

Validasi Kelengkapan Knowledge Graph (KG) Menggunakan SHACL = Validating Knowledge Graph Completeness Using SHACL

Muhammad Jilham Luthfi, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20529239&lokasi=lokal>

Abstrak

Perkembangan knowledge graph serta penggunaannya meningkat secara masif akhir-akhir ini. Namun, pertumbuhan knowledge graph nyatanya tidak sebanding dengan kualitas data yang dimiliki. Sementara itu, aspek kelengkapan dalam kualitas data knowledge graph memiliki peranan penting dalam pemakaian data yang sesuai dengan kebutuhan. Sayangnya, hingga saat ini belum ada penelitian tentang penggunaan SHACL dalam melakukan validasi kelengkapan data. Penelitian ini hadir dengan tujuan untuk membuat pola-pola kelengkapan berdasarkan kondisi knowledge graph beserta instance-nya. Penelitian ini mengidentifikasi enam pola kelengkapan yang sering muncul pada knowledge graph. Dari keenam pola kelengkapan tersebut, penelitian ini berhasil membuat 360.162 instance pola kelengkapan dalam bentuk SHACL shape. Pembuatan instance pola kelengkapan menggunakan lima pendekatan, yakni pendekatan manual, otomatis, spreadsheet, ontologi, dan statistik. Semua shape tersebut kemudian digunakan untuk melakukan validasi terhadap 928.310 entitas pada knowledge graph dunia nyata, yakni Wikidata dan DBpedia. Hasil validasi population, no-value, dan label and description completeness pada Wikidata dan DBpedia mencapai skor lebih dari 90%. Sementara itu, hasil validasi property dan interlinking completeness pada Wikidata dan DBpedia berada di rentang skor 60-70%. Terakhir, validasi schema completeness pada Wikidata dan DBpedia mencapai hasil 70-90%. Selain itu, penelitian ini menyediakan aplikasi web bernama SHACL on Completeness Knowledge (SoCK) untuk mengelola dan menyediakan pola kelengkapan beserta instance-nya.

.....The development of knowledge graphs and their use has increased massively recently. In fact, the growth of the knowledge graph is not proportional to the quality of the data it has. Meanwhile, the completeness aspect in the quality of the knowledge graph data has an essential role in the use of data according to needs. Unfortunately, nobody has previously done research on the use of SHACL in validating data completeness. Therefore, this research focused on identifying completeness patterns based on the condition of the knowledge graph and its instances. This study identified six completeness patterns that often appear in knowledge graphs. Of the six completeness patterns, this study successfully created 360.162 completeness pattern instances in the form of SHACL shape. The instantiation of the completeness pattern used five approaches, that is, manual, automated, spreadsheet, ontology, and statistical approaches. All these shapes are used to validate 928.310 entities in a real-world knowledge graph, namely Wikidata and DBpedia. Based on the population, no-value, and label and description completeness validation, the completeness of Wikidata and DBpedia reached a score of more than 90%. Moreover, the validation results of property and interlinking completeness stated that the Wikidata and DBpedia completeness are 60-70%. Finally, schema completeness validation on Wikidata and DBpedia reached 70-90%. In addition, this research provided a web application called SHACL on Completeness Knowledge (SoCK) to manage and accommodate completeness patterns and their instances.