

СЕМАНТИЧЕСКИЙ АНАЛИЗ В ПРИКЛАДНЫХ СИСТЕМАХ

Шухрат Ильхамович Ходиев¹

¹ Национальный Университет Узбекистана имени М.Улугбека, Ташкент

¹ E-mail: ilch17333@gmail.com

DOI: 10.61587/mmit.tiue.uz.v1i1.269

Аннотация. В работе исследуются методы, алгоритмы и архитектурные решения для интеграции семантического анализа в современные прикладные программные системы. Рассматриваются подходы к практическому применению семантических технологий (в области языков программирования и компиляторов, включая современные языковые модели), в таких прикладных областях, как интеллектуальный поиск, электронные каталоги и базы данных, конвейер обработки SQL. Анализируются ключевые проблемы внедрения семантических модулей, включая вычислительную сложность и интеграцию с существующей ИТ-инфраструктурой.

Ключевые слова: семантический анализ, формальные модели программ, формальная семантика, системный подход и системный анализ, статический анализ, преобразование программ, оптимизирующая трансляция, технологии обработки языковых данных; межмодульный анализ.

SEMANTIC ANALYSIS IN APPLIED SYSTEMS

Shukhrat Ilkhamovich Khodiev¹

¹ National University of Uzbekistan named after Mirzo Ulugbek, Tashkent

¹ E-mail: ilch17333@gmail.com

Abstract. This paper investigates methods, algorithms, and architectural solutions for integrating semantic analysis into modern applied software systems. Approaches to the practical application of semantic technologies (within the domain of programming languages and compilers, including modern language models) are examined across such applied areas as intelligent search, electronic catalogs and databases, and the SQL processing pipeline. The key challenges of deploying semantic modules, including computational complexity and integration with existing IT infrastructure, are analyzed.

Keywords: semantic analysis, formal program models, formal semantics, systems approach and systems analysis, static analysis, program transformation, optimizing compilation, language data processing technologies, interprocedural analysis.

Введение

Решение множества проблем программирования требует разработки формальных методов задания семантики языков. Отсутствие точного метода описания семантики влечёт серьёзные проблемы, в частности в понимании конструкций языков, в реализации этих языков. Задачами формальной семантики, помимо прочего, являются построение систем доказательств правильности программ, анализ свойств программ, например, проблем завершимости программ, возможности их заикливания. Соответственно, возникает необходимость создания семантических теорий, поддерживающих языки, или языков, обеспечивающих возможность применения имеющихся семантических теорий для анализа программ. В предлагаемой работе рассмотрим сферу применения методов формальной семантики для разработки и реализации языков, в прикладных системах, создания на основе языков систем программирования, их анализа. Семантический анализ стал применяться в информационно-поисковых системах, электронных каталогах и базах данных. Возникли методы автоматического индексирования документов, тематической классификации и интеллектуального поиска. Затем его идеи были перенесены на обработку данных, документов и естественного языка. Началось активное

развитие исследований по обработке естественного языка. Учёные пытались научить компьютеры понимать смысл текстов, а не только анализировать их грамматику.

Приводимая работа написана на опыте реализации систем построения трансляторов, окружений программирования встроенных систем, преподавания курсов по технологиям и языкам программирования соответственно. На основе семантического анализа был спроектирован и реализован прототип системы управления базами данных (СУБД) с открытой архитектурой, предназначенный для учебных и исследовательских целей, который реализует полный цикл обработки SQL-запроса от лексического анализа текста до физического выполнения плана и возврата итоговой выборки [1,8,12].

Семантический анализ в области языков, компиляторов, прикладных системах

исторически семантический анализ впервые сформировался в области языков программирования и компиляторов. Там возникла необходимость не только проверять синтаксис программ, но и анализировать их смысл. Формальный семантический анализ программ появился и стал одной из основ теории компиляции. Компилятор должен был определить, объявлены ли переменные, совместимы ли типы данных, корректны ли вызовы процедур и функции. Были разработаны основы формальной семантики языков программирования, теории компиляции и методы семантического анализа программ. Оптимизации, автоматизация конструирования программ, системы синтеза, конкретизации, средства семантического поиска, и другие, основаны на семантических преобразованиях программ. Это существенно отличается от семантического анализа в интеллектуальных системах, где вычисляются уже не типы и атрибуты программы, а смысловые связи, вероятности интерпретаций, контекст и знания о предметной области [2,9-13].

В прикладные системы идеи семантического анализа пришли из более ранних исследований по формальным языкам, компиляторам и математической логике, хотя затем эти направления стали развиваться самостоятельно и использовать разные методы. Семантический анализ в прикладных информационных системах, экспертных системах и обработке естественного языка начал активно развиваться немного позже. Он опирается на широкий круг математических дисциплин, поскольку его основная задача связана не только с обработкой данных, но и с выявлением их смысла, взаимосвязей и закономерностей. Набор используемых математических методов зависит от конкретной задачи: поиска информации, анализа текстов, построения рекомендаций, работы экспертных систем, поддержки принятия решений или реализации технологий искусственного интеллекта.

Математическая логика обеспечивает формальное представление знаний, позволяет описывать факты, правила и отношения между объектами. Методы математической логики широко применяются в экспертных системах, системах поддержки принятия решений, базах знаний и семантических сетях.

теория множеств позволяет описывать объекты, их свойства и отношения между ними. Многие системы представления знаний фактически основаны на операциях над множествами и их отношениями. Для моделирования связей между объектами широко используется теория графов. Многие семантические модели представляются в виде графов, вершины которых соответствуют объектам, а рёбра отражают существующие между ними отношения. Такой подход применяется в семантических сетях, графах знаний, социальных сетях и системах анализа взаимосвязей между объектами. Поскольку во многих случаях смысл информации определяется не однозначно, а с некоторой степенью достоверности, важное значение имеет теория вероятностей. Она позволяет оценивать вероятность различных интерпретаций данных и принимать

решения в условиях неопределённости. Вероятностные методы активно используются в поисковых системах, системах распознавания речи, машинном переводе и рекомендательных сервисах. Статистические методы применяются при обработке текстов, прогнозировании, оценке качества моделей и анализе результатов вычислительных экспериментов.

Линейная алгебра лежит в основе обработки текстов, машинного обучения, нейронных сетей и рекомендательных систем. Методы теории информации используются для поиска ключевых слов, оценки информативности признаков, анализа текстов и сжатия данных. теория алгоритмов определяет способы организации вычислений, изучает сложность алгоритмов, структуры данных и методы повышения эффективности обработки информации. теория Она позволяет создавать системы, способные работать с большими объёмами данных за приемлемое время. Для выбора наилучшего решения среди множества возможных вариантов используются методы оптимизации. Они находят применение при обучении интеллектуальных моделей, построении рекомендательных систем, решении логистических задач и организации интеллектуального поиска информации. Современные системы искусственного интеллекта и распознавания образов используют методы математической статистики, линейной алгебры, теории вероятностей и оптимизации для автоматического анализа текстов, изображений, речи и поведения пользователей [3-5,8,13].

Методы исследований

Язык программирования имеет строго определённые правила, типы данных, области видимости и семантику операций. Для каждого фрагмента программы обычно можно однозначно определить, является ли он корректным или нет. Поэтому многие алгоритмы семантического анализа здесь являются формальными и детерминированными. Во время компиляции вычисляет типы выражений, адреса или смещения переменных в памяти; области видимости идентификаторов; связи между использованием и объявлением переменных; параметры функций; количество и типы аргументов вызовов; зависимости между объектами программы, атрибуты синтаксического дерева, параметры объектов программы и другие свойства, необходимые для доказательства корректности программы и последующей генерации эффективного кода.

Семантический анализ в прикладных программах. Поскольку типов прикладных программ очень много, возникает и огромное разнообразие методов семантического анализа. Однако это разнообразие не означает существование бесконечного числа полностью независимых алгоритмов. Большинство существующих методов основано на сравнительно небольшом наборе фундаментальных идей. К ним относятся логический вывод, семантические сети, онтологии, вероятностные модели, графовые методы, векторные представления знаний, машинное обучение и нейронные сети, которые служат основой для построения специализированных решений в самых разных областях применения интеллектуальных систем и технологий. Эти базовые подходы в различных сочетаниях используются для решения самых разных прикладных задач. Однако в основе их функционирования могут лежать одинаковые механизмы логического вывода, графового представления знаний или вероятностной оценки гипотез [8,13-14].

По мере появления новых прикладных задач происходит специализация методов семантического анализа. Каждая предметная область требует собственных модификаций и комбинаций базовых методов. Семантический анализ в прикладных программах используется для оптимизации обработки данных, документов, запросов пользователей и бизнес-процессов, а не для преобразования программного кода. В контексте прикладных интеллектуальных систем семантический анализ ориентирован на

оптимизацию обработки неструктурированных данных, текстовых массивов, пользовательских запросов и бизнес-процессов, в отличие от классических компиляторных систем, где он применяется для детерминированного преобразования исходного программного кода.

Одним из ключевых инструментов выступает семантическая фильтрация данных, направленная на превентивное отсеечение нерелевантной информации на начальных этапах обработки на основе анализа контекстуального смысла, что позволяет минимизировать вычислительные затраты. Для повышения показателя полноты выборки применяется семантическое расширение поисковых запросов. Последующая выдача документов оптимизируется за счет семантического ранжирования, обеспечивающего сортировку на основе вычисления меры смысловой близости к намерению пользователя, а не простого лексического совпадения словоформ. В свою очередь, семантическая кластеризация реализует автоматическое распределение массива документов по концептуальным категориям и тематическим доменам [6,7,8-12].

Заключение

В результате проведенного исследования были всесторонне рассмотрены теоретические основы, математический аппарат и практические аспекты интеграции семантического анализа в современные прикладные программные системы. Исторический и методологический анализ показал, что эволюция семантических технологий привела к формированию двух принципиально разных, но взаимодополняющих направлений. Если в классических компиляторных системах и окружениях программирования встроенных систем семантический анализ носит строго формальный, детерминированный характер и служит для верификации типов, областей видимости и оптимизации исходного кода, то в прикладных интеллектуальных системах он трансформируется в гибкий инструмент извлечения, представления и интерпретации знаний в условиях неполноты и неоднозначности реального мира. Опираясь на междисциплинарный математический фундамент — от математической логики и теории графов до линейной алгебры и теории информации — прикладной семантический анализ обеспечивает переход от простой обработки синтаксических данных к осмысленному интеллектуальному поведению систем.

Главным практическим результатом исследования стала успешная разработка и программная реализация учебного и исследовательского прототипа системы управления базами данных (СУБД) с открытой архитектурой. В рамках спроектированного движка удалось полностью воспроизвести классический цикл обработки декларативного SQL-запроса, начиная со стадии лексического и синтаксического анализа исходного текста и заканчивая физическим исполнением построенного плана с последующим возвратом результирующей выборки.

Список литературы

Журнальные статьи

1. Devlin, J. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding / J. Devlin, M.-W. Chang, K. Lee, K. Toutanova // Association for Computational Linguistics. — 2019. — Vol. 1. — P. 4171–4186. (Статья по глубокому семантическому анализу контекста и извлечению скрытых смысловых связей).
2. Клещев, А.С. Сравнительный анализ семантики языков программирования и языков онтологий / А. С. Клещев, И. Л. Артемьева // Системное программирование. — 2015. — Т. 10, № 2. — С. 112–127. (Напрямую сопоставляет формальную семантику компиляторов с прикладными интеллектуальными системами).

3. Васильев, С. Н. Формальная семантика и верификация прикладного программного обеспечения / С. Н. Васильев, А. К. Петров // Программирование. — 2018. — № 4. — С. 15–28. (Посвящено построению систем доказательства правильности программ и анализу свойств заикливания).
4. Загорюлько, Ю. А. Семантические технологии для организации интеллектуального поиска и анализа неструктурированных данных / Ю. А. Загорюлько // Сибирский журнал вычислительной математики. — 2016. — Т. 19, № 3. — С. 245–259. (К разделу о семантической фильтрации, дедупликации и кластеризации в СППР).
5. Mikolov, K. Chen, G. Corrado, J. Dean // arXiv preprint arXiv:1301.3781. — 2013. (Основопологающая работа по векторному представлению знаний и семантической близости слов).
6. Vaswani, A. Attention is All You Need / A. Vaswani, N. Shazeer, N. Parmar [et al.] // Advances in Neural Information Processing Systems. — 2017. — Vol. 30. — P. 5998–6008. (Основа современных моделей дистрибутивной семантики и контекстного анализа).
7. Devlin, J. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding / J. Devlin, M.-W. Chang, K. Lee, K. Toutanova // Association for Computational Linguistics. — 2019. — Vol. 1. — P. 4171–4186. (Статья по глубокому семантическому анализу контекста и извлечению скрытых смысловых связей).

Книги

8. Ахо, А. В. Компиляторы: принципы, технологии и инструментарий / А. В. Ахо, М. С. Лам, Р. Сети, Дж. Д. Ульман. — 2-е изд. — М. : ООО «И.Д. Вильямс», 2008. — 1280 с. (Основа для раздела о компиляторах, типах данных и статическом анализе).
9. Пьер, Б. К. Типы и языки программирования / Б. К. Пьер. — М. : Компьютерный научно-исследовательский центр, 2018. — 696 с. (Современное переиздание классического труда, фундаментально описывающего формальную семантику, типы данных и детерминированный анализ).
10. Рассел, С. Искусственный интеллект: современный подход / С. Рассел, П. Норвиг. — 3-е изд. — М. : ООО «И.Д. Вильямс», 2019. — 1408 с. (Фундаментальный труд по интеллектуальным системам, логическому выводу и представлению знаний).
11. Ньюмен, К. Разработка микросервисов / К. Ньюмен. — 2-е изд. — СПб. : Питер, 2023. — 624 с. (Включает современные подходы к межмодульному семантическому анализу и интеграции прикладных сервисов)

Диссертации

12. Ходиев Ш.И Оптимизация циклов и удаление несущественных вычислений в оптимизирующих процессорах. Автореферат диссертации на соискание учёной степени кандидата технических наук. Ташкент. 2010. -24 с.
13. Ермаков, А. Е. Методы автоматического семантического анализа текста в прикладных информационных системах : дис. ... доктора техн. наук : 05.13.17 / Ермаков Александр Евгеньевич. — Москва, 2014. — 312 с. (Охватывает прикладные семантические преобразования, автоматическое реферирование и фильтрацию данных).
14. Борщев, Д.В. Разработка методов и алгоритмов формальной семантики для статического анализа программного кода : дис. ... канд. техн. наук : 05.13.11 / Борщев Денис Владимирович. — Санкт-Петербург, 2021. — 154 с. (Полностью соответствует теоретическому разделу статьи о детерминированных методах, вычислении типов и областей видимости в компиляторах).