

## ОТЗЫВ

на автореферат диссертации Казаковцева Владимира Львовича  
«Алгоритмы ускоренного поиска в векторных базах данных»,  
представленной на соискание ученой степени кандидата технических  
наук по специальности 2.3.1 – Системный анализ, управление и  
обработка информации, статистика

Приближенный поиск ближайших соседей (Approximate Nearest Neighbor search, ANN) - это оптимизированный подход для нахождения «достаточно близких» точек к запросу в высокоразмерных пространствах, когда точный поиск становится непрактичным из-за больших объемов данных. Вместо гарантии точного нахождения ближайшего соседа алгоритм ANN достигает значительного ускорения, жертвуя точностью ради производительности. В многомерных пространствах традиционные структуры данных и некоторые алгоритмы, такие как алгоритм Элкана, становятся неэффективными. Это явление известно как "проклятие размерности", что делает поиск точных решений вычислительно сложным и неприменимым на практике.

Как следует из автореферата, в диссертации разрабатываются новые алгоритмы приближенного поиска ближайших соседей. Актуальность таких исследований связана с ростом объема и размерности данных, обрабатываемых в современных информационных системах. Эта проблема четко обозначена автором в автореферате. Актуальность исследований не вызывает сомнений.

Теоретическая и практическая значимость результатов работы связана с новым адаптивным алгоритмом поиска ближайших соседей в векторной базе данных на основе IVF-индекса. Этот алгоритм обеспечивает ускорение выполнения запросов на 10–30% на наборах данных до 1 миллиарда объектов. Такие алгоритмы вносят вклад в развитие методов поиска в векторных базах данных, а их практическая применимость подтверждается возможностью использования в системах управления векторными базами данных, работающими с сотнями миллионов объектов. Поддержка исследований грантами и хозяйственным договором с технологической компанией свидетельствует о востребованности этой темы исследований.

Работа не лишена недостатков. Из замечаний к автореферату можно выделить следующие:

1. Принцип работы ускоренной агломеративной процедуры (алгоритм 5) строится на предположении о зависимости прироста

целевой функции от размерности кластера, определяемого удаляемым центроидом. Это предположение не подтверждается в автореферате теоретическими выкладками.

2. На стр. 18 для формулы расчета Score Function – меры качества кластеризации, не приводятся определения переменных и параметров, что затрудняет понимание.

3. В Алгоритме 2 из второй главы выбор параметра  $n_{res}$ , как единственного параметра определения сложности запроса, требует более четкого обоснования. В Алгоритме 4 из третьей главы появляются «магические числа» 1.5, 1.25, 0.1 и др.

4. Не приводится явного сравнения со state-of-the-art алгоритмами ANN поиска.

Несмотря на перечисленные замечания, считаю, что работа соответствует всем требованиям, предъявляемым ВАК РФ к кандидатским диссертациям по специальности 2.3.1, а ее автор, Казаковцев В.Л., заслуживает присуждения ему ученой степени кандидата технических наук по данной специальности.

Доктор физико-математических наук,  
профессор, главный научный  
сотрудник Института математики им.  
С.Л. Соболева СО РАН

Л

Кочетов Юрий  
Андреевич

Адрес организации:  
Просп. Академика Коптюга, 4.  
г. Новосибирск, 630090, Россия  
e-mail: jkochet@math.nsc.ru  
Телефон: +7(383)3297583

Подпись Кочетова Ю.А. заверяю.

*Рук. отдела ДОУ ИИИ СО РАН*

27.04.2026

