

УДК 004.822

ЭКСПОРТ СЕМАНТИЧЕСКИ БЛИЗКИХ К ЗАДАННОЙ ПРЕДМЕТНОЙ ОБЛАСТИ УЗЛОВ СЕМАНТИЧЕСКОЙ СЕТИ

Канторов В.С. (Университет ИТМО)

Научный руководитель – старший преподаватель факультета программной инженерии
и компьютерной техники, Цопа Е.А.
(Университет ИТМО)

Введение. Семантическая сеть — структура данных, состоящая из узлов, соответствующих понятиям, и связей, указывающих на взаимосвязи между узлами. При помощи семантической сети появляется возможность оперировать смыслами, а не словами при работе с текстами на естественном языке.

В случае с заранее известной предметной областью текста использование семантической сети, которая описывает более широкую предметную область, невыгодно по причине обработки заведомо не подходящих данных. С целью увеличения эффективности взаимодействия с семантической сетью в рамках конкретной предметной области необходимо создать возможность экспорта подсети, которая описывает заданную предметную область, что и легло в основу представленной работы.

Экспорт необходимой подсети имеет большое практическое значение, поскольку это позволяет уменьшить объем семантической сети, а значит сократить потребление памяти и увеличить скорость обработки запросов.

Основная часть. Первостепенной задачей, решаемой в рамках данной работы, стало определение подхода к определению семантической близости узла к заданной предметной области. На основе проведенного анализа семантической сети как структуры данных было выявлено, что понятие семантической близости является контекстно зависимым. К примеру, для некоторых задач антонимы понятия, входящего в предметную область, являются семантически близкими и должны быть учтены при экспорте. В некоторых же задачах наоборот антонимы понятия не считаются семантически близкими и не должны содержаться в полученной подсети. В связи с этим был сделан вывод о нецелесообразности поиска универсального алгоритма, что вынудило принять решение о создании возможности гибкой настройки механизма определения семантической близости в разрабатываемом модуле для экспорта.

Следующим шагом стала разработка метода определения семантической близости. В процессе выполнения данного этапа был предложен алгоритм, способный определить подсеть, описывающую общую предметную область для заданных узлов и учитывающий заданные настройки механизма определения семантической близости. Вышеупомянутый алгоритм основан на определении кратчайших путей между заданными основными узлами предметной области и использует различные алгоритмы обхода ориентированного графа.

После этого с учетом ранее определенных методов и подходов был реализован программный модуль на языке программирования C++, который был интегрирован в состав существующей системы хранения данных на основе семантической сети и позволил осуществлять экспорт подсети понятий конкретной предметной области. Также был развит язык запросов к семантической сети, добавлением возможности вызывать реализованный модуль. Проведённое функциональное и нагрузочное тестирование доказало пригодность разработанного решения к использованию в производственных системах обработки текста на естественном языке.

Выводы. В результате исследования проведен анализ семантической сети как структуры данных и предложен способ определения семантической близости узла к предметной области, на основе которого разработан и интегрирован модуль для экспорта подсети понятий конкретной предметной области, что увеличило эффективность взаимодействия с

семантической сетью в задачах данной предметной области. Проведённое тестирование доказало возможность использования в реальных системах.

Список использованных источников:

1. John F. Sowa. Semantic Networks. [Электронный ресурс]. Режим доступа: <https://www.jfsowa.com/pubs/semnet.htm> (дата обращения: 10.02.2023).
2. Клименков С.В., Цопа Е.А., Жмылёв С.А., Покид А.В., Ткешелашвили Н.М. Метод быстрого поиска узлов семантической сети по точному совпадению словоформы. Известия высших учебных заведений. Приборостроение. 2017. Т. 60. № 10. С. 932-939. [Тип: Статья, Год: 2017]
3. Письмак А.Е., Харитонов А.Е., Цопа Е.А., Клименков С.В. Метод автоматического формирования семантической сети из слабоструктурированных источников // Программные продукты и системы - 2016. - № 3. - С. 74-78