

## 1. Pendahuluan

### 1.1 Latar Belakang

*Knowledge graph* adalah representasi dari informasi relasional terstruktur dalam bentuk Entitas dan hubungan di antara keduanya. Graf ini adalah graf multi rasional yang mana entitas dan hubungan saling terhubung. Entitas adalah objek nyata atau informasi abstrak. Entitas dan hubungan direpresentasikan sebagai *triple*[1]. Sebagai contoh bentuk data triple adalah (“Bandung”, “ibukota”, “Jawa Barat”). “Bandung” dan “Jawa Barat” adalah entitas yang terhubung oleh hubungan yang bernama “ibukota”.

Pada tahun 2012, Google memperkenalkan istilah *Knowledge graph* pada mesin pencari mereka sebagai peningkatan semantik dari mesin pencari tersebut. Hal ini membuat mesin pencari Google dapat mencari sesuatu yang lebih berorientasi dengan objek di dunia nyata (*real-world object*). Sejak 2012, istilah *Knowledge graph* juga digunakan untuk aplikasi serupa. Beberapa diantaranya seperti DBPedia, YAGO (Yet Another Great Ontology), Freebase, Wikidata[4]. *Knowledge graph* bisa dibangun dengan Natural Language Programming (NLP) [6][7] dan modul dari graph database. Terdapat beberapa framework graph database seperti Ontotext GraphDB, Neo4j, Cayley, dan yang lainnya.

Sebuah *knowledge graph* yang telah dibangun biasanya masih belum sempurna. Terdapat beberapa informasi yang masih belum lengkap dan masih bisa untuk ditambahkan. Maka hal tersebut bisa diatasi dengan *Knowledge graph completion*, sebuah metode untuk menemukan hubungan yang hilang atau salah dengan tujuan untuk meningkatkan kualitas sebuah *knowledge graph* [10]. Dengan metode tersebut diharapkan *knowledge graph* memiliki lebih banyak informasi didalamnya.

Selain digunakan untuk merepresentasikan dan mengelola pengetahuan umum dan ensiklopedia, *knowledge graph* juga dapat digunakan untuk mengelola data berdomain scholar atau bibliografi yaitu scholarly *knowledge graph* [8]. Data *scholar* tersebut didapat dari makalah-makalah ilmiah. Data yang bisa kita temukan di dalamnya seperti pengarang makalah ilmiah, insitusi asal pengarang, nama jurnal/konferensi, dan topik penelitian [9]. Penggunaan *knowledge graph* untuk mengelola data berdomain *scholar* juga masih merupakan sebuah pendekatan yang baru. [8].

Data berdomain *scholar* di Telkom University telah didata dengan baik oleh Direktorat Penelitian dan Pengabdian Masyarakat (PPM). Data tersebut berisi judul makalah ilmiah, penulis makalah ilmiah, dan kelompok keahlian dari dosen tersebut. Data kelompok keahlian menggambarkan fokus bidang penelitian dari tiap-tiap dosen dalam melakukan penelitiannya. Namun terdapat kekurangan pada data, walaupun bidang penelitian dosen digambarkan dari kelompok keahliannya, masih terdapat dosen yang melakukan penelitian tidak sesuai dengan kelompok keahlian dosen tersebut. Sehingga masih sulit untuk menentukan fokus bidang penelitian dosen yang sebenarnya. Di sisi lain, Google Scholar yang merupakan salah satu layanan yang paling lengkap terkait data peneliti-peneliti dan bibliografi di internet juga belum dapat untuk mengategorisasikan bidang penelitian dari peneliti dan fokus bidang penelitian yang diterapkan dari perguruan tinggi masing-masingnya. Kondisi tersebut membuat PPM sulit untuk mengidentifikasi bidang penelitian dosen yang sesuai untuk diikuti dalam sebuah

skema atau hibah penelitian. Oleh karena itu, penelitian ini bertujuan membantu PPM dalam mengatasi masalah tersebut dengan cara menerapkan *knowledge graph completion* pada *scholarly knowledge graph* nya Telkom University.

## **1.2 Perumusan dan Batasan Masalah**

Rumusan masalah pada penelitian ini yaitu bagaimana melengkapi *missing parts* (informasi yang belum lengkap) pada sebuah *scholarly knowledge graph* dengan menggunakan *knowledge graph completion* dan menghilangkan data null yang terdapat di dalamnya.

Pada penelitian ini terdapat beberapa batasan masalah. Pertama, dataset yang digunakan adalah data publikasi di lingkup Prodi Informatika Telkom University tahun 2018. Kedua, data yang dilengkapi ke *scholarly knowledge graph* hanya makalah lain dan *research field* dari penulis 1 serta deskripsi dari sebuah konferensi atau jurnal.

## **1.3 Tujuan**

Tujuan dari penelitian ini adalah untuk melengkapi *missing parts* (informasi yang belum lengkap) pada sebuah *scholarly knowledge graph*. Sebagaimana diketahui pada data domain scholar terdapat banyak data null, maka untuk menghilangkan data null tersebut digunakan sebuah framework khusus graph yaitu GP2.