

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
СУМСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ

ІНФОРМАТИКА, МАТЕМАТИКА,
АВТОМАТИКА

ІМА :: 2017

**МАТЕРІАЛИ
та програма**

НАУКОВО-ТЕХНІЧНОЇ КОНФЕРЕНЦІЇ

(Суми, 17–21 квітня 2017 року)



Суми
Сумський державний університет
2017

Применение технологии NoSQL для хранения и обработки библиографической информации

Блинков Ю.А., *заведующий кафедрой*; Панкратов И.А., *магистрант*
Саратовский национальный исследовательский государственный университет имени Н.Г. Чернышевского, г. Саратов, Россия

В настоящей работе построена концепция информационной системы для документо-ориентированного хранения и обработки информации публикациях различного типа с помощью технологии NoSQL [1]. При использовании NoSQL не требуется создавать несколько связанных друг с другом таблиц. Применение технологии NoSQL позволяет отказаться от жёсткой структуры базы данных. Легко можно будет при необходимости добавить новый, пока ещё отсутствующий тип публикации, не затрагивая ранее внесённые записи.

Для заполнения указанной базы данных был написан «паук» (web crawler) на языке Python с использованием свободно распространяемого фреймворка Scrapy. Информация о публикациях извлекается с сайта национальной библиографической базы данных научного цитирования (РИНЦ, <http://elibrary.ru>). Созданное программное обеспечение позволяет получить, например, список публикаций того или иного автора, зарегистрированного в РИНЦ. При этом в качестве формата выходных данных был выбран широко известный JavaScript Object Notation (JSON), удобный для чтения человеком и компьютером. JSON-файл, содержащий библиографическую информацию, легко импортируется в документо-ориентированную базу данных MongoDB [2]. Для работы с MongoDB был использован модуль pymongo языка программирования Python, позволяющий не только загружать информацию в базу данных, но и писать к ней различные запросы.

1. Э. Редмонд, Д.Р. Уилсон, *Семь баз данных за семь недель. Введение в современные базы данных и идеологию NoSQL* (Москва: ДМК Пресс: 2013).
2. К. Бэнкер, *MongoDB в действии* (Москва: ДМК Пресс: 2012).