IBIMA Publishing
Communications of International Proceedings
https://ibimapublishing.com/p-articles/40AI/2022/4058722/
Vol. 2022 (30), Article ID 4058722

The Architecture Concepts for Building Highly Scalable Crawling Cluster For Data-Driven On-Page Optimization*

Krystian MAGDZIARZ and Damian FRĄSZCZAK

Military University of Technology, Warsaw, Poland

Correspondence should be addressed to: Krystian MAGDZIARZ; krystianmagdziarz@gmail.com

* Presented at the 40th IBIMA International Conference, 23-24 November 2022, Seville, Spain

Abstract

The paper presents the architecture concepts for building crawling clusters for data-driven on-page optimization tasks. The aim of the study is to develop a base architecture capable of continuous monitoring of on-page parameters in terms of SEO on a very large scale. The issue of building an efficient data crawling mechanism has been addressed in the literature since the spread of the Internet, however, the problem is not thoroughly described in relation to tools designed for SEO. This paper offers the concept of building a highly scalable environment designed to analyze key on-page metrics, the analysis of which will provide critical knowledge in the context of their optimization.

Keywords: crawling, spider cluster, content extraction, SEO, crawling cluster architecture

Cite this Article as: Krystian MAGDZIARZ and Damian FRĄSZCZAK, Vol. 2022 (30) " The Architecture Concepts for Building Highly Scalable Crawling Cluster For Data-Driven On-Page Optimization " Communications of International Proceedings, Vol. 2022 (30), Article ID 4058722.