

ОТЗЫВ

Научного руководителя

На выпускную квалификационную работу бакалавра на тему «Выявление типов объектов в графовой базе данных на основе кластеризации», представленную Соковицкой Светланой Алексеевной.

Тема работы Соковицкой С.А. относится к направлению, связанному с выявлением структуры в слабоструктурированных данных и может быть отнесена к широкому классу задач, объединяемых термином «извлечение знаний», в данном случае – представленной структурой. Более конкретно, в работе решается задача выявления типов объектов, представленных в нетипизированной слабоструктурированной форме в тернарном представлении графа.

Выделение типов объектов обеспечивает размещение данных в СУБД с учетом этой структуры, что может радикально ускорить выполнение запросов к такому набору данных.

В работе Соковицкой реализован инструмент для выделения структуры из набора нетипизированных слабоструктурированных данных. Основным применением этого инструмента могут быть данные в формате RDF или XML без схемы. Выделение типов реализовано применением алгоритма кластеризации к представлению данных в виде набора битовых шкал, представляющих наличие или отсутствие поименованных атрибутов объекта.

Работоспособность этого метода и разработанного инструмента проверялась на относительно большом наборе данных (библиографической коллекции DBLP), содержащей небольшое количество известных типов, однако эта информация использовалась только для оценки качества результата выделения типов.

Работа над минимальным вариантом проекта была успешно завершена, однако не удалось провести экспериментальное сравнение с альтернативными методами, в частности, использующие другие алгоритмы кластеризации. Не были также проведены эксперименты по сравнению производительности при выполнении поисковых запросов. Изложение материала в тексте работы отличается лаконичностью, некоторое незрелостью и содержит значительное количество неточных утверждений.

В связи с вышеизложенным считаю, что работа заслуживает оценку «хорошо».

Научный руководитель



Б.А. Новиков

Д.ф.-м.н., проф.

Проф. Кафедры информационно-аналитических систем СПбГУ