

Секция «Вычислительная математика и кибернетика»

Исследование и программная реализация алгоритма кластеризации в терминологической сети

Новиков Константин Викторович

Студент

Московский государственный университет имени М.В. Ломоносова, Факультет вычислительной математики и кибернетики, Москва, Россия

E-mail: novkostya+ctc_msu@gmail.com

За всю историю существования человечество, начиная с самых ранних времен, накапливало свои знания. Для умелого обращения с накопленной информацией необходимо разумно представлять, структурировать, хранить полученные знания, поэтому важно сопоставлять различные методы представления знаний и по заданным критериям определять наиболее полезные. Правильно выбранный метод представления знаний позволяет значительно упростить сложные задачи, практически не поддающиеся решению при иных подходах. Решение таких задач и включает в себя проблема наиболее корректной организации и хранения знаний. Для хранения и представления знаний используются глоссарии, словари, тезаурусы, энциклопедии, семантические сети. Глоссарии представляют собой словари узкоспециализированных терминов конкретной рассматриваемой предметной области знаний с пояснением смысла, переводом на другой язык, комментариями и примерами. Глоссарии представляют особый интерес, так как познавательная сила глоссария состоит в том, что расположенные рядом статьи являются «близкими родственниками»; статьи глоссария человек может читать подряд, не выходя за пределы интересующей его темы [3]. Семантические сети возникли для представления информации для баз знаний искусственного интеллекта. По определению универсальное терминологическое пространство (УТП) есть абстрактное хранилище всех терминологических статей, связанных семантическими отношениями [1]. Фактически УТП — семантическая сеть на множестве абсолютно всех терминов, выработанных человечеством. Одной из наиболее актуальных задач формирования УТП является задача кластеризации понятий и формирования структуры связей между кластерами так, чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров существенно отличались, и при этом связь осуществлялась наиболее эффективно. Решение этой задачи позволит пользователю удобно обозревать и выбирать интересующие его проблемные области [2]. Рассматриваемая задача заключается в построении алгоритма, реализация которого позволит конечному пользователю обозревать проблемные области и выбирать необходимые, являющиеся наиболее интересными для него в текущий момент времени. Необходимо разрешить проблему организации эффективного и удобного использования определенной семантической сети для конечного пользователя. Исследована проблема реализации скорейшего попадания пользователя к интересующему его фрагменту сети. Существуют различные алгоритмы разбиения графов на кластеры, на основе этих алгоритмов предложен метод кластеризации терминологической сети с возможностью корректировки редактором. Разработан алгоритм кластеризации понятий и формирования структуры связей между кластерами для разрешения задачи кратчайшего «вывода» пользователя на нужный ему фрагмент сети. Предложенный алгоритм реализован в виде программы, обеспечивающей кластеризацию понятий

и формирование структуры связей между кластерами. Таким образом, пользователь легко может выбрать интересующие его проблемные области среди всех удобно обозреваемых разработанной программой, и получить к ним скорейший доступ. Программа применяется для работы с универсальным терминологическим пространством. На основе полученных результатов сделан вывод об эффективности применения подобного подхода к решению задачи кластеризации понятий и формирования структуры связей между кластерами. Программа может быть использована для работы с любыми семантическими сетями с бинарными отношениями, что будет безусловно полезно с точки зрения организации доступа к запрашиваемой информации.

Литература

1. Мальковский М.Г., Соловьев С.Ю. Универсальное терминологическое пространство. Труды международного семинара: Компьютерная лингвистика и интеллектуальные технологии, М: Наука, 2002. Т.1, 266-270.
2. Мальковский М.Г., Соловьев С.Ю. Технология формирования универсального терминологического пространства. Сб. Информационные компьютерные технологии и Интернет в образовании и науке. Москва: изд-во МИИ для инвалидов с нарушением ОДС, 2002, 54-55.
3. Глоссарий.py: <http://glossary.ru>