется термином, обозначающим эту предметную область, и представляемый в виде многоуровневой сети иерархически связанных между собой фреймов, которые, в свою очередь, именуются с помощью терминов, обозначающих понятия этой предметной области.

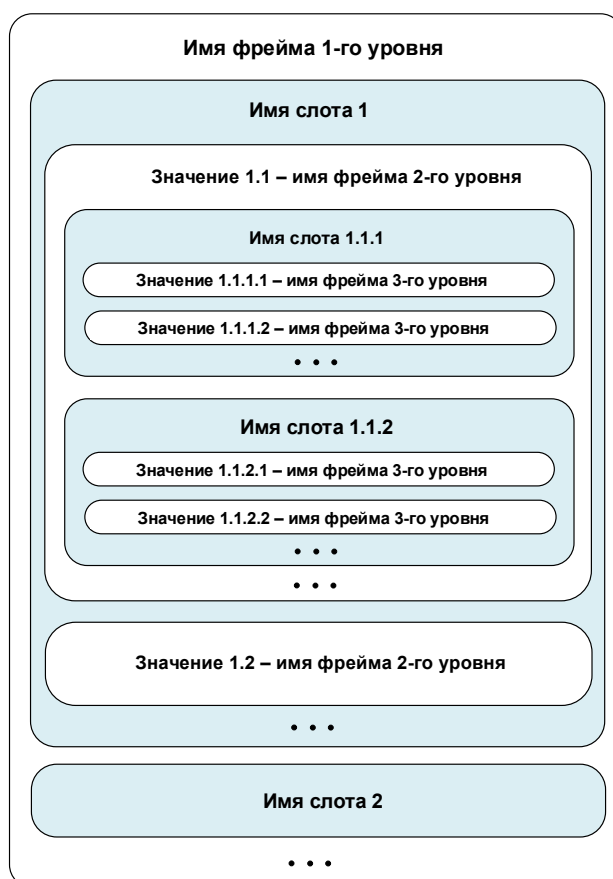


Рисунок 1. Элементы фреймовой модели

На Рисунке 2 представлен фрагмент терминологической фреймово-сетевой модели предметной области «Компьютерная лингвистика».

На верхнем (первом) уровне модели располагается фрейм «Компьютерная лингвистика». Слот этого фрейма, именуемый «Направления компьютерной лингвистики», содержит термины-значения (Большакова, Воронцов, Ефремова и др., 2017; Хобсон, Ханнес, Коул, 2020):

1. «Машинный перевод».
2. «Электронные словари, тезаурусы и онтологии».
3. «Распознавание символов».
4. «Обработка и анализ текстов».
5. «Распознавание и синтез речи».
6. «Диалоговые системы».
7. «Корпусная лингвистика».
8. «Компьютерная количественная лингвистика».
9. «Информационный поиск».
10. «Анализ мнений (анализ тональности)».

Значение «Машинный перевод» в рассматриваемой сети представляет собой отдельный фрейм второго уровня (Рисунок 3), который содержит слоты «Виды машинного перевода, различаемые по признаку участия человека» и «Технологические разновидности машинного перевода».

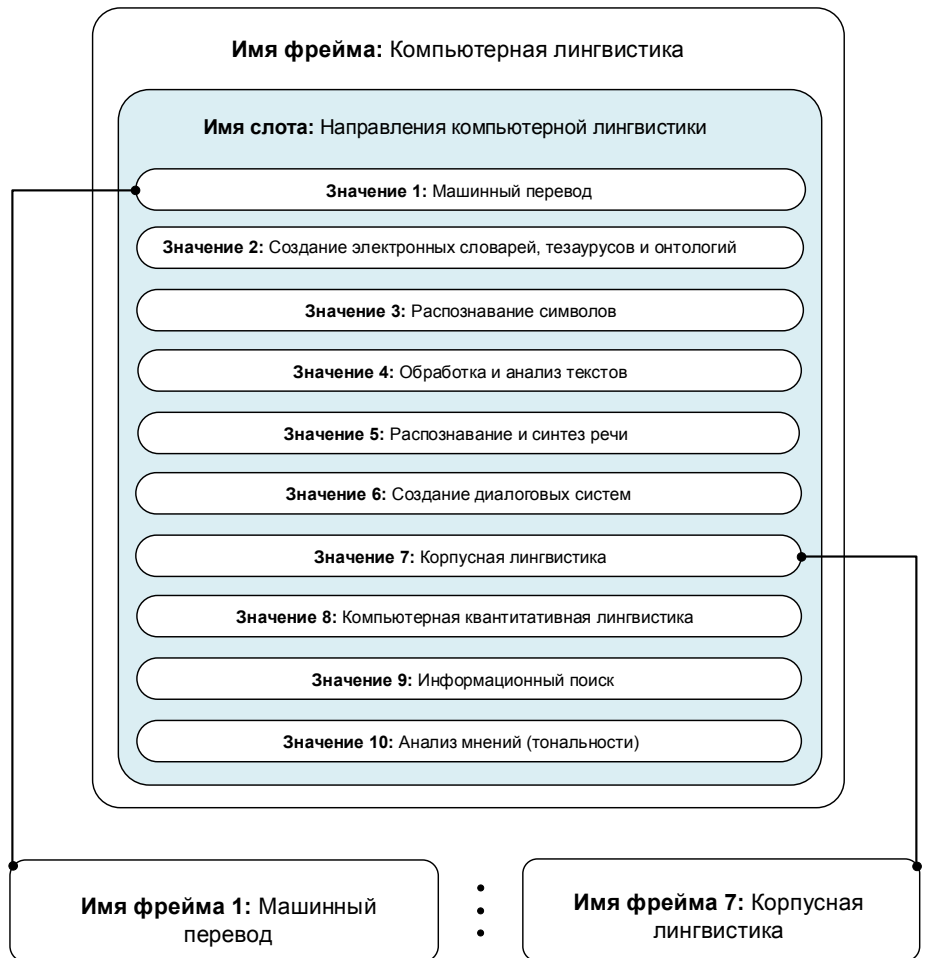


Рисунок 2. Фрагмент терминологической фреймово-сетевой модели предметной области «Компьютерная лингвистика»

Слот, именуемый «Виды машинного перевода, различаемые по признаку участия человека», имеет значения, именуемые следующими терминами:

1. «Автоматический машинный перевод».
2. «Автоматизированный машинный перевод».

Слот с именем «Технологические разновидности машинного перевода» включает термины-значения:

1. «Машинный перевод, основанный на правилах».
2. «Машинный перевод, основанный на примерах».
3. «Статистический машинный перевод».
4. «Гибридный машинный перевод».
5. «Нейросетевой машинный перевод».

Каждое из вышеуказанных значений образует соответствующий фрейм нижестоящего уровня. В качестве примера можно привести значение, именуемое «Статистический машинный перевод», которое в предлагаемой сети выносится в виде одноименного фрейма на третий уровень модели.

Фрейм «Статистический машинный перевод» содержит слот «Модели статистического машинного перевода», который имеет значения, обозначаемые следующими терминами:

1. «Машинный перевод по словам».
2. «Машинный перевод по фразам».
3. «Машинный перевод по синтаксису».

Системная упорядоченность такого представления обеспечивается благодаря реализации важнейшей особенности фреймов, состоящей в наследовании их свойств. Эта особенность проявляется, в частности, в дефинициях терминов компьютерной лингвистики. Так, например, фрейм второго уровня рассматриваемой модели именуется термином *корпусная лингвистика* (Рисунок 2), определяемым как направление компьютерной лингвистики, изучающее разработку общих принципов построения и использования лингвистических корпусов с применением компьютерных технологий (Захаров, 2005, с. 3). Представленная выше дефиниция содержит имя фрейма вышестоящего (первого) уровня, что указывает на наследование понятием «корпусная лингвистика» ключевой особенности понятия «компьютерная лингвистика», состоящей в использовании компьютерных информационных технологий для исследования естественного языка.

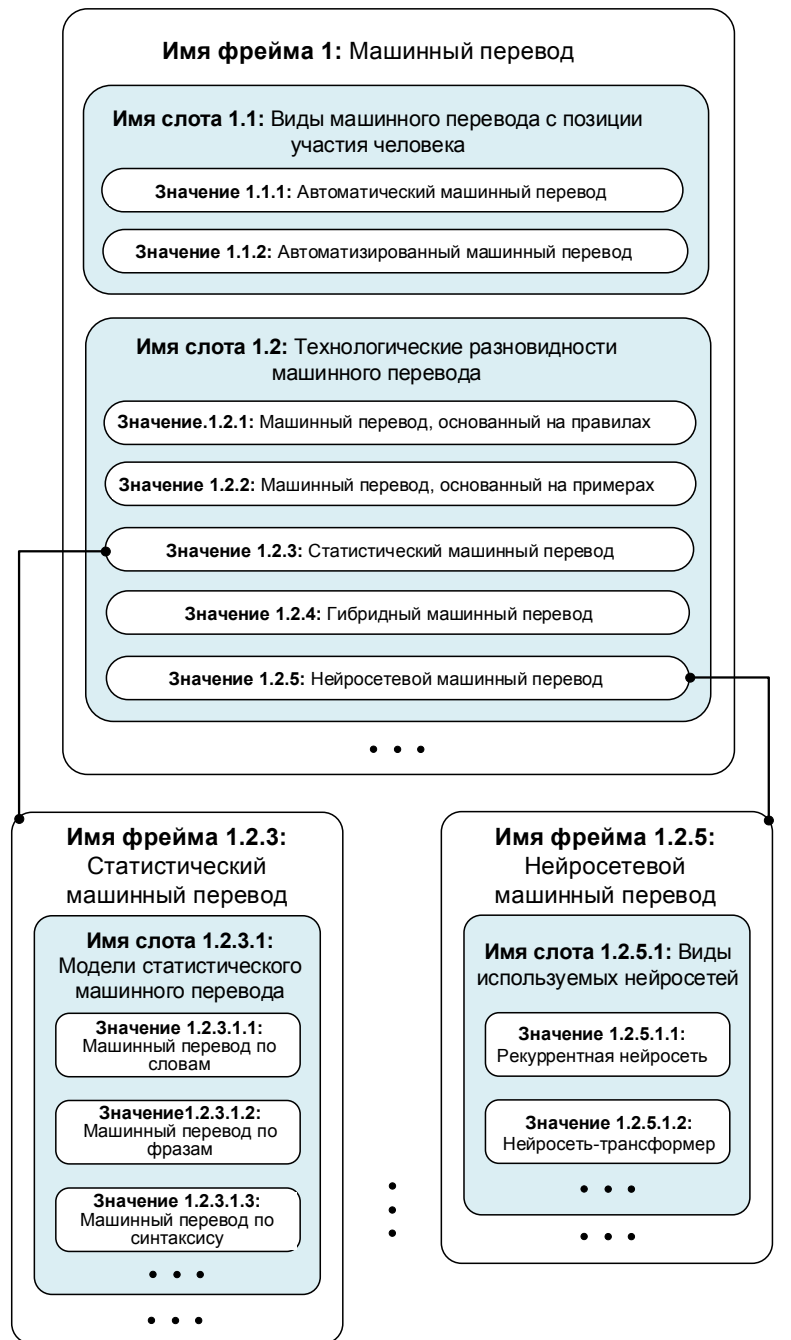


Рисунок 3. *Фрагмент фреймово-сетевой модели, ассоциированный с термином «машинный перевод»*

Предложенная фреймовая модель может применяться не только для анализа концептуально-семантической организации понятий предметной области «Компьютерная лингвистика», но и для систематизации семантических терминообразовательных процессов и лексико-смысловых отношений в исследуемой терминологии.

На Рисунке 4 представлен фрагмент терминологической сети фреймов предметной области «Синонимия терминов компьютерной лингвистики». Фрейм верхнего уровня модели именуется «Термины-синонимы компьютерной лингвистики».

В виде примера показаны два слота этого фрейма, именуемые «Типы терминов-синонимов, различаемые по близости значений» и «Типы терминов-синонимов, различаемые по структурному признаку». Первый слот содержит три значения, которые являются лингвистическими терминами:

1. «Абсолютные термины-синонимы».
2. «Относительные термины-синонимы».
3. «Комплексные термины-синонимы».

Значение «Абсолютные термины-синонимы» в рассматриваемой сети может быть представлено в виде отдельного фрейма второго уровня (Рисунок 5), который содержит слот «Разновидности абсолютных терминов-синонимов, различаемые по морфологической природе».



Рисунок 4. Фрагмент терминологической сети фреймов предметной области «Синонимия терминов компьютерной лингвистики»

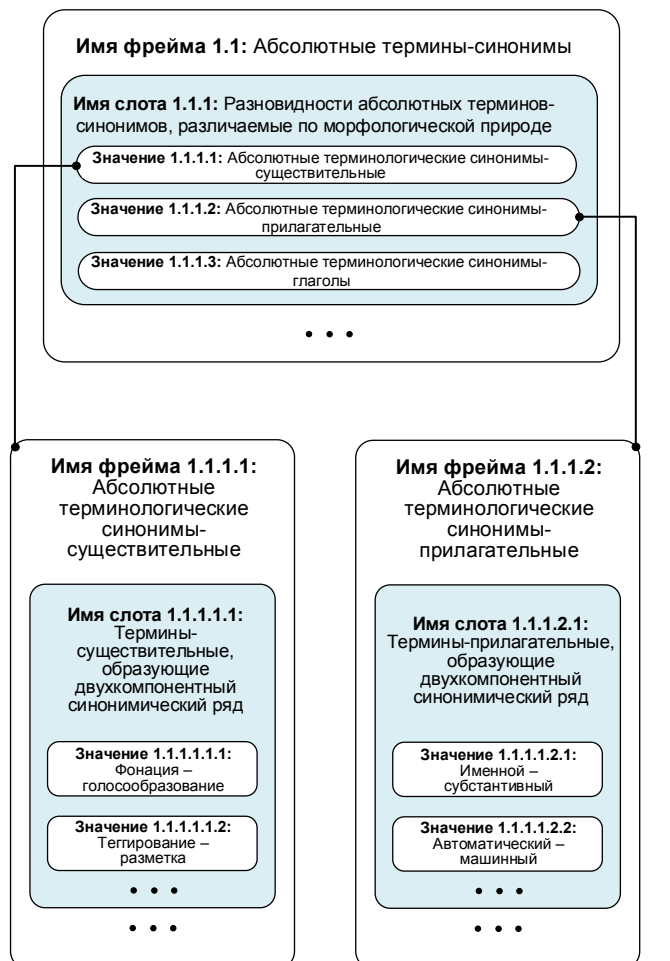


Рисунок 5. Фрагмент сети фреймов, соответствующий понятию «абсолютные термины-синонимы»

Слот «Разновидности абсолютных терминов-синонимов, различаемые по морфологической природе» имеет значения, именуемые следующими лингвистическими терминами:

1. «Абсолютные терминологические синонимы-существительные».
2. «Абсолютные терминологические синонимы-прилагательные».
3. «Абсолютные терминологические синонимы-глаголы».

Каждое из вышеуказанных значений образует соответствующий фрейм нижестоящего уровня. В качестве примера на Рисунке 5 приводится значение, именуемое «Абсолютные терминологические синонимы-существительные», которое в предлагаемом фрагменте выносится в виде одноименного фрейма на третий уровень модели. Данный фрейм содержит слот «Термины-существительные, образующие двухкомпонентный синонимический ряд», который имеет множество значений, именуемых терминами компьютерной лингвистики. В рассматриваемом фрагменте представлены два термина-значения:

1. «Фонация – голосообразование».
2. «Теггирование – разметка».

Таким образом, построение фреймово-сетевой модели дает возможность проанализировать логико-понятийную структуру предметной области «Компьютерная лингвистика», а также систематизировать организацию семантических терминообразовательных процессов и лексико-семантических отношений в исследуемой терминологии.

Заключение

Выполненное исследование позволило сделать следующие выводы:

1. Система терминов, применяемая для обозначения структуры знаний предметной области «Компьютерная лингвистика», может быть представлена в виде иерархически организованной терминологической фреймово-сетевой модели. Согласно правилам построения такой модели, связанные в многоуровневой сетевой структуре фреймы-термины содержат слоты, которые являются носителями определенных когнитивных признаков соответствующих понятий. Каждый термин, с помощью которого именуется значение слота, выносится на нижестоящий уровень модели в виде фрейма с соответствующим именем.

2. Представленные фрагменты терминологической сети фреймов иллюстрируют возможность использования предложенной модели для иерархической структуризации понятий, обозначаемых с помощью терминов рассматриваемой предметной области. Полученные результаты показали, что применение предложенной фреймовой модели может быть рекомендовано не только для анализа концептуально-семантической организации понятий предметной области «Компьютерная лингвистика», но и систематизации семантических терминообразовательных процессов и лексико-смысловых отношений в исследуемой терминологии.

3. Представленная в статье модель терминологической сети фреймов не содержит субфреймов, субслогов, микрослогов как подобластей, включаемых в главные области. Вместо подобных иерархий, изложенных в других публикациях, предложена однородная сетевая фреймово-слововая структура, которая на каждом уровне строится в соответствии с универсальными правилами. Это отличие, во-первых, дает возможность снять ограничения в наращивании системных связей «в глубину» исследуемой предметной области и, во-вторых, позволяет упростить алгоритмизацию построения иерархической структуры, что имеет принципиальное значение для представления терминологической системы в виде разветвленной компьютерной базы данных. Данные преимущества определяют теоретическую и практическую пользу предложенной модели.

Перспективы дальнейших исследований состоят в использовании представленных результатов при разработке информационно-терминологической базы данных предметной области «Компьютерная лингвистика».

Источники | References

1. Агузумцян Р. В., Великанова А. С., Польщиков К. А., Игитян Е. В., Лихошерстов Р. В. О применении интеллектуальных технологий обработки естественного языка и средств виртуальной реальности для поддержки принятия решений при подборе исполнителей проектов // Экономика. Информатика. 2021. Т. 48. № 2.
2. Евсюкова Т. В., Карнась А. А. Фреймовая структура терминологии предметной области «Индустрия туризма» (на материале русского и английского языков) // Вестник Адыгейского государственного университета. Серия 2: Филология и искусствоведение. 2021. № 2 (277).
3. Захаров В. П. Корпусная лингвистика. СПб.: Изд-во СПбГУ, 2005.
4. Ли Ян. Фрейм как средство описания туристических объектов // Вестник Воронежского государственного университета. Серия: Лингвистика и межкультурная коммуникация. 2020. № 3.
5. Луканин А. В. Автоматическая обработка естественного языка. Челябинск: Издательский центр ЮУрГУ, 2011.
6. Минский М. Фреймы для представления знаний. М.: Энергия, 1979.
7. Польщиков К. А., Лазарев С. А., Константинов И. С., Польщикова О. Н., Свойкина Л. Ф., Игитян Е. В., Балакшин М. С. Модель для оценки эффективности выполнения робототехнической системой коммуникативных функций // СТИН. 2020. № 6.
8. Польщиков К. А., Польщикова О. Н., Игитян Е. В., Балакшин М. С. Алгоритм поддержки принятия решений по выбору средств обработки больших массивов естественно-языковых данных // Научные ведомости Белгородского государственного университета. Серия: Экономика. Информатика. 2019. Т. 46. № 3.

9. Трофимова Н. А., Щитова О. Г. Когнитивно-фреймовое моделирование терминосистемы предметной области «строительные материалы» в русском языке XXI века // Вестник Томского государственного педагогического университета. 2022. № 1 (219).
10. Туманян Р. Г. Фреймовое моделирование терминологии предметной области «Энергетика» как способ представления специальных знаний (на материале английского и русского языков) // Современный ученый. 2022. № 1.
11. Khurana D., Koli A., Khatter K., Singh S. Natural language processing: State of the art, current trends and challenges // Multimedia Tools and Applications. 2023. No. 82.
12. Velikanova A. S., Polshchikov K. A., Likhosherstov R. V., Polshchikova A. K. The use of virtual reality and fuzzy neural network tools to identify the focus on achieving project results // Journal of Physics. 2021. Vol. 2060.

Информация об авторах | Author information

RU**Прохорова Ольга Николаевна¹**, д. филол. н., проф.**Польщикова Ольга Николаевна²**, к. филол. н., доц.^{1,2} Белгородский государственный национальный исследовательский университет**EN****Prokhorova Olga Nikolaevna¹**, Dr**Polshchikova Olga Nikolaevna²**, PhD^{1,2} Belgorod National Research University¹ prokhorova@bsu.edu.ru, ² polshchikova@bsu.edu.ru

Информация о статье | About this article

Дата поступления рукописи (received): 17.12.2023; опубликовано online (published online): 02.02.2024.

Ключевые слова (keywords): компьютерная лингвистика; терминология; фреймовая модель; сеть фреймов; слот; computational linguistics; terminology; frame model; frame network; slot.