

К идентификации ситуативных ролей сущностей в контексте задачи семантического информационного поиска

А. С. Гаврилкина, Н. В. Максимов, О. Л. Голицына

Национальный исследовательский ядерный университет «МИФИ»

asgavrilkina@yandex.ru, nv-maks@yandex.ru, olgolitsina@yandex.ru

Аннотация

В статье рассматриваются подходы к идентификации семантических ролей сущностей в контексте задачи семантического информационного поиска. Проанализированы различные определения ролей. Предложен подход к назначению ролей в соответствии с классами онтологии отношений, построенной на основе расширенной функциональной модели. Проведён эксперимент по оценке качества автоматической идентификации ролей Причина, Основание, Условие.

Ключевые слова: семантические роли, информационный поиск, автоматическая обработка текста, полнотекстовое индексирование

Библиографическая ссылка: Гаврилкина А. С., Максимов Н. В., Голицына О. Л. К идентификации ситуативных ролей сущностей в контексте задачи семантического информационного поиска // Компьютерная лингвистика и вычислительные онтологии. Выпуск 7 (Труды XXVI Международной объединённой научной конференции «Интернет и современное общество», IMS-2023, Санкт-Петербург, 26–28 июня 2023 г. Сборник научных статей). — СПб: Университет ИТМО, 2023. С. 21–31. DOI: 10.17586/2541-9781-2023-7-21–31

1. Введение

Целесообразность использования указателей роли для повышения точности отбора документов при информационном поиске была рассмотрена и реализована ещё на заре становления автоматизированных информационно-поисковых систем (см., например, [1, 2]). Этот очевидно целесообразный подход, однако, не получил развития и не нашёл в последующем практического применения не в последнюю очередь вследствие неоднозначности подходов к унификации типологии связей и усложнения языка запросов, а, соответственно, возникновения трудностей в использовании. Косвенным, но в итоге определяющим фактором были сравнительно малые объёмы информации и слабые вычислительные возможности. Отметим, что «минимальной основой для построения исчисления семантики является иерархия классов сущностей, отношений и свойств, базирующаяся на общей системе характеристических признаков. Однако обзор существующих решений показывает, что нет строгой и всеобщей классификации, хотя есть ряд вполне самодостаточных решений» [3].

Развитие ИТ и ВТ обеспечило существенное увеличение возможностей по обработке информации и, в частности, текстов документов, а онтологический подход позволил реализовать глубокий семантический поиск [4]. В этом случае поисковые образы представляют полномасштабную концептуальную модель объекта, отражая существо его функционирования — это не только понятийный состав (традиционные ключевые слова), но все именованные сущности и ситуативные отношения. Построение основывается на выделении сущностей и связей между ними при помощи шаблонов, связывающих сущности

и отношения с их вербальными формами представления [5]. Выражение имён сущностей и отношений на знаковом уровне позволяет индексировать элементарный факт, как триплет — последовательность знаков, в которой представлены не только имена, но и типы сущностей и отношений. Таким образом, могут быть построены как традиционные индексы (по ключевым словам), так и индексы, представляющие семантические связи. Наличие таких индексов позволяет в рамках традиционной теоретико-множественной модели информационного поиска (и средствами традиционного дескрипторного ИПЯ) реализовать отбор документов с учётом имманентных и ситуативных отношений между сущностями. ИПЯ при этом попадает в семантической классификации в класс языков, имеющих средства выражения (и различения) и имманентных, и ситуативных отношений.

При этом связи между сущностями (указывающие на взаимодействие/соотношение объектов действительности) могут быть идентифицированы как через отношения, которые выражаются преимущественно глагольными и предложными конструкциями, так и через роли, выделяемые на основе имён сущностей (то есть на понятийном уровне присутствует условное разделение на отношения и роли). Отметим, что понятие роли вводится и в языке онтологического моделирования Gellish стандарта ISO-15926 [6].

Таким образом, роли могут использоваться как параметр для поиска (квалификатор в условии отбора), позволяющий выделить множество имён сущностей с общими для связанных с определенным классом предикатов свойствами, или в качестве одного из параметров при поиске на графах — при построении отображения графа, например, на основе функциональной модели, или при задании аспектной проекции [4].

В настоящей работе будут рассмотрены различные определения ролей и описан подход к автоматической идентификации ролей сущностей в тексте в соответствии с классификацией отношений, построенной на основе расширенной функциональной модели.

2. К определениям понятия «роль»

Рассмотрим некоторые определения ролей.

Роль — это функциональное назначение, которое объект (нечто, anything) выполняет в данной ассоциации по отношению к связи, или наоборот, связь по отношению к объекту [7].

В [8] семантическая роль имени при предикате определяется как часть семантики предиката, которая отражает общие свойства участников определенных групп ситуаций.

В [9] семантические роли рассматриваются как классы, к которым сводятся лексические партиципаны. Предполагается, что таких классов должно быть небольшое обозримое число.

Исходя из приведённых определений можно сказать, что у сущности существует роль относительно предиката, ситуации или фрейма, которая отражает основные свойства для группы. Здесь семантический фрейм, согласно [10], представляет собой совокупность фактов, определяющих «характерные черты, атрибуты и функции денотата, а также его характерные взаимодействия с вещами, обязательно или типично с ним связанными». В лингвистике фрейм ассоциируется с определенным предикатом (а чаще несколькими близкими по смыслу предикатами), при котором выражаются участники фрейма (аргументы) [11].

В настоящей работе роль рассматривается как проекция отношения на объект, причём отношения могут быть как связями между взаимодействующими объектами, так и между объектами и обстоятельствами. То есть роль выступает как форма объективизации взаимосвязи. Это особенно заметно, например, в случае отношения «вход», когда «поглощение» входного объекта для обработки является скорее событием, чем действием, и его очевидно удобнее идентифицировать ролью «исходный материал» [12].

Сущность может быть соотнесена с несколькими ролями, так как количество ролей зависит непосредственно от количества отношений, в которых она состоит. Например, подлежащему, как главному члену предложения, являющемуся наиболее очевидным

вариантом для выбора в пару «сущность — отношение» может быть назначено несколько ролей, так как оно связывается с обстоятельствами и дополнениями и таким образом получает роль для идентификации в каждом отношении. То есть роль представляет основные свойства сущности в контексте конкретного отношения. И наоборот, идентификатора роли должно быть достаточно, чтобы сказать, каким типом отношения связана сущность.

3. Типизация отношений и ролей

В последние десятилетия активно решается задача автоматической разметки ролей, которая заключается в автоматической идентификации участников ситуаций, обозначаемых предикатами (глаголами, существительными, прилагательными и т. д.), и разметка способа их выражения — вне зависимости от того, связаны ли обозначающие участников лексические единицы с предикатом синтаксически или нет. [11]

Определение ролевой структуры высказывания позволяет соотносить похожие по смыслу предложения независимо от их синтаксических структур. Ролевые структуры высказываний используются во многих прикладных задачах: в вопросно-ответном поиске, машинном переводе, оценке смысловой близости фрагментов текстов на естественном языке и др. [13].

Однако, учитывая многообразие предикатов в языке, в том числе со схожими в некоторых аспектах значениями, требуется определить признаки для обобщения.

С точки зрения [14] подходы к определению ролей можно разделить согласно уровню обобщения: при максимальном обобщении роли сводятся к двум — Actor и Undergoer, где роль Actor включает роли агенса, экспериментера, инструмента и др., а Undergoer — роли пациенса, темы, приёмника и т.д.; с противоположной стороны спектра находятся предикатно-специфичные роли, то есть уникальные для каждого предиката; все наборы ролей с некоторым обобщением находятся в середине спектра.

К наборам ролей из середины спектра можно отнести предложенный Ч. Филмором классический набор ролей, который он считал универсальным для всех языков: агентив, объектив, датив, инструменталис, фактив, локатив [15].

При решении задачи семантической разметки ролей в большинстве случаев используются предикатно-специфичные роли, что связано с популярностью методов машинного обучения и ограниченностью размеченных корпусов для их применения. Наиболее известны изначально англоязычные аннотированные семантическими ролями корпуса текстов PropBank [16] и FrameNet [17], которые позже были сопоставлены и продублированы на других языках, первый из которых ориентирован на аннотацию глаголов, а второй — на заполнение фреймов (в которых предикаты могут быть сгруппированы в типовых ситуациях). Для русского языка был создан FrameBank, основанный на концепции фреймов, но всё же сосредоточенный на описании конкретных глаголов и их окружения (то есть анализируется возможная сочетаемость глагола с предлогами и назначаются роли участникам ситуации). Однако хотя FrameNet и FrameBank и допускают разные уровни обобщения, классификационные признаки для группировки фреймов и ролей чётко не обозначены.

Рассмотрим более подробно корпус FrameBank [11], который рассматривается создателями с одной стороны, как ресурс для тренировки систем искусственного интеллекта, с другой — как инструмент для изучения конструкционных свойств русской лексики.

FrameBank включает словарь лексических конструкций и корпус примеров их употребления. В шаблонах лексических конструкций указываются морфосинтаксические характеристики элементов конструкции, синтаксический ранг участника, роль (экспликация) участника, лексико-семантические ограничения на заполнение слота конструкции, статус участника: обязательный или факультативный, буква, маркирующая

участника в кратком паттерне. Ядро системы FrameBank составляют 2200 частотных русских глаголов. В корпусной части ресурса FrameBank представлено приблизительно по 100 примеров из НКРЯ на каждый глагол. В работах [13, 18] описывается использование FrameBank для автоматической разметки ролей с применением машинного обучения.

В FrameBank на основе ролей согласно [19] выделены 88 базовых экспликаций семантических ролей, которые разделены на следующие группы [11]:

- блок Агенса;
- блок Пациенса;
- блок Экспериенцера;
- блок Инструмента;
- блок Адресата;
- блок обстоятельственных характеристик (группы Места, Времени, Параметров, Признаков, Причины, Цели);
- группа посессивных ролей;
- группа Источников и Ресурсов.

В настоящей работе к классификации отношений применяется подход на основе расширенной функциональной модели. Акт процесса (реализация операции функциональной обработки) целенаправленной и управляемой деятельности (познавательной, производственной и т.д.) рассматривается с точки зрения теории систем и может быть представлен средствами функционального моделирования [12].

Для типизации ситуативных (функциональных) отношений (и соотнесённых с ними ролей) используется ориентированная на научно-техническую сферу онтология отношений, основанная на трехуровневой иерархической классификации. Первый уровень отражает соотношение реальность/модель, второй — комбинации соотношений отдельного (часть) и агрегатного (целое), третий уровень построен по признаку формы проявления отношения — действие-ориентированные, объект-ориентированные и результат-ориентированные. Листьями иерархического дерева являются классы отношений, обладающие комбинацией свойств верхних уровней. Каждому классу поставлено в соответствие множество лингвистических конструкций, которые выражают его семантику. Классы нижнего уровня открыты для пополнения и содержимое их может зависеть от вида/жанра текстового массива, подлежащего обработке. Для удобства восприятия и эксплуатации онтология отношений представлена в виде иерархически упорядоченной таксономии. [5]

На верхнем уровне классификации находятся 27 классов отношений, среди которых Создание, Воздействие, Изменение, Идентификация, Отождествление, Зависимость-связь, Размещение, Разъединение и т.д. [3]. Каждый класс отношений соотносится с набором ролей, назначаемых его подклассам, например, сущностям, связанным отношениями из класса Зависимость-связь, могут быть назначены следующие роли: Вход, Выход, Цель, Результат, Причина, Следствие, Ресурс, Ограничение, Средство, Инструмент, Управление, Условие, Помеха, Основание, Назначение.

Так как в данной работе роли рассматриваются как свойства связи и соотнесены с классами отношений, то для идентификации ролей используются классифицированные лингвистические конструкции отношений (преимущественно глаголы, отглагольные части речи и предлоги), при помощи которых отношение может быть представлено в тексте, их морфологические характеристики (возвратность, наклонение глаголов, вид причастий), а также место сущности в триплете. Уровни обобщения ролей соответствуют уровням таксономии отношений, но при этом есть основания полагать, что они могут группироваться иначе, чем классы отношений.

Кроме того, следует учитывать и взаимозависимость сущностей и отношений, т. е. возможность совершения действия того или иного характера над объектами определённых классов. Например, некоторые действия можно выполнить только с одушевлёнными объектами или только с абстрактными и т.д. При этом часто встречается и перенос

буквальных значений с одушевлённого на неодушевлённое, например, для «дышать», «есть», «жить».

Как отмечалось выше для идентификации ролей использовалась онтология отношений, ориентированная на научно-техническую сферу. Поэтому для сравнительной оценки адекватности идентификации ролей сопоставим результаты с получаемыми на основе подхода FrameBank:

- 1) в тексте *Земля представляет собой сфероид* у сущности *Земля* определена роль Идентифицируемое, связанная с классом отношений Идентификация, в то время как в FrameBank — это Статус в блоке Инструмент, который является частью Блока обстоятельственных характеристик;
- 2) в предложении *Билет стоил тысячу рублей* у сущности «тысяча рублей» определена роль Отождествляемое, связанная с классом отношений Отождествление, в FrameBank — Цена в блоке Параметры, который является частью Блока обстоятельственных характеристик;
- 3) в результате обработки предложения *Он разломал кусок хлеба на 4 части* будет построено два триплета <Он> разломал <кусок хлеба>, <кусок хлеба> разломал на <4 части>, где для сущности *кусок хлеба* определена роль Разъединяемое, для сущности *4 части* — Части разъединённого, соотнесённые в с классом отношений Разъединение, во FrameBank сущности *4 части* назначается роль Результат в блоке Пациент.

Таким образом, можно сделать вывод, что в естественно-научной классификации отношений некоторые классы более детализированы (например, класс отношений Изменение включает подклассы Изменение состояния, Изменение формы, Изменение состава и пр. и соответствующие роли) или ключевой признак сущности, по которому определена роль, отличается, что связано с задачей и предметной областью. Для разметки научно-технических текстов FrameBank избыточен, кроме того, в словаре лексических конструкций не были обнаружены некоторые характерные для научно-технических текстов производные предлоги (в частности «с целью», «по причине», «в соответствии с», «в условиях», «в связи с», «вследствие»). Однако для обработки художественных текстов множество лингвистических конструкций отношений классификации может быть пополнено предикатами из FrameBank.

4. Эксперимент по оценке качества идентификации ролей

Для оценки качества идентификации ролей использовались 19 текстов общим объёмом 150 Кб (19 426 слов), после обработки которых с помощью опытного образца сервиса визуального онтологического анализа научно-технических текстов [20] было получено 3619 имён сущностей (уникальных в рамках каждого текста) и 4234 триплетов с отношениями из текста.

Для экспертной оценки были выбраны роли Причина, Основание, Условие, связанные с подклассами класса Зависимость-связь: Быть причиной, Быть основанием, Быть условием соответственно. В классификации лингвистических конструкций отношений для идентификации роли Причина предназначено 20 лингвистических конструкций отношений, Основание — 22, Условие — 16, из которых использовалось 12, 7 и 15, соответственно. В эксперименте идентифицировалась роль только первой сущности. В результате для 40 сущностей была определена роль Причина, для 62 — роль Основание, для 59 — роль Условие. Применённый метод автоматической обработки текста допускает формирование некорректных триплетов, однако при оценке были рассмотрены только триплеты с корректно определёнными отношением и сущностью, которой назначена роль.

Таблица 1. Количество и процентное соотношение сущностей с корректно и некорректно определёнными ролями Причина, Основание, Условие

	Причина	Основание	Условие
Количество имён сущностей, у которых корректно определена роль	40 (100 %)	48 (77%)	58 (98%)
Количество имён сущностей, у которых некорректно определена роль	0	14 (23%)	1 (2%)

Были проанализированы случаи некорректного определения ролей и выявлены следующие причины:

- многозначность предлога «по», по которому идентифицировано 93% сущностей с ролью Основание: в сочетаниях «по мнению экспертов», «по ветру», «по плану» и других случаях, когда предлог «по» может быть заменён «согласно» или «в соответствии с», назначение роли считалось корректным в отличие от случаев «завод по обогащению урана», «программе по развитию процессов памяти» и прочих, где речь идёт о теме или предмете деятельности;
- возможность формирования триплетов, в которых некорректно связаны имена сущностей, которые в качестве членов предложения идентифицируются как дополнения.

Первая проблема, может быть, в некоторой мере решена пополнением множества классифицированных лингвистических конструкций отношений распространёнными сочетаниями с предлогами (преимущественно для повышения точности идентификации при обработке текстов художественного стиля), вторая — введением ограничений на падежи главных слов в словосочетаниях, связываемых определёнными отношениями.

В таблицах 2-4 приведены примеры триплетов с корректно определённой ролью сущности и фрагменты текста, в результате обработки которых они составлены. Имена сущностей и отношений в триплетах приведены в нормальной форме (пословная нормализация). Код указывает на класс отношения в таксономии отношений¹. В правой колонке курсивом выделена лингвистическая конструкция отношения, указывающая на роль сущности, полужирным — сущность, которой присваивается роль.

Таблица 2. Примеры триплетов с отношением из класса Быть причиной и сущностями с ролью Причина

Триплет, включающий сущность с ролью Причина	Фрагмент текста
<поддержание межличностный отношение> из-за 12315[из-за] <особенность мировосприятия>	Чрезмерное пребывание за компьютером в ущерб всему остальному, трудности в установлении и поддержании межличностных отношений <i>из-за особенностей мировосприятия</i> и познавательных процессов, обусловленных взаимодействием с виртуальной реальностью компьютерных технологий, приводят к упрощению системы отношений «человек - человек» у подростков, включённых в компьютеризированную деятельность

¹ Упомянутые в работе коды соотносятся с классами отношений следующим образом: 12315 — Быть причиной, 1231Е — Быть основанием, 1231С — Быть условием, 1121 — Передача, 11212 — Излучение, 1133 — Явление, факт, 1211 — Изменение, 12113 — Изменение состояния, 1213 — Усложнение, 1232 — Зависимость-соотношение, 1322 — Определение, 2111 — Присоединение, 22212 — Локативность в пространстве, 2223 — Обладание.

Продолжение таблицы 2

Триплет, включающий сущность с ролью Причина	Фрагмент текста
<характеристика> обладать в связи с 12315[в связь с] 2223[обладать] <опосредованностью компьютер>	Все названные формы общения <i>в связи с его опосредованностью компьютером</i> обладают такой характеристикой, как анонимность, которая имеет целый ряд последствий
<словарь русский язык> пополняться за счёт 12315[за счёт] 2111[пополнять] <англицизм>	Наблюдается картина, где словарь русского языка, в основном, пополняется <i>за счёт англицизмов</i>
<дефицит вод> вызывать 12315[вызывать] <снижение работоспособность>	Дефицит воды в организме <i>вызывает</i> снижение работоспособности, потеря воды в количестве 10% от массы тела приводит к нарушению обмена веществ, потеря 15-20% смертельна при температуре воздуха 30°C, а потеря 25% абсолютно смертельна
<актуальность водоснабжение населённый пункт> кардинально возрастать в случай 12315[в случай] 1211[возрастать] <разрушение инфраструктура жизнеобеспечение население>	Следовательно, <i>в случае разрушения инфраструктуры жизнеобеспечения населения</i> актуальность водоснабжения населённых пунктов кардинально возрастает
<феномен профессиональный деятельность> в результат 12315[в результат] <стресс>	В новой редакции Международной классификации болезней (МКБ-11), разработанной ВОЗ (вступит в силу с 1 января 2022 года), выгорание официально зарегистрировано как феномен профессиональной деятельности <i>в результате стресса</i>
<применение технология реверсивный обучение> являться по причина 12315[по причина] 1133[являть] <переворачивание традиционный модель обучение>	Применение технологии реверсивного обучения при обучении иностранным языкам, на наш взгляд, является эффективной технологией обучения <i>по причине переворачивания традиционной модели обучения</i> , вследствие которого домашнее задание выполняется в аудитории, тогда как классная работа выносится на самостоятельное изучение
<они> могут восприниматься вследствие 11212[воспринимать] 12315[вследствие] <распад связь время>	Массовым сознанием <i>вследствие распада связи времён</i> они могут восприниматься как новые

Класс отношений Быть причиной содержит преимущественно производные предлоги, вследствие однозначности которых точность определения роли Причина высока.

Таблица 3. Примеры триплетов с отношением из класса Быть основанием и сущностями с ролью Основание

Триплет, включающий сущность с ролью Основание	Фрагмент текста
<агрегирование> всегда основываться на 1231E[основывать на] <абстрагирование>	Агрегирование всегда <i>основывается на абстрагировании</i> , т. е. отвлечении от несущественных моментов и выделения наиболее значимых, существенных, типичных черт, закономерностей экономических процессов и явлений
<вершина эволюция> благодаря 1231E[благодаря] <память>	Человек поднялся на «вершину эволюции», <i>благодаря</i> своей памяти и её постоянному совершенствованию
<память> определить по 1322[определить] 1231E[по] <мнение Р.С. Немова>	<i>По мнению Р. С. Немова</i> , память можно определить, как способность к получению, хранению и воспроизведению жизненного опыта

Продолжение таблицы 3

Триплет, включающий сущность с ролью Основание	Фрагмент текста
<сводный результат> быть получить следовать по 1231E[по] 1121[получить] 8[получить] 12315[следовать] <все четыре методика>	В процессе проведённого исследования (<i>по всем четырём методикам</i>) были получены следующие сводные результаты
<неопытный команда> быть полностью поразить по 1231E[по] <ветер>	В Бискайском заливе судно столкнулось с сильнейшими штормами и шло <i>по ветру</i> , а неопытная команда была полностью поражена морской болезнью
<психологический основа конструирование различный вариант УПО студент> на основа 1231E[на основа] <компонент>	В данной статье обсуждается вопрос о психологической основе конструирования различных вариантов УПО студентов <i>на основе компонентов</i> в его структуре
<жизнь> развиваться по 1231E[по] 1213[развивать] <единый закон эволюция>	Жизнь на них развивается <i>по единому закону эволюции</i> , восходя ко все более сложным формам и видам, достигая постепенно своего высшего уровня, разумной жизни

Класс отношений Быть основанием включает многозначный предлог «по», из-за которого требуется дополнительный анализ окружения для более точного определения роли присоединяемой им сущности.

Таблица 4. Примеры триплетов с отношением из класса Быть условием и сущностями с ролью Условие

Триплет, включающий сущность с ролью Условие	Фрагмент текста
<вероятность применение> подтверждать в условие 1231C[в условие] 1232[подтверждать] <рост геополитический напряжение международный отношенье>	<i>В условиях роста геополитического напряжения международных отношений</i> в настоящее время подтверждает вероятность применения в современной войне оружия массового поражения
<физиологический потребность> зависеть от 1231C[зависеть от] <возраст>	Физиологическая потребность в воде <i>зависит от возраста</i> , характера работы, пищи, профессии, климата и т.д.
<название> обусловить 1231C[обусловить] <особенность ритмика электроэнцефалограмма>	Существует две фазы сна: медленная и быстрая. Эти названия <i>обусловлены особенностями ритмики электроэнцефалограммы (ЭЭГ)</i> во время сна медленной активностью в ФМС и более быстрой в ФБС
<мозг человек> переключаться при 12113[переключать] 1231C[при] 22212[при] <недостаток сон>	<i>При недостатке сна</i> мозг человека переключается на более примитивные формы деятельности и не в состоянии нормально управлять эмоциями, эмоциональные зоны мозга становятся реактивными
<изменение> определяться в рамка 1231C[в рамка] 1322[определять] <обогащающей модели>	В связи с этим эффективность обучения студентов <i>в рамках «обогащающей модели»</i> определяется изменениями в их интеллекте как в специфической форме организации индивидуального познавательного (ментального) опыта, обеспечивающей возможность эффективного включения в учебно-познавательную деятельность студентов, так и в учебно-познавательном опыте как результате этой деятельности

5. Заключение

В данной работе оценивалась адекватность автоматического назначения ролей, определенных в соответствии с классами отношений, применительно к онто-графовому представлению содержания текста.

Использование для идентификации отношений и сущностей классификации с явно заданными характеристическими признаками делает возможным количественно определять семантическую близость между сущностями и поисковыми образами, а использование ролей сущностей повышает точность поиска и позволяет при поиске обращаться к ролям сущностей как к самостоятельному поисковому атрибуту.

Проведённый эксперимент на текстах разного стиля показал минимальную точность 77% и максимальную — 100% (для ролей Основание и Причина, соответственно) и, таким образом, подтверждается целесообразность использования предложенного подхода для автоматизированного индексирования и поиска научно-технической литературы. Однако для текстов художественного стиля множество лингвистических конструкций для идентификации отношений и ролей должно быть расширено.

Литература

- [1] Скороходько Э. Ф. Лингвистические проблемы обработки теистов в автоматизированных информационно-поисковых системах // Вопросы информационной теории и практики. 1974. № 25. С. 5–120.
- [2] СИНТОЛ // Сборник переводов по вопросам информационной теории и практики. М.: ВИНТИ, 1968. С. 36–47; 50–52; 66–72; 76–80.
- [3] Максимов Н. В., Гаврилкина А. С., Андропова В. В., Тазиева И. А. Систематизация и идентификация семантических отношений в онтологиях научно-технических предметных областей // Научно-техническая информация. Сер. 2: Информационные процессы и системы. 2018. № 11. С. 32–42.
- [4] Максимов Н. В., Голицына О. Л. От семантического к когнитивному информационному поиску. Основные положения и модели глубинного семантического поиска // Научно-техническая информация. Сер. 2: Информационные процессы и системы. 2022. № 6. С. 1–16.
- [5] Голицына О. Л., Гаврилкина А. С. Об одном подходе к выделению имён сущностей и связей в задаче построения семантического поискового образа // Научно-техническая информация. Серия 2: Информационные процессы и системы. 2021. № 3. С. 17–26.
- [6] Van Renssen A. Gellish: A Generic Extensible Ontological Language. Delft: Delft University Press, 2005. 238 с.
- [7] Kozaki K., Sunagawa E., Kitamura Y., Mizoguchi R. Role Representation Model Using OWL and SWRL // Roles'07. Proceedings of the 2nd Workshop on Roles and Relationships in Object Oriented Programming, Multiagent Systems, and Ontologies. Berlin, 2007. С. 39–46.
- [8] Плунгян В. А. Основные синтаксические грамлеммы имени // Общая морфология: Введение в проблематику: Учебное пособие. Изд. 2-е, исправленное. М.: Едиториал УРСС, 2003. 384 с.
- [9] Семантическая роль как элемент метаязыков общей и специальной типологии // 40 лет Санкт-Петербургской типологической школе. М.: Знак, 2004. С. 233–252.
- [10] Allan K. Natural language semantics. Oxford, UK; Malden, Mass.: Blackwell, 2001. 529 p.
- [11] Кашкин Е. В., Ляшевская О. Н. Типы информации о лексических конструкциях в системе ФреймБанк // Труды института русского языка им. В.В. Виноградова. 2015. Т. 6. С. 464–556.
- [12] Максимов Н. В. Методологические основы онтологического моделирования документальной информации // Научно-техническая информация. Сер. 2: Информационные процессы и системы. 2018. №3. С. 6–22.

- [13] Шелманов А. О., Каменская М. А. Обучение анализатора для определения ролевых структур высказываний в текстах на русском языке на автоматически размеченном корпусе // Труды Института системного анализа Российской академии наук. 2017. Т. 67, №. 2. С. 104–120.
- [14] Van Valin R. D. Generalized semantic roles and the syntax-semantics interface // Empirical issues in formal syntax and semantics. 1999. Vol. 2. P. 373–389.
- [15] Fillmore Ch. The case for case. // Universals in linguistic theory / Ed by E. Bach and R. T. Harms. N. Y., Chicago, San Francisco, 1968.
- [16] Kingsbury P. R., Palmer M. From TreeBank to PropBank // LREC. 2002. P. 1989–1993.
- [17] Fillmore C. J. et al. FrameNet in action: The case of attaching // International journal of lexicography. 2003. Vol. 16 (3). P. 297–332.
- [18] Апресян Ю. Д., Богуславский И. М., Иомдин Л. Л., Санников В. З. Теоретические проблемы русского синтаксиса: взаимодействие грамматики и словаря. М., 2010.
- [19] Автоматическая разметка семантических ролей в русском языке: автореферат дис. ... кандидата филологических наук: 10.02.21 / Кузнецов Илья Олегович; [Место защиты: Моск. гос. ун-т им. М.В. Ломоносова]. М., 2016. 25 с.
- [20] Свидетельство о государственной регистрации программы для ЭВМ № 2021610648, 15.01.2021 / Авторы: Максимов Н.В., Голицына О.Л., Монанков К.В., Гаврилкина А.С.

On the Identification of Situational Roles of Entities in the Context of the Semantic Information Retrieval Task

Anastasiia S. Gavrilkina, Nikolay V. Maksimov, Olga L. Golitsina

National Research Nuclear University MEPhI (Moscow Engineering Physics Institute)

The article discusses approaches to identifying the semantic roles of entities in the context of the task of semantic information retrieval. Various definitions of roles are analyzed. An approach is proposed to the distribution of roles in accordance with the classes of the ontology of relations built on the basis of an extended functional model. An experiment was carried out to assess the quality of automatic determination of the roles Problem, Foundation, Condition.

Keywords: semantic roles, information retrieval, automatic text processing, full-text indexing

Reference for citation: Gavrilkina A.S., Maksimov N.V., Golitsina O.L. On the Identification of Situational Roles of Entities in the Context of the of Semantic Information Retrieval Task // Computational Linguistics and Computational Ontologies. Vol. 7 (Proceedings of the XXVI International Joint Scientific Conference «Internet and Modern Society», IMS-2023, St. Petersburg, June 26–28, 2023). - St. Petersburg: ITMO University, 2023. P. 21–31. DOI: 10.17586/2541-9781-2023-7-21–31

Reference

- [1] Skorohod'ko E. F. Lingvisticheskie problemy obrabotki teistov v avtomatizirovannykh informacionno-poiskovykh sistemah // Voprosy informacionnoj teorii i praktiki. 1974. № 25. S. 5–120. (in Russian).
- [2] SINTOL // Sbornik perevodov po voprosam informacionnoj teorii i praktiki. M.: VINITI, 1968. S. 36–47; 50–52; 66–72; 76–80. (in Russian).
- [3] Maksimov N. V., Gavrilkina A. S., Andronova V. V., Tazieva I. A. Sistematizaciya i identifikaciya semanticheskikh otnoshenij v ontologiyah nauchno-tehnicheskikh predmetnyh

- oblastej // Nauchno-tekhnicheskaya informaciya. Ser. 2: Informacionnye processy i sistemy. 2018. № 11. S. 32–42. (in Russian)
- [4] Maksimov N. V., Golicyna O. L. Ot semanticheskogo k kognitivnomu informacionnomu poisku. Osnovnye polozheniya i modeli glubinnogo semanticheskogo poiska // Nauchno-tekhnicheskaya informaciya. Ser. 2: Informacionnye processy i sistemy. 2022. № 6. S. 1–16. (in Russian)
- [5] Golicyna O. L., Gavrilkina A. S. Ob odnom podhode k vydeleniyu imyon sushchnostej i svyazej v zadache postroeniya semanticheskogo poiskovogo obraza // Nauchno-tekhnicheskaya informaciya. Seriya 2: Informacionnye processy i sistemy. 2021. № 3. S. 17–26. (in Russian)
- [6] Van Renssen A. Gellish: A Generic Extensible Ontological Language. Delft: Delft University Press, 2005. 238 p.
- [7] Kozaki K., Sunagawa E., Kitamura Y., Mizoguchi R. Role Representation Model Using OWL and SWRL // Roles'07. Proceedings of the 2nd Workshop on Roles and Relationships in Object Oriented Programming, Multiagent Systems, and Ontologies. Berlin, 2007. P. 39–46.
- [8] Plungyan V. A. Osnovnye sintaksicheskie grammemy imeni // Obshchaya morfologiya: Vvedenie v problematiku: Uchebnoe posobie. Izd. 2-e, ispravlennoe. M.: Editorial URSS, 2003. 384 s. (in Russian)
- [9] Semanticheskaya rol' kak element metazykov obshchej i special'noj tipologii // 40 let Sankt-Peterburgskoj tipologicheskoy shkole. M.: Znak, 2004. S. 233–252. (in Russian)
- [10] Allan K. Natural language semantics. Oxford, UK; Malden, Mass.: Blackwell, 2001. 529 p.
- [11] Kashkin E. V., Lyashevskaya O. N. Tipy informacii o leksicheskikh konstrukciyah v sisteme Frejmbank // Trudy instituta russkogo yazyka im. V.V. Vinogradova. 2015. T. 6. S. 464–556. (in Russian)
- [12] Maksimov N. V. Metodologicheskie osnovy ontologicheskogo modelirovaniya dokumental'noj informacii // Nauchno-tekhnicheskaya informaciya. Ser. 2: Informacionnye processy i sistemy. 2018. №3. S. 6–22. (in Russian)
- [13] Shelmanov A. O., Kamenskaya M. A. Obuchenie analizatora dlya opredeleniya rolevyh struktur vyskazyvanij v tekstah na russkom yazyke na avtomaticheski razmechennom korpuse // Trudy Instituta sistemnogo analiza Rossijskoj akademii nauk. 2017. T. 67, №. 2. S. 104–120. (in Russian)
- [14] Van Valin R. D. Generalized semantic roles and the syntax-semantics interface // Empirical issues in formal syntax and semantics. 1999. Vol. 2. P. 373–389.
- [15] Fillmore Ch. The case for case. // Universals in linguistic theory. Ed by E. Bach and R. T. Harms. N. Y., Chicago, San Francisco, 1968.
- [16] Kingsbury P. R., Palmer M. From TreeBank to PropBank // LREC. 2002. P. 1989–1993.
- [17] Fillmore C. J. et al. FrameNet in action: The case of attaching // International journal of lexicography. 2003. Vol. 16. № 3. P. 297–332.
- [18] Apresyan Yu. D., Boguslavskij I. M., Iomdin L. L., Sannikov V. Z. Teoreticheskie problemy russkogo sintaksisa: vzaimodejstvie grammatiki i slovarya. M., 2010. (in Russian)
- [19] Avtomaticheskaya razmetka semanticheskikh rolej v russkom yazyke: avtoreferat dis. ... kandidata filologicheskikh nauk: 10.02.21 / Kuznecov Il'ya Olegovich; [Mesto zashchity: Mosk. gos. un-t im. M. V. Lomonosova]. M., 2016. 25 s. (in Russian)
- [20] Svidetel'stvo o gosudarstvennoj registracii programmy dlya EVM № 2021610648, 15.01.2021 / Avtory: Maksimov N. V., Golicyna O. L., Monankov K. V., Gavrilkina A. S. (in Russian)