

Методы организации хранения иерархически структурированных данных в документоориентированном хранилище с поддержкой полнотекстового поиска

А. М. Неганов

Московский физико-технический институт (государственный университет)

Специализированные документоориентированные хранилища представляют собой мощное средство для организации информационного поиска по слабоструктурированным данным.

Одной из важных проблем, которая может решаться таким образом, является задача хранить иерархически связанные данные при условии, что в качестве основного сценария использования хранилища выступает полнотекстовый поиск.

Работа посвящена исследованию способов хранения иерархически структурированных данных и организации поисковых запросов на примере системы информационного поиска Elasticsearch. Целью исследования было определение наиболее производительных и хорошо масштабируемых при увеличении объёмов данных и нагрузки методов, для чего производился анализ размера поискового индекса и скорости выполнения запросов.

Литература

1. *Han J., E H., Le G., Du J.* Survey on NoSQL database. // Proceedings of 6th International Conference on Pervasive Computing and Applications (ICPCA), 2011
2. *Sierra J., Fernandez-Valmayor A., Fernandez-Manjon B.* A Document-Oriented Paradigm for the Construction of Content-Intensive Applications // The Computer Journal. 2006. V. 49, N. 5. P. 562-584.
3. *Gormley C., Tong Z.* Elasticsearch: The Definitive Guide. Sebastopol: O'Reilly Media, 2015. 719 с.