

АЛГОРИТМ АГРЕГАЦИИ ДАННЫХ НА ГРАФАХ

Халмакшинов Е.А.¹

¹) Восточно-Сибирский государственный университет технологий и управления,
Улан-Удэ, Россия
E-mail: poeobot3r@gmail.com

ALGORITHM FOR DATA AGGREGATION ON GRAPHS

Khalmakshinov E.A.¹

¹) East Siberian state university of technology and management, Ulan-Ude, Russia

This paper proposes an algorithm for aggregation of graph data, based on the results of analysis of other algorithms, such as: ranking, shortest path search and clustering. The developed algorithm provides a facilitation of data analysis process on graphs with attribute composition.

В настоящее время количество электронных данных растёт всё более быстрыми темпами, как по размеру, так и по возможности подключения. Данный рост способствует увеличению интереса к взаимосвязям между данными, тем самым повышая спрос на технологии, которые могут обрабатывать подобные данные [1].

Перспективность исследования данных на графах обусловлена ростом количества информации в Интернете, пополняемой пользователями сети, а это, тем самым приводит к её постепенному превращению в большую базу знаний, данные которой легче всего представить в виде взаимосвязанных блоков информации.

Основная проблема исследования заключается в визуальном представлении результатов анализа графовых данных, так как при отображении графа возникают такие проблемы как:

1. Наложение сущностей и связей. Это приводит к проблеме анализа результатов, отображаемых на графе, так как появляются сложности идентификации сущности.

2. Фильтрация по атрибутному составу сущности. Скорость выполнения вычислений на графах существенно замедляется, если включать в запрос на получение данных атрибуты сущности.

Решением данных проблем является достижение структурированного вида данных, путём выполнения распределенного агрегирования результатов анализа данных графа.

Основной процесс обработки графов изображен на рисунке 1.

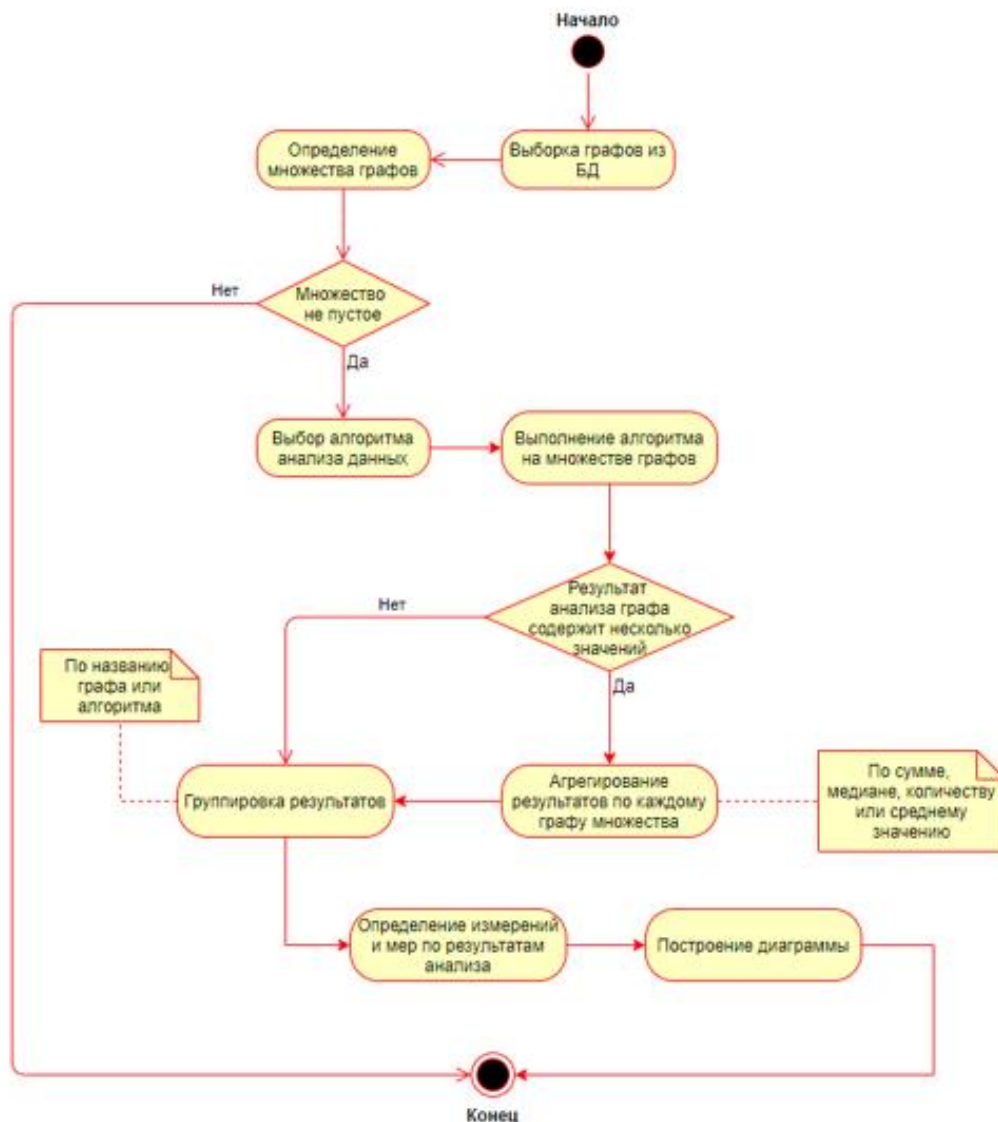


Рис. 1. Диаграмма действий алгоритма агрегации данных графа

Описание шагов выполнения:

1. Выборка графов представляет собой получение графа из базы данных.
2. Подразумевается, что в графе базы данных, полученном на первом шаге, может содержаться некоторое множество однотипных графов, то есть которые относятся к одному классу.
3. Граф базы данных может быть пустым, поэтому необходима проверка на наличие вершин, если их нет, то алгоритм завершает работу.
4. Выбор алгоритма анализа данных является одним из самых важных этапов, так как именно по его результатам воспроизводится дальнейший процесс.
5. Выбранный в предыдущем шаге алгоритм должен обрабатывать на каждом из графов выборки (пункт 2).
6. Если результат анализа на одном из графов содержит несколько значений необходимо выполнить операцию агрегирования, которая включает в себя суммирование, среднее, количество значений, а также поиск медианы. В ином случае

сразу выполняется группировка (объединение) результатов анализа по каждому графу в одно множество по названию графа или алгоритма.

7. Определение измерений и мер производится для представления значений множества из пункта 6 в виде OLAP – куба, который является структурой диаграммы.

8. Построение диаграммы производится по результатам из пункта 7.

Разработанный алгоритм предоставляет наибольшую гибкость для анализа данных, исключая проблемы наложения вершин и связей, а также отображения атрибутного состава сущностей в атрибутно-ориентированных графах.

1. Линник Е.В. Графовая аналитика для решения ключевых проблем в банковской сфере [Текст] / Молодой ученый. – 2018. – №52 (238). – С. 128-134.