

ОБОБЩЕНИЕ УЗЛОВ ЧИСЛИТЕЛЬНЫХ В СЕМАНТИЧЕСКОЙ СЕТИ

Чупанов А.А. (Университет ИТМО)

Научный руководитель – старший преподаватель факультета программной инженерии
и компьютерной техники Цопа Е.А.

(Университет ИТМО)

Введение. Семантические сети — это один из наиболее распространенных способов представления знаний в искусственном интеллекте и компьютерных науках. Они представляют собой графические модели, в которых узлы соответствуют понятиям, а дуги определяют отношения между этими понятиями. Одной из основных проблем, связанных с построением семантических сетей, является обобщение узлов, которое позволяет упрощать структуру сети и повышать ее обобщающую способность. В данной работе рассмотрена проблема обобщения узлов числительных в семантических сетях.

Основная часть. Одной из ключевых проблем, с которой можно столкнуться при обработке текста на естественном языке с использованием семантической сети – это сложности с обработкой числительных, которые могут принимать различные формы в зависимости от контекста.

Числительные являются важной частью естественного языка и используются для обозначения количественных значений. Числительные могут быть конкретными (например, "один", "два", "три"), или абстрактными (например, "много", "несколько", "немного"), или например, словоформы "два" и "двое" могут означать одно и то же число, но при этом могут относиться к разным узлам семантической сети, что затрудняет работу с данными и усложняет их анализ. Таким образом, возникает задача “обобщения” числительных в структуре семантической сети. Успешное решение этой задачи позволит лучше понимать, как эти значения связаны друг с другом и как они могут быть использованы в различных контекстах.

На первом этапе исследования были проанализированы типичные сценарии использования числительных в тексте и возможные подходы к отображению словоформ числительных на узлы семантической сети (смыслы). В результате была выявлена ключевая проблема в механике такого отображения: так как множество чисел бесконечно, решение задачи путем непосредственного использования отдельных узлов под каждую словоформу приводит к созданию семантической сети, “перегруженной” узлами-числительными, что приводит как к неэффективности работы с числительными, так и к снижению производительности семантической сети в целом.

На следующем этапе были проанализированы пути решения выявленной проблемы. В результате был предложен подход, позволяющий эффективно обобщать несколько числительных в рамках одного узла семантической сети без потери их идентичности.

После этого был разработан программный модуль на языке программирования C++, реализующий предложенный механизм обобщения числительных на практике. Реализованный модуль был интегрирован в состав семантической сети и в настоящий момент показал свою эффективность в решении практических задач автоматической обработки текста на естественном языке.

Выводы. В ходе исследования были проанализированы особенности использования числительных в текстах на естественном языке и рассмотрены подходы к представлению числительных в семантической сети. По итогам исследования была сформулирована задача обобщения узлов числительных в структуре семантической сети, после чего был предложен подход к решению поставленной задачи и разработан программный модуль, интегрированный в состав семантической сети.

Список использованных источников:

1. Шульга М. В., Динамика форм имени числительного в современном русском языке // Вестник ВолГУ. Серия 2: Языкознание. 2019. №4.
2. Агаев Мугадин КОЛИЧЕСТВЕННО-ИМЕННЫЕ СВЯЗАННЫЕ СЛОВСОЧЕТАНИЯ // Евразийский Союз Ученых. 2020. №9-4 (78).
3. Белякова С. М. О СЕМАНТИЧЕСКОМ ПОТЕНЦИАЛЕ ПОРЯДКОВЫХ ЧИСЛИТЕЛЬНЫХ В РУССКОМ ЯЗЫКЕ // Вестник ВГУ. Серия: Лингвистика и межкультурная коммуникация. 2020. №4.
4. Гринченко Н. Н., Цыбулько Дмитрий Михайлович МЕТОД ОБРАБОТКИ КОЛИЧЕСТВЕННЫХ ЧИСЛИТЕЛЬНЫХ ЕСТЕСТВЕННЫХ ЯЗЫКОВ И ЕГО ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ // Известия ТулГУ. Технические науки. 2014. №9-2.