TESIS

PERANGKINGAN ENTITAS LINKED OPEN DATA MENGGUNAKAN PENDEKATAN RESOLUSI ENTITAS UNTUK PENINGKATAN RELEVANSI PENCARIAN ENTITAS PADA DATASET HALAL NUTRITION FOOD

ENTITY RESOLUTION APPROACH FOR ENTITY RANKING ON LINKED OPEN DATA: A CASE STUDY ON HALAL NUTRITION FOOD

AHMAD CHOIRUN NAJIB
NRP 05211850010006

Dosen Pembimbing:
Nur Aini Rakhmawati, S.Kom., M.Sc.Eng., Ph.D.
NIP 198201302005012001

PROGRAM MAGISTER
DEPARTEMEN SISTEM INFORMASI
FAKULTAS TEKNOLOGI ELEKTRO DAN INFORMATIKA CERDAS
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2020

Halaman ini sengaja dikosongkan

TESIS

PERANGKINGAN ENTITAS LINKED OPEN DATA MENGGUNAKAN PENDEKATAN RESOLUSI ENTITAS UNTUK PENINGKATAN RELEVANSI PENCARIAN ENTITAS PADA DATASET HALAL NUTRITION FOOD

AHMAD CHOIRUN NAJIB
NRP 05211850010006

Dosen Pembimbing:
Nur Aini Rakhmawati, S.Kom., M.Sc.Eng., Ph.D.
NIP 198201302005012001

PROGRAM MAGISTER
DEPARTEMEN SISTEM INFORMASI
FAKULTAS TEKNOLOGI ELEKTRO DAN INFORMATIKA CERDAS
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2020

Halaman ini sengaja dikosongkan

MASTER THESIS - IS185401

ENTITY RESOLUTION APPROACH FOR ENTITY RANKING ON LINKED OPEN DATA: A CASE STUDY ON HALAL NUTRITION FOOD

AHMAD CHOIRUN NAJIB
NRP 05211850010006

Supervisor:
Nur Aini Rakhmawati, S.Kom., M.Sc.Eng., Ph.D.
NIP 198201302005012001

MASTER PROGRAM
DEPARTMENT OF INFORMATION SYSTEM
FACULTY OF INTELLIGENT ELECTRICAL AND INFORMATICS TECHNOLOGY
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2020

Halaman ini sengaja dikosongkan

LEMBAR PENGESAHAN

PERANGKINGAN ENTITAS LINKED OPEN DATA MENGGUNAKAN PENDEKATAN RESOLUSI ENTITAS UNTUK PENINGKATAN RELEVANSI PENCARIAN ENTITAS PADA DATASET HALAL NUTRITION FOOD

TESIS

Diajukan Guna Memenuhi Salah Satu Syarat
Memperoleh Gelar Magister Komputer

pada

Bidang Studi Akuisisi Data dan Diseminasi Informasi
Program Studi S2 Departemen Sistem Informasi
Fakultas Teknologi Elektro dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember

Oleh :

AHMAD CHOIRUN NAJIB
NRP: 05211850010006

Surabaya, Maret 2020

KEPALA
DEPARTEMEN SISTEM INFORMASI

Dr. Mudjahidin, ST, MT.
NIP. 19701010 200312 1 001

Halaman ini sengaja dikosongkan

LEMBAR PERSETUJUAN

PERANGKINGAN ENTITAS LINKED OPEN DATA MENGGUNAKAN PENDEKATAN RESOLUSI ENTITAS UNTUK PENINGKATAN RELEVANSI PENCARIAN ENTITAS PADA DATASET HALAL NUTRITION FOOD

TESIS

Diajukan Guna Memenuhi Salah Satu Syarat
Memperoleh Gelar Magister Komputer
pada
Bidang Studi Akuisisi Data dan Diseminasi Informasi
Program Studi S2 Departemen Sistem Informasi
Fakultas Teknologi Elektro dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember

Oleh :

AHMAD CHOIRUN NAJIB
NRP: 05211850010006

Disetujui Tim Penguji: Tanggal Ujian: 13 Januari 2019
Periode Wisuda: Maret 2020

Nur Aini Rakhmawati, S.Kom., M.Sc.Eng., Ph.D. **(Pembimbing 1)**

Dr. Apol Pribadi S., S.T, M.T **(Penguji 1)**

Dr. Eng. Febriliyan Samopa, S.Kom, M.Kom **(Penguji 2)**

Halaman ini sengaja dikosongkan

PERANGKINGAN ENTITAS LINKED OPEN DATA MENGGUNAKAN PENDEKATAN RESOLUSI ENTITAS UNTUK PENINGKATAN RELEVANSI PENCARIAN ENTITAS PADA DATASET HALAL NUTRITION FOOD

Nama Mahasiswa : Ahmad Choirun Najib

NRP : 05211850010006

Pembimbing I : Nur Aini Rakhmawati, S.Kom., M.Sc.Eng., Ph.D.

ABSTRAK

Banyaknya jumlah keragaman data memberikan peran penting dalam menyajikan informasi secara komprehensif dari sebuah entitas agar dapat memberikan informasi kepada pengguna dengan sebaik mungkin. Halal Nutrition Food http://halal.addi.is.its.ac.id merupakan produk riset berupa aplikasi yang menyediakan informasi produk makanan halal yang telah diintegrasikan dengan dataset 10 lembaga halal dunia dan OpenFoodFacts yang merupakan situs penyedia informasi produk makanan dengan platform crowdsourcing. Dampak integrasi data yang telah dilakukan mengakibatkan duplikasi entitas dan menurunkan relevansi pada pencarian entitas. Sehingga penguna harus melakukan pengecekan berulang pada setiap entitas yang muncul pada hasil pencarian untuk mendapatkan informasi yang dicari. Pada tesis ini melakukan peningkatan relevansi pencarian dengan pendekatan Entity Resolution (ER). Pendekatan ER dilakukan dengan penggunaan graph embedding *Node2vec* yang melakukan transformasi setiap node menjadi luaran berupa sekumpulan vektor yang representatif. Luaran ini digunakan untuk mengetahui pasangan antar node yang memiliki kesamaan dengan node sumber. Pasangan antar node ini akan dijadikan sebagai kandidat untuk dilakukan resolusi dan prediksi link. Hasil luaran akan digunakan sebagai input dalam proses perangkingan berupa PageRank dan LinkCount. Proses perangkingan dilakukan dengan pendekatan query-independent (QI) dan query-dependent (QD). QI dihitung menggunakan algoritma PageRank dan LinkCount. Sedangkan QD dihitung menggunakan pembobotan Term Frequency Inverse Entity Frequency (TF-IEF) yang menghitung kemiripan dokumen berdasarkan kata kunci tertentu, kemudian fungsi perangkingan dihitung menggunakan BM25. Kombinasi perhitungan QI dan QD akan menghasilkan skor final yang akan dijadikan bobot acuan dalam perangkingan entitas untuk meningkatkan relevansi pencarian. Hasil pada tesis ini menunjukkan bahwa faktor yang paling dominan yaitu nilai skor dependen, kemudian diikuti oleh skor independen. Semakin besar nilai skor dependen maka semakin besar peluang muncul sebuah dokumen untuk ditampilkan pada sistem dan menjadi hasil pencarian yang relevan. Semakin besar nilai skor independen akan mempengaruhi peringkat rangking dari dokumen pada hasil pencarian. Penggunaan ER untuk perangkingan memiliki peran yang penting dalam melakukan pembobotan rangking pada setiap dokumen melalui relasi-relasi yang terbentuk (owl:sameAs dan rdfs:seeAlso) pada

proses *graph embedding* melalui similarity pada hasil embedding dan koten dokumen. Hasil pendekatan ER terbukti dapat meningkatkan relevansi pencarian entitas pada linked open data.

Kata kunci:Rangking entitas, resolusi entitas, linked open data, rangking produk, produk makanan, halal

ENTITY RESOLUTION APPROACH FOR ENTITY RANKING ON LINKED OPEN DATA: A CASE STUDY ON HALAL NUTRITION FOOD

Name : Ahmad Choirun Najib

NRP : 05211850010006

Supervisor I : Nur Aini Rakhmawati, S.Kom., M.Sc.Eng., Ph.D.

ABSTRACT

A large amount of data diversity provides an essential role in presenting information comprehensively from an entity to provide information to users as well as possible. Halal Nutrition Food http://halal.addi.is.its.ac.id is a research product in the form of an application that provides information on halal food products that have been integrated with a dataset of 10 global halal institutions and OpenFoodFacts which is a food product information provider site with a crowdsourcing platform. The impact of data integration that has been done results in duplication of entities and decreases relevance in entity search. Hence, users must repeatedly check on each entity that appears in the search results to get the information sought. In this thesis, the relevance of searching increases the Entity Resolution (ER) approach. The ER approach is carried out using graph embedding called Node2vec, which transforms each node into output in the form of a representative set of vectors. This output is used to find pairs between nodes that have similarities to the source node. This pair of nodes will be used as a candidate for link resolution and prediction. The outputs are used as input in the ranking process to produce PageRank and LinkCount scores. The ranking process is carried out with a query-independent (QI) and query-dependent (QD) approach. QI is calculated using the PageRank and LinkCount algorithm. Whereas QD is calculated using the Weighting Frequency Inverse Entity Frequency (TF-IEF), which calculates the similarity of documents based on certain keywords, then the ranking method uses the BM25 ranking function. The combination of QI and QD calculations will produce a final score, which will be used as a reference weight in ranking the entities to increase the relevance of the search. The results of this thesis show that the most dominant factor is the dependent score, followed by an independent score. The higher the value of the dependent score, the more elevated the chance of a document to appear on the system and become relevant search results. The higher the value of the independent score will affect the ranking rank of the document in the search results. The use of the entity resolution method for ranking has an essential role in the weighting the ranking score of each document through the relations formed (owl: sameAs and rdfs: seeAlso) in the graph embedding proses through similarity process using the results of vector embedding and the content of documents. The results of the ER approach are proven to be able to increase the relevance of entity search in linked open data.

Keywords: *Entity Ranking, Entity Resolution, Linked Open Data, Product Ranking*

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Alhamdulillah, segala puji syukur kepada Allah SWT yang telah memberi kesempatan serta kemudahan bagi penulis sehingga penulis dapat menyelesaikan tesis dengan baik dan tepat pada waktunya. Tesis ini disusun untuk memenuhi salah satu syarat menyelesaikan pendidikan pascasarjana di Departemen Sistem Informasi, Fakultas Teknologi Elektro dan Informatika Cerdas, Institut Teknologi Sepuluh Nopember. Penyusunan tesis ini tidak lepas dari bantuan berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada:

- Ibu Nur Aini Rakhmawati, S.Kom., M.Sc.Eng., Ph.D. selaku Dosen Pembimbing yang telah meluangkan waktu, tenaga dan pikiran, serta memberikan ilmu, dukungan, dan kesabaran selama membimbing penulis dari awal hingga tesis ini selesai.
- Bapak Dr. Apol Pribadi Subriadi, S.T, M.T dan Bapak Dr. Eng. Febriliyan Samopa, S.Kom, M.Kom selaku Dosen Penguji yang telah bersedia menguji dan memberikan masukan untuk penelitian ini.
- Kedua orang tua saya yang selalu memberikan doa dan dukungannya.
- Seluruh responden yang telah menyediakan waktu untuk mengikuti serangkaian eksperimen pada penelitian ini.
- Seluruh Bapak dan Ibu dosen beserta staf karyawan di Departemen Sistem Informasi, Fakultas Teknologi Elektro dan Informatika Cerdas, Institut Teknologi Sepuluh Nopember.
- Teman-teman keluarga besar S2 SI Angkatan 2018 yang telah menemani suka, duka serta dukungannya selama menempuh pendidikan pascasarjana.
- Teman belajar, dan diskusi selama kuliah S2 yang sudah membantu pengerjaan tesis dan selalu menyemangati penulis hingga pengerjaan tesis ini selesai.
- Semua pihak yang tidak dapat disebutkan satu-persatu.

Akhir kata, penulis mengucapkan terimakasih dengan segala hormat dan kerendahan hati. Penulis berharap semoga tesis ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan untuk semua pihak. Apabila pada tesis ini terdapat kata-kata yang kurang berkenan di hati para pembaca sekalian, maka penulis memohon maaf yang sebesar-besarnya. Tesis ini juga masih jauh dari kata sempurna, sehingga penulis sangat terbuka terkait masukan dan kritik dari pembaca. Pembaca dapat mengirimkan masukan dan saran melalui email ahmadchoirunnajib@gmail.com.

Halaman ini sengaja dikosongkan

DAFTAR ISI

DAFTAR TABEL

Halaman ini sengaja dikosongkan

DAFTAR GAMBAR

DAFTAR KODE

Halaman ini sengaja dikosongkan

BAB 1

PENDAHULUAN

Pada bab ini akan dibahas terkait latar belakang dilakukannnya penelitian, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan penelitian, kontribusi penelitian (mencakup kontribusi teoritis dan kontribusi praktis), serta sistematika penulisan laporan tesis.

1.1 Latar Belakang

Jumlah Muslim di dunia semakin berkembang setiap tahun. Pada tahun 2010 jumlah Muslim mencapai 24% atau 1,6 milyar dari jumlah penduduk di dunia (Kettani 2010). Jumlah ini diperkirakan akan terus meningkat dengan tingkat pertumbuhan sekitar 1,705%. Sementara populasi pertumbuhan jumlah penduduk hanya sekitar 1,194%. Meningkatnya jumlah Muslim berakibat pada permintaan akan kebutuhan produk halal seperti makanan halal yang digunakan untuk konsumsi atau promosi dari makanan halal tersebut (Sham et al. 2017). Permintaan ini juga dikaitkan dengan mutu kualitas, kebersihan dan keamanan produk seperti yang telah ditetapkan oleh prinsip syariah Islam (Jaafar et al. 2011).

Halal merupakan sebuah terminologi yang biasa digunakan dalam agama Islam. Halal berarti merujuk kepada segala aktivitas yang diperbolehkan dalam syariah Islam (Rezai et al. 2012). Syariah mengatur garis pedoman terkait dengan hal yang diperbolehkan atau dilarang. Hal ini berlaku dalam berbagai aspek kehidupan, seperti makanan, kehidupan berkeluarga, transaksi bisnis, dan lainnya (Jaafar et al. 2011). Dalam aspek makanan, Islam mengatur makanan yang secara jelas dilarang untuk dikonsumsi dalam Al-Qur'an dan Hadits seperti babi, khamr atau alkohol, bangkai, darah atau daging hewan yang disembelih bukan karena Allah SWT. Oleh karena itu, Muslim harus memastikan dan memiliki kesadaran terhadap sifat halal dari makanan yang hendak dimakan.

Indonesia merupakan negara dengan penduduk Muslim terbesar di dunia. Persentase dari penduduk Muslim Indonesia mencapai 12,7% dari populasi Muslim

dunia dan 88,1% dari 205 juta penduduk Indonesia (Indrawan 2015). Oleh karena itu, pemerintah Indonesia mendirikan Lembaga Pengkajian Pangan Obat obatan dan Kosmetika Majelis Ulama Indonesia (LPPOM MUI) yang bertujuan untuk melakukan pengawasan dan sertifikasi terhadap produk halal yang beredar di Indonesia.

Saat ini, LPPOM MUI menyediakan sebuah situs web http://www.halalmui.org yang memudahkan pengguna dalam mencari sertifikat halal berdasarkan nama produk, manufaktur atau nomor sertifikat. Hasil pencarian menyajikan informasi berupa nama produk, nama manufaktur, nomor sertifikat, dan kedaluwarsa dari sertifika halal. LPPOM MUI menyediakan kumpulan sertifikat produk halal secara lengkap dalam bentuk dokumen PDF. Hal ini mengakibatkan sulitnya ekstraksi data sertifikat produk halal dan integrasi dengan dataset lain. Selain itu, LPPOM MUI tidak menyediakan informasi terkait komposisi produk dari produk yang telah tersertifikasi halal.

Dari permasalahan tersebut, Rakhmawati (Rakhmawati et al. 2019) mengembangkan sebuah situs web Halal Nutrition Food http://halal.addi.is.its.ac.id yang memanfaatkan teknologi Linked Data untuk menyajikan informasi produk halal dilengkapi dengan informasi nomor sertifikat, manufaktur, dan komposisi produk secara mendetail yang diperoleh dari berbagai dataset. Situs web ini mengintegrasikan dataset dari 10 lembaga sertifikasi halal dunia, OpenFoodFacts, Dbpedia, Mesh, Pubchem, Chebi, E-number dan Crowd-sourcing Halal Nutrition Food dengan menggunakan halal food vocabulary dan menyimpannya dalam dokumen berbentuk Resouce Description Framework (RDF) dengan sintaks .ttl (Turtle). Sebelumnya, integrasi data hanya dilakukan dengan dataset lembaga halal LPPOM MUI, Dbpedia, Mesh, Pubchem, Chebi, E-number dan Crowd-sourcing Halal Nutrition Food. Selanjutnya, integrasi dilakukan dengan dataset OpenFoodFacts dan lembaga halal lainnya.

Dengan jumlah lebih dari 390 ribu data yang telah terintegrasi mengharuskan adanya fitur pencarian untuk mempermudah pengguna dalam mencari produk. Tantangan pada fitur pencarian produk yaitu relevansi hasil pencarian. Oleh karena itu, fitur pencarian produk harus memiliki sistem pencarian dengan sistem perang-

kingan produk yang dapat menghasilkan relevansi hasil pencarian yang baik. Sebelum proses integrasi data dengan OpenFoodFacts dan lembaga halal lainnya, sistem perangkingan produk pada fitur pencarian dikembangkan menggunakan pendekatan Query-Independent (QI) dan Query-Dependent (QD) (Delbru, Rakhmawati & Tummarello 2010). QI merupakan pembobotan yang dilakukan sebelum pencarian dilakukan, sedangkan QD merupakan pembobotan yang dilakukan berdasarkan query pengguna. Pembobotan QI dihitung berdasarkan jumlah relasi yang mengarah pada entitas, sedangkan QD dihitung berdasarkan term frequency-inverse entity frequency (TF-IEF) yang menilai pentingnya term yang ada pada query terhadap setiap entitas. Kedua pembobotan ini dilakukan normalisasi menggunakan fungsi sigomid (Craswell et al. 2005) yang akan menghasilkan final-score sebagai skor terakhir yang digunakan dalam perangkingan produk dalam hasil pencarian.

Setelah proses integrasi data dengan berbagai dataset, terdapat duplikasi entitas pada dataset Halal Nutrition Food. Duplikasi pada entitas berasal dari sumber yang berbeda (entity profiles) namun sebenarnya merujuk pada entitas yang sama di dunia nyata (real-world entity) (Christophides et al. 2015). Hal ini menyebabkan berkurangnya kualitas informasi pada real-world entity dan relevansi pada pencarian entitas. Sehingga, pengguna harus melakukan pengecekan berulang untuk mendapatkan informasi yang dibutuhkan dari setiap entitas yang dimunculkan pada hasil pencarian. Entity Resolution (ER) merupakan metode untuk mengidentifikasi dan meleburkan informasi dari entity profiles ke dalam real-world entity untuk memperkaya informasi dan kualitas data (Christophides et al. 2015). ER dapat menjadi solusi untuk menyelesaikan permasalahan duplikasi dan relevansi pencarian entitas.

Dalam permasalahan duplikasi, ER sering digunakan untuk menyelesaikan permasalahan ini. Pada penelitian sebelumnya, ER dikelompokkan menjadi content-based dengan pendekatan blocking dan relational similarity dengan memanfaatkan keberadaan relasi pada entitas. Pendekatan blocking dilakukan dengan cara mengelompokkan entitas berdasarkan token/term atau kata yang terdapat pada nilai dari atribut entitas (Papadakis et al. 2013). Setelah terkelompokkan pada

blok-blok, dilakukan penghitungan similarity antar entitas dengan nilai batas tertentu untuk dilakukan resolusi entitas. Teknik blocking dilakukan dalam rangka memperkecil jumlah perbandingan dan jumlah matching yang keliru. Metode ini kemudian ditingkatkan dengan mengadopsi pendekatan blocking graph yang menggunakan graf berbobot untuk meningkatkan performa blocking yang disebut dengan meta-blocking (Papadakis et al. 2014). Pendekatan ini kemudian dikembangkan lagi dengan mengadopsi Shannon entropy (Shannon 1948) yang menghasilkan aggregate entropy untuk masing-masing gugus atribut dengan tujuan untuk menggandakan koefisien dalam pembobotan blocking graph guna meningkatkan performa dan efisiensi blocking (Simonini et al. 2019).

Sementara dalam permasalahan relevansi pencarian entitas, relevansi dikatikan dengan tantangan berupa rangking entitas pada hasil pencarian (result ranking) dan keberagaman hasil pencarian (result diversity) (Arnaout & Elbassuoni 2018). Result ranking menunjukkan tingkat relevansi atau pentingnya sebuah entitas pada hasil pencarian. Sementara result diversity menunjukkan beragamnya entitas pada hasil pencarian untuk memudahkan pengguna dalam menjelajahi informasi terkait lainnya yang terdapat pada knowledge graph.

Pada penelitian sebelumnya, Hogan ((Hogan et al. 2006) mengusulkan metode rangking Reconrank yang memanfaatkan analisa hirarki links pada entitas berupa algoritma PageRank dengan melakukan modifikasi pada iterasi pada penghitungan bobot PageRank. Metode ini kemudian dikembangkan oleh Delbru (Delbru, Toupikov, Catasta, Tummarello & Decker 2010) dengan menghitung bobot PageRank pada tingkat lapisan dataset kemudian menghitung bobot entitas pada lapisan entitas dengan mempertimbangkan bobot PageRank yang telah dihitung dari dataset asal entitas terkait. Metode serupa dikembangkan kembali oleh Delbru (Delbru, Rakhmawati & Tummarello 2010) dengan pendekatan query-independent dan query-dependent. Query-independent dilakukan dengan memanfaatkan analisa hirarki links sementara query-dependent dilakukan dengan pembobotan term-frequency inverse entity frequency (TF-IEF). Kedua pendekatan ini dikombinasikan untuk menghasilkan skor final yang menentukan rangking entias pada hasil pen-

carian. Kemudian, metode perangkingan TRank diusulkan oleh Tonon yang merangking tipe entitas (entity types) (Tonon et al. 2016). Seringkali entitas memiliki berbagai jenis tipe entitas sehingga diperlukan rangking tipe entitas untuk menampilkan tipe entitas yang paling relevan dari hasil pencarian terhadap sebuah kueri. Metode TRank dikembangkan kembali oleh Tonon menjadi TRank++ dengan menambahkan pendekatan yang berbeda pada saat melakukan rangking tipe entitas yaitu hierarchy based, textual-context, entity-context, entity-based (Tonon et al. 2016). Metode perangkingan lainnya juga dikembangkan oleh Arnaout (Arnaout & Elbassuoni 2018) berupa framework umum yang dapat digunakan mencari entitas pada dataset. Framework ini melakukan ekstensi pada RDF dengan menambahkan bobot dan keyword beserta bobot pada setiap entitas. Metode perangkingan dilakukan dengan memanfaatkan tripple-pattern queries with dengan menambahkan keywords diakhir query untuk memperluas hasil pencarian. Untuk menyajikan keberagaman hasil pencarian, framework ini memanfaatkan Maximal Marginal Relevance.

Pendekatan-pendekatan pada peneletian sebelumnya, pendekatan ER umumnya dilakukan dengan content-based similarity menggunakan pendekatan blocking. Sementara penelitian yang memanfaatkan pendekatan relational similarity belum memanfaatkan graph embedding untuk mengetahui keberadaan relasi dan konteks entitas pada graph. Sejauh ini masih belum ada penelitian yang memanfaatkan graph embedding untuk melakukan ER maupun rangking entitas pada kasus data semi-terstruktur RDF atau Linked Data. Sehingga, pada penelitian ini akan diajukan sebuah teknik perangkingan entitas dengan pendekatan ER yang memanfaatkan graph embedding. Pendekatan ER dengan graph embedding akan menjawab permasalahan duplikasi dan meningkatkan pencarian dengan memberikan skor pembobotan yang akan menjadi masukan pada proses perangkingan pendekatan query-independent. Pendekatan rangking dengan query-independent dan query-dependent akan melakukan pembobotan perangkingan pada setiap entitas untuk menjawab tantangan rangking entitas untuk meningkatkan relevansi hasil pencarian.

Pendekatan ER menggunakan graph embedding dilakukan menggunakan Node2vec. Node2vec merupakan salah satu algoritma graph embedding yang meng-

ubah graf menjadi vektor yang selanjutnya dapat diolah ke dalam machine learning (Grover & Leskovec 2016). Node2vec akan mempelajari fitur berupa node sekitar sumber node dengan metode RandomWalks dan merubahnya kedalam vektor. Setelah berhasil diubah ke dalam vektor, maka vektor tersebut akan dijadikan sebagai luaran embedding yang merepresentasikan graf. Luaran berupa embedding vektor tersebut akan dihitung node similarity melalui vektornya untuk mengetahui node-node sekitar yang memiliki kemiripan dan merepresentasikan node sumber misalnya dengan menggunakan cousine similarity. Jika dikaitkan dengan pendekatan ER khususnya metode token blocking atau meta blocking, luaran Node2vec berperan seperti halnya token blocking atau meta blocking, yaitu menghasilkan node-node yang memiliki kemiripan ini selanjutnya akan menjadi kandidat untuk diresolusi. Resolusi entitas dilakukan dengan memanfaatkan string similarity (misal: Jaccard similarity) yang membandingkan atribut pada masing-masing kandidat dengan konfigurasi pembobotan dan batas tertentu. Hasil kandidat string similarity yang memenuhi batas tertentu akan dileburkan dengan menambahkan properti owl:sameAs pada masing-masing kandidat, sebaiknya kandidat yang lain akan dilakukan prediksi link dengan ditambahkan properti rdfs:seeAlso yang menunjukkan entitas terkait. Dengan penambahan properti owl:sameAs dan rdfs:seeAlso akan mepengaruhi bobot PageRank dan LinkCount dari sebuah entitas. Kedua bobot ini akan dijadikan sebagai parameter pembobotan pada proses perangkingan pendekatan query-independent.

Pendekatan rangking dengan query-independent dan query-dependent dilakukan seperti halnya penelitian sebelumnya (Rakhmawati et al. 2019). Namun perbedaannya terletak pada pendekatan query-independent. Penelitian sebelumnya menggunakan pembobotan dengan LinkCount dengan menghitung relasi yang terhubung dari sebuah entitas yang dilakukan secara manual. Namun pada penelitian ini akan dilakukan penambahan pembobotan dengan kombinasi algoritma PageRank dan LinkCount dari entitas dengan pendekatan ER yang menggunakan graph embedding Node2vec. Perangkingan enttias dengan pendekatan ER yang menggunakan graph embeeding, diharapkan akan meningkatkan relevansi hasil pencarian

enttias pada sistem Halal Nutrition Food.

1.2 Rumusan Masalah

Berdasarkan alasan dan peluang penelitian yang menjadi latar belakang penelitian, maka masalah utama yang ingin diselesaikan melalui penelitian ini adalah perlu adanya sistem yang dapat melakukan perangkingan entitas yang akan meningkatkan relevansi pencarian Halal Nutrition Food. Maka rumusan masalah yang ingin dijawab melalui penelitian ini diuraikan sebagai berikut :

1. Bagaimana melakukan perangkingan entitas pada dataset Halal Nutrition Food menggunakan pendekatan entity resolution yang menggunakan graph embedding?
2. Bagaimana memodelkan entitas pada dataset Halal Nutrition Food menggunakan graph embedding ode2vec?
3. Bagaimana evaluasi pendekatan tersebut dalam meningkatkan relevansi pada sistem pencarian Halal Nutrition Food?

1.3 Tujuan

Berdasarkan latar belakang masalah dan rumusan masalah yang telah didefinisikan sebelumnya, maka tujuan dari penelitian ini diuraikan sebagai berikut:

1. Melakukan perangkingan entitas pada dataset Halal Nutrition Food dengan pendekatan entity resolution menggunakan graph embedding untuk meningkatkan relevansi pencarian pada sistem Halal Nutrition Food.
2. Memodelkan entitas pada dataset sistem Halal Nutrition Food menggunakan graph embedding Node2vec untuk perangkingan entitas.
3. Mengukur kinerja pendekatan perangkingan entitas yang telah dilakukan dengan pendekatan entity resolution menggunakan graph embeeding.

1.4 Manfaat

Manfaat dari penelitian ini adalah memberikan kemudahan kepada pengguna dalam mencari produk dan komposisinya, baik yang memiliki informasi status halal

maupun tidak. Kemudahan ini dicapai melalui sistem yang menyediakan informasi produk halal dengan fitur pencarian yang dapat menyajikan hasil pencarian yang relevan. Hasil pencarian yang relevan akan memberikan pengalaman yang baik bagi pengguna dalam melakukan pencarian informasi produk.

1.5 Kontribusi Penelitian

Kontribusi atau sumbangsih yang diberikan oleh peneliti kepada bidang penelitian yang terkait yaitu berupa kontribusi teoritis serta kontribusi praktis yang dijabarkan sebagai berikut.

1.5.1 Kontribusi Teoritis

Kontribusi teoritis dikaitkan perspektif baru yang original (novelty) untuk memajukan pengetahuan dan dapat digunakan dalam praktik. Kontribusi teoritis yang diberikan oleh peneliti diuraikan sebagai berikut :

1. Perangkingan entitas linked open data menggunakan pendekatan entity resolution yang memanfaatkan graph embedding. Pemanfaatan graph embedding digunakan untuk resolusi entitas dengan prediksi link atau penambahan properti owl:sameAs dan rdfs:seeAlso pada entitas terkait.
2. Penambahan properti ini akan mempengaruhi bobot PageRank dan LinkCount yang berdampak pada skor perangkingan entitas pada hasil pencarian. Selanjutnya, perangkingan entitas dapat meningkatkan relevansi pencarian.

1.5.2 Kontribusi Praktis

Suatu penelitian bisa mempunyai kontribusi praktis jika permasalahan atau isu penelitian yang dipilihnya dapat dimanfaaatkan atau diterapkan dengan kondisi lingkungan yang ada. Kontribusi praktis yang diberikan oleh peneliti diuraikan sebagai berikut :

1. Meningkatkan relevansi pencarian pada sistem pencarian Halal Nutrition Food.
2. Menyelesaikan permasalahan duplikasi entitas pada dataset Halal Nutrition

Food.

1.6 Keterbaruan

Berdasarkan penyusunan penelitian dari pendahuluan, perumusan masalah, tujuan penelitian dan manfaat penelitian dapat ditentukan keterbaruan (novelty) penelitian ini. Keterbaruan pada penelitian ini diuraikan sebagai berikut :

1. Berdasarkan studi literatur yang telah dilakukan, penelitian sebelumnya terkait pemanfaatan graph embedding untuk perangkingan entitas linked open data belum banyak dieksplorasi. Pada umumnya perangkingan entitas linked open data dilakukan dengan penghitungan bobot entitas tanpa memperhatikan representasi/konteks/relasi dari entitas tersebut seperti entitas lain di lingkungan sekitar yang akan mempengaruhi entitas sumber, sehingga permasalahan seperti kemungkinan munculnya entitas yang duplikat dapat terjadi. Hal ini menyebabkan penurunan relevansi pencarian yaitu pengguna harus melakukan pengecekan berulang terhadap entitas yang muncul untuk mendapatkan informasi yang ingin dicari. Pada penelitian ini diusulkan perangkingan entitas dengan pendekatan entity resolution yang memanfaatkan graph embedding. Dengan pemanfaatan graph embedding, representasi dari sebuah entitas dapat diketahui. Sehingga entity resolution dan prediksi link dapat dilakukan. Entity resolution akan menyelesaikan permasalahan duplikasi sekaligus prediksi link akan memberikan penambahan bobot yang akan mempengaruhi skor perangkingan untuk meningkatkan relevansi pencarian.
2. Resolusi entitas diselesaikan dengan prediksi link atau penambahan properti owl:sameAs pada entitas yang memiliki kemiripan diatas batas (threshold) dan sebaliknya untuk entitas yang lain dengan penambahan properti rdfs:seeAlso untuk menunjukkan entitas lain yang memiliki kemiripan/keterkaitan dengan entitas sumber.
3. Penambahan properti akan mempengaruhi bobot PageRank dan LinkCount dari setiap entitas. Bobot ini akan mempengaruhi skor perangkingan pada hasil pencarian.

1.7 Batasan Penelitian

Agar dalam penulisan penelitian ini dapat terarah dan terfokuskan serta tidak melebar membahas permasalahan diluar pembahasan yang ada, maka diberikan batasan penelitian sebagai berikut:

1. Data yang digunakan berasal dari kumpulan dataset yang telah diintegrasikan dengan sistem Halal Nutrtition Food.
2. Data entitas yang dirangking berasal dari entitas produk.
3. Integrasi data dilakukan pada tingkat produk, bukan varian produk.
4. Pencarian produk tidak dilakukan secara semantik.
5. Perangkingan tidak dilakukan untuk merangking produk mana yang lebih halal
6. Sistem pencarian tidak melakukan klasifikasi/prediksi halal
7. Objek penelitian dilakukan pada fitur sistem pencarian Halal Nutrition Food

1.8 Sistematika Penulisan Laporan

Agar penulisan laporan proposal penelitian bersifat sistematis, maka sistematika penulisan laporan proposal penelitian yang dilakukan adalah sebagai berikut:

Bab 1: Pendahuluan

Bab ini terdiri dari latar belakang dilakukannnya penelitian, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan penelitian, kontribusi penelitian (kontribusi teoritis dan kontribusi praktis), dan sistematika penulisan.

Bab 2 : Kajian Pustaka

Bab ini berisi kajian terhadap teori dan penelitian-penelitian yang sudah ada sebelumnya terkait dengan bidang data dan semantik. Kajian pustaka ini bertujuan untuk memperkuat dasar dan pemahaman terkait dengan bidang yang akan diteliti.

Bab 3 : Metodologi Penelitian

Bab ini membahas mengenai rancangan penelitian, lokasi dan tempat penelitian, dan juga tahapan-tahapan sistematis yang digunakan selama melakukan penelitian.

Bab 4 : Hasil dan Pembahasan

Bab ini membahas mengenai hasil penelitian yang diperoleh mengacu pada tahapan metode penelitian serta analisis dan pembahasan terkait dengan hasil yang diperoleh.

Bab 5 : Kesimpulan dan Saran

Bab ini membahas mengenai kesimpulan dari seluruh kegiatan penelitian dan saran penelitian kedepannya.

Halaman ini sengaja dikosongkan

BAB 2

TINJAUAN PUSTAKA

Bab ini menjelaskan mengenai teori-teori yang digunakan dalam penyusunan thesis serta kajian pustaka yang diambil dari penelitian-penelitian sebelumnya yang relevan. Kajian pustaka ini selanjutnya akan dibangun sebagai landasan dalam melakukan penelitian ini.

2.1 Kajian Teori

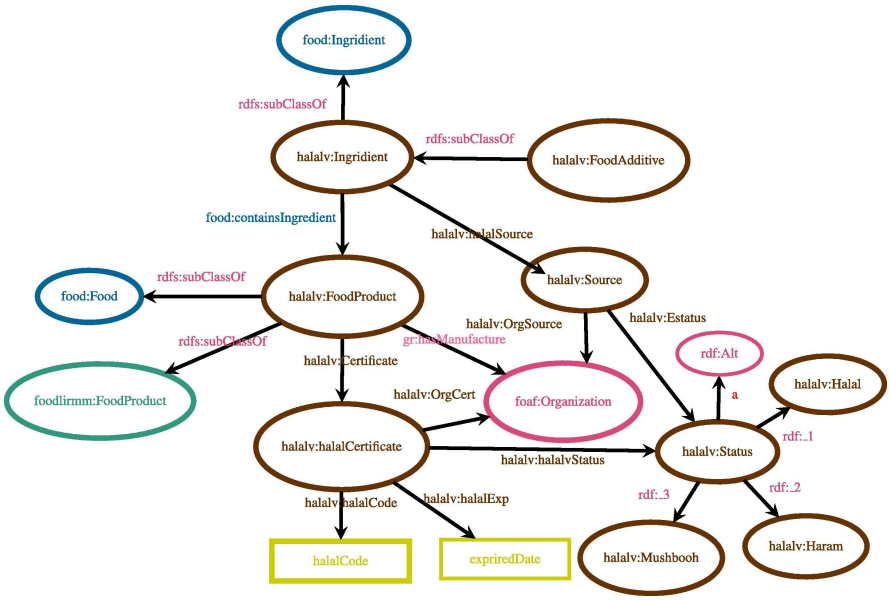
2.1.1 Produk Halal

Produk adalah barang dan/atau jasa yang terkait dengan makanan, minuman, obat, kosmetik, produk kimiawi, produk biologi, produk rekayasa genetik, serta barang gunaan yang dipakai, digunakan, atau dimanfaatkan oleh masyarakat. Halal merupakan sebuah terminologi yang biasa digunakan dalam agama Islam. Halal berarti merujuk kepada segala aktivitas yang diperbolehkan dalam syariah Islam (Rezai et al. 2012). Produk Halal adalah produk yang telah dinyatakan halal sesuai dengan syariat Islam. Dalam Al-Qur'an dan Hadits disebutkan bahwa haram mengkonsumsi babi, khamr atau alkohol, bangkai, darah atau daging hewan yang disembelih bukan karena Allah SWT. Salah satu cara untuk mengetahui status Halal produk dilihat dari ada atau tidaknya sertifikasi produk tersebut. Namun, di Indoensia hanya 20% produk di pasar yang telah tersertifikasi halal (*Produk Bersertifikasi Halal di Indonesia Baru 20 Persen, Malaysia Sudah 90 Persen* 2018). Lembaga yang berperan sebagai lembaga sertifikasi halal di Indonesia yaitu Lembaga Pengkajian Pangan Obat obatan dan Kosmetika Majelis Ulama Indonesia (LPPOM MUI) (*Lembaga Pengkajian Pangan Obat-obatan dan Kosmetika MUI* 2019). Untuk mendukung sosialiasai produk halal kepada masyarakat, LPPOM MUI menyediakan situs web http://www.halalmui.org yang memudahkan masyarakat dalam melakukan pencarian produk halal. Informasi yang ditampilkan berupa nama produk, manufaktur, sertifikat halal dan kedaluwarsa dari sertifikat. LPPOM MUI juga menyediakan keseluruhan sertifikasi produk halal dalam dokumen format PDF. Na-

mun, dokumen PDF sulit untuk diekstraksi dan dintegrasikan dengan dataset lain.

2.1.2 Dataset Halal Nutrition Food

Dataset Halal Nutrition Food terdiri dari beberapa gabungan dataset yang berasal dari institusi halal misalnya LPPOM MUI, data crowd-sourcing, Pubchem, Chebi, E-number, Mesh, Dbpedia, dan OpenFoodFacts. Dalam mengintegrasikan data produk, Halal Nutrition Food membuat vocabulary baru untuk produk halal (Rakhmawati et al., 2019). Vocabulary ini menambahkan objek entitas dan properti baru yaitu entitas halalv:FoodProduct, halalv:Ingredient, halalv:Source, halalv:Status dan properti halalv:Certificate, halalv:halalCode, halalv:halalvStatus, halalv:halalExp, halalv:halalSource, dan halalv:OrgCert. Vocabulary ini dibuat dengan melakukan ekstensi terhadap vocabulary yang sudah ada misalnya seperti food:Food, foodlirmm:FoodProduct, food:Ingredient, dan foaf:Organization. Ilustrasi vocabulary diilustrasikan seperti Gambar 2.1.



Gambar 2.1: Vocabulary Halal Nutrition Food (Rakhmawati et al. 2019)

Saat ini dataset Halal Nutrition Food terdiri dari lebih dari 380 ribu produk, 200 ribu komposisi produk, 2584 manufaktur, dan 2976 sertifikat seperti Tabel 2.1. Dataset Halal Nutrition Food bertambah dengan ditambahkannya dataset institusi halal lainnya.

Tabel 2.1: Statistik Dataset Halal Nutrition Food

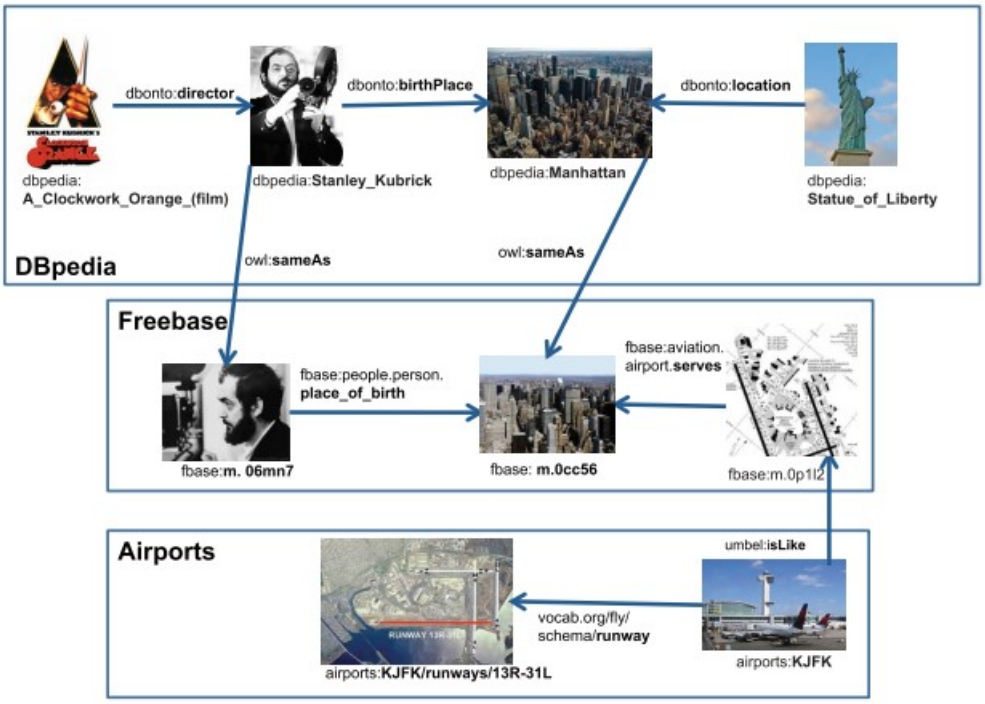
Entitas	Jumlah
Produk	389,145
Komposisi	222,024
Manufaktur	2,584
Sertifikat	2,976

2.1.3 Entity Resolution

Permasalahan duplikasi pada entitas berasal dari sumber yang berbeda (entity profiles) namun sebenarnya merujuk pada entitas yang sama di dunia nyata (real-world entity) seperti Gambar 2.2. Entity Resolution (ER) merupakan metode untuk mengidentifikasi dan meleburkan informasi dari entity profiles ke dalam real-world entity untuk memperkaya informasi dan kualitas data . Pendekatan ER yang sering digunakan yaitu blocking. Blocking dilakukan dengan cara mengelompokkan entitas pada blok-blok tertentu berdasarkan token atau kata yang didapat dari nilai atribut enttias. Entitas-entitas yang terkelompokkan pada setiap blok akan dilakukan perhitungan similarity pada nilai atributnya dengan konfigurasi dan batas tertentu untuk menentukan apakah entitas tersebut merujuk pada entitas sama. Blocking bertujuan untuk memperkecil jumlah perbandingan dan jumlah matching yang keliru. Hingga saat ini terdapat dua jenis pendekatan blocking yaitu token blocking Papadakis et al. (2013) dan meta blocking Papadakis et al. (2014), Simonini et al. (2019).

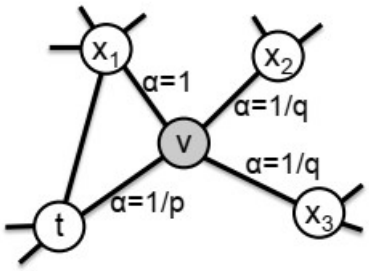
2.1.4 Node2vec

Node2vec merupakan salah satu algoritma graph embedding yang mengubah graf menjadi vektor yang selanjutnya dapat diolah ke dalam machine learning (Grover and Leskovec, 2016). Node2vec menggunakan strategi sampling untuk mencari node sekitar dari sumber node dengan mengakomodasi dua pendekatan yaitu Breadth-first sampling dan Depth-first sampling. Metode sampling ini dinamakan biased RandomWalks. RandomWalks dilakukan dengan konfigurasi hyperparameter p dan q. Hyperparameter p akan menentukan probabilitas perpindahan

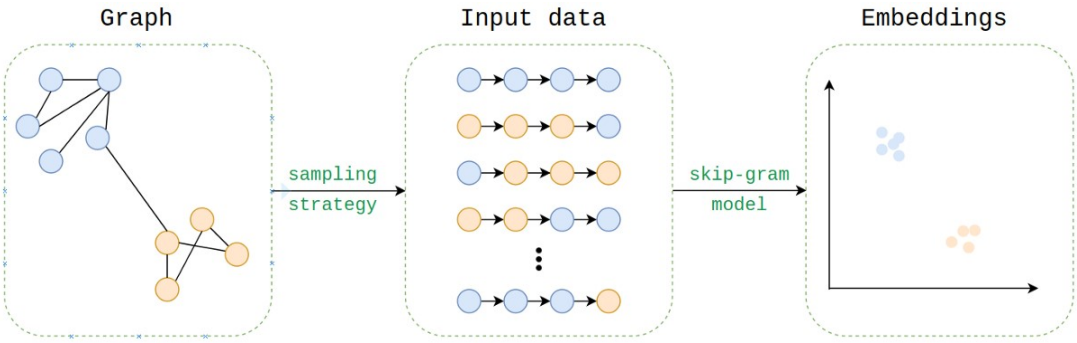


Gambar 2.2: Entity Resolution (Christophides et al. 2015)

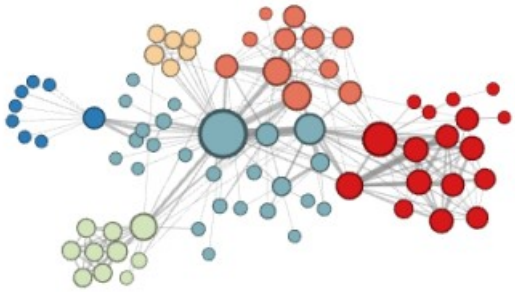
dari sebuah node kembali ke node sebelumnya. Sedangkan hyperparameter q akan menentukan propabilitas perpindahan dari sebuah node untuk melangkah lebih jauh dari node sumber seperti Gambar 2.3. Luaran dari RandomWalks yaitu kumpulan node yang telah diubah kedalam vektor seperti Gambar 2.4. Vektor tersebut akan merepresentasikan node dari graf misalnya berupa node-node sekitar yang berdekatan, memiliki kemiripan dan merepresentasikan node sumber seperti Gambar 2.5.



Gambar 2.3: RandomWalks Node2vec (Grover & Leskovec 2016)



Gambar 2.4: Proses graph embedding node2vec (Cohen n.d.)



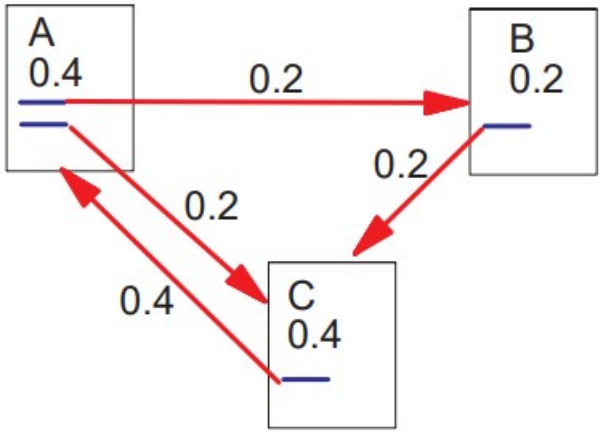
Gambar 2.5: Luaran node2vec dengan homophily (Grover & Leskovec 2016)

2.1.5 Query-independent Ranking

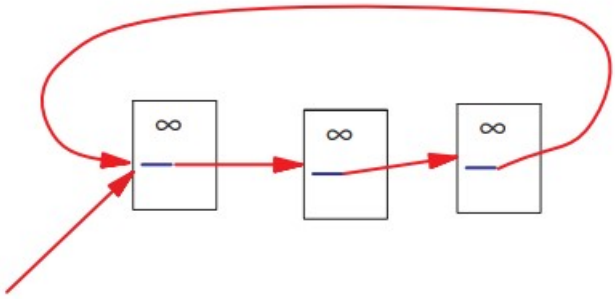
Query-Independent ranking disebut juga sebagai static ranking (Delbru, Toupikov, Catasta, Tummarello & Decker 2010) merupakan salah satu skema ranking yang menggunakan fitur tertentu yang telah diatur sebelum kueri pencarian dilakukan (Craswell et al. 2005). Fitur yang sering digunakan adalah PageRank (Page et al. 1999) dan LinkCount (Latifi & Nematbakhsh 2014).

PageRank (PR) menghitung popularitas berdasarkan iterasi pembagian nilai PR dengan jumlah link yang keluar dari sebuah node hingga mencapai nilai yang diharapkan seperti Gambar 2.6 dan Gambar 2.7. Rumus PageRank didefinisikan pada Persamaan 2.1. Sementara LinkCount menghitung popularitas berdasarkan jumlah link yang mengarah pada sebuah entitas (incoming links). Rumus LinkCount didefinisikan pada Persamaan 2.2.

$$PR(A) = (1-d) + d\left(\frac{PR(T1)}{C(T1)} + .. + \frac{PR(Tn)}{C(Tn)}\right) \tag{2.1}$$



Gambar 2.6: Perhitungan PageRank (Page et al. 1999)



Gambar 2.7: Iterasi hingga mencapai nilai batas tertentu (Page et al. 1999)

$$r(j) = \sum_{l_{\sigma,i,j}} w\left(l_{\sigma,i,j}\right) \tag{2.2}$$

2.1.6 Query-dependent Ranking

Query-dependent ranking merupakan salah satu skema ranking yang menggunakan fitur tertentu yang berdasarkan kueri pencarian dilakukan (Delbru, Rakhmawati & Tummarello 2010). Fitur yang sering digunakan seperti penggunaan TF-IDF. Di dalam RDF, penentuan skor dari subjek atau entitas (e) didapatkan melalui penyatuan skor dari skor predikat (a) dan skor objek (v) yang didapatkan dari TF-IEF seperti Persamaan 2.4-2.7. Namun, pada tesis kali ini, penggunaan TF-IEF hanya digunakan pada proses indexing. Adapun rumus TF-IEF seperti Persamaan

2.3.

$$tf.ief(t,e) = \sum_{k} n(t,e) \times log \frac{|V^E|}{1 + n(t,e)} \tag{2.3}$$

dimana :

n(t,e) = jumlah term t pada dokumen e

V^E = jumlah dokumen

Sedangkan skor dari subjek, predikat, dan objek didapatkan melalui rumus berikut :

$$score(q,v) = \sum_{t \in q} tf.ief(t,e) \tag{2.4}$$

$$score(q,a) = \sum_{c \in p} score(q,v) \tag{2.5}$$

$$score(q,e) = \sum_{p \in s} score(q,a) \times spread(q,e) \tag{2.6}$$

dengan

$$spread(q,e) = \frac{(|q| - |v_t/v_n|) + 1}{|q|} \tag{2.7}$$

dimana $|q|$ menunjukkan jumlah dari term yang berbeda pada query q dan $|v_t/v_n|$ menunjukkan jumlah pembagian antara jumlah objek yang mengandung term dengan jumlah total objek yang terdapat pada sebuah entitas. Faktor normalisasi $spread$ berlaku pada term kueri untuk objek tunggal. Jika term kueri tersebar pada beberapa objek, maka faktor normalisasi $spread$ bernilai lebih kecil. Sebagai contoh jika entitas memiliki 3 objek dan kueri terbentuk dari 3 term, dan semua term terdapat pada objek tunggal, maka faktor normalisasi $spread$ sama dengan $\frac{(3-1/3)+1}{3} = 1,23$. Sebaliknya jika term tersebar pada beberapa objek yang berbeda, maka faktor normalisasi $spread$ sama dengan $\frac{(3-3/3)+1}{3} = 1$.

Sementara itu, proses perangkingan entitas berdasarkan *query* tertentu menggunakan BM25 (*Best Matching*), merupakan sebuah metode untuk menghitung rangking pada kumpulan dokumen (*dataset*) berdasarkan sebuah *query* (Robertson et al. 2009). Rumus BM25 dirumuskan seperti Persamaan 2.8.

$$score(D,Q) = \sum_{i=1}^{n} \text{IDF}(q_i) \cdot \frac{f(q_i,D) \cdot (k_1+1)}{f(q_i,D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{\text{avgdl}}\right)} \tag{2.8}$$

dimana $f(q_i, D)$ merupakan q_i's [[term frequency]] pada dokumen *D*, $|D|$ merupakan panjang dokumen *D* dalam jumlah kata, dan $avgdl$ merupakan rata-rata dari panjang dokumen berdasarkan koleksi dokumen. k_1 b merupakan parameter bebas untuk proses optimasi, dimana $k_1 \in [1.2, 2.0]$ dan $b = 0.75$.

2.1.7 Semantic Web

Semantic Web merupakan sebuah fungsi tambahan dari sebuah web dimana memberikan cara yang lebih mudah untuk menemukan, berbagi, menggunakan kembali, dan menggabungkan informasi. Kemampuan ini dibentuk dengan menggabungkan kemampuan teknologi XML untuk membentuk tagging schemes dan RDF's (Resource Descripton Framework) sebagai pendekatan fleksible yang mewakili data (Webopedia 2016). Semantic web menyediakan format umum untuk pertukaran data. Selain itu Semantic web juga menyediakan bahasa umum untuk merekam bagaimana data berelasi dengan obyek-obyek dunia nyata, memungkinkan orang atau sebuah mesin memulai pada satu database kemudian berhubungan degan database lain dan terkoneksi satu sama lain.

2.1.8 Linked Data

Linked data merupakan salah satu bagian dari pembangunan web semantik. Linked data adalah sebuah pendakatan dimana menghubungkan dan membagikan data pada web. Dengan linked data sebuah website yang memiliki padanan yang sama bisa dihubungkan satu sama lain dengan menggunakan semantic queries. Se-

bagai contoh untuk mendapatkan deskripsi kota surabaya, dengan menghubungkan dengan dataset Dbpedia maka resource dari Dbpedia dapat digunakan kembali dan tidak perlu menuliskannya lagi. Kriteria-kriteria yang terdapat data yang dapat dihubungkan adalah sebagai berikut:

1. Tersedia di internet
2. Memiliki struktur data yang dapat dimengerti oleh mesin
3. Tersedia dalam format non-proprietary
4. Menggunakan standar dari W3C untuk open data
5. Terhubung dengan sumber data lainnya di internet

2.1.9 RDF

Resource Description Framework (RDF) adalah kerangka untuk menerangkan informasi dari sumber-sumber data. Sumber-sumber tersebut dapat berupa apapun, termasuk dokumen, orang, benda fisik, dan konsep-konsep abstrak. RDF ini muncul saat ini dimana Web perlu di proses oleh aplikasi, bukan hanya ditampilkan kepada orang. RDF menyediakan framework umum untuk menginformasikan data sehingga dapat dilakukan pertukaran data antar apliaksi tanpa kehilangan makna (Schreiber et al. 2014).

RDF data model mirip dengan model konseptual sederhana seperti *entity relationship model* atau *class diagram*, namun paada RDF didasarkan pada pembuatan model berdasrkan pernyataan tentang sumber daya / resources (pada web) ke dalam bentuk subject-predicate-obyek. Bentuk ini dikenal dengan nama triples pada terminologi RDF. Subyek menunjukkan sumber daya / resources, predikat menunjukkan ciri-ciri atau aspek sumber daya dan menghubungkan antara subyek dan obyek. Untuk lebih jelasnya dapat dilihat ilustrasi di bawah ini:

halalf:Energen_Ras_Kacang_Hijau	rdf:type	halalv:FoodProduct .
(subyek)	(predikat)	(obyek)

Subyek merupakan suatu hal yang dideskripsikan. Sedangkan obyek merupakan data berupa angka, string, tanggal, ataupun URI dari suatu hal atau benda lain yang memiliki hubungan dengan subjek. Predikat merupakan merupakan suatu URI yang

digunakan untuk mendeskripsikan hubungan antara subjek dengan objek. URI dari predikat diambil dari vocabularies, suatu kumpulan URI yang dapat digunakan untuk merepresentasikan informasi terkait bidang tertentu. RDF triples memiliki dua tipe, sebagai berikut:

- Literal Triples, merupakan triples dengan RDF literal berupa string, angka, atau tanggal sebagai objek. Literal triples digunakan untuk mendeskripsikan sifat / properti dari suatu hal / data.
- RDF Links, merepresentasikan hubungan antara dua sumber data. RDF links terdiri dari tiga referensi URI. URI yang digunakan pada subjek dan objek untuk mengidentifikasi sumber data yang saling terkait, serta URI pada predikat untuk mendefinisikan keterkaitan antar data.

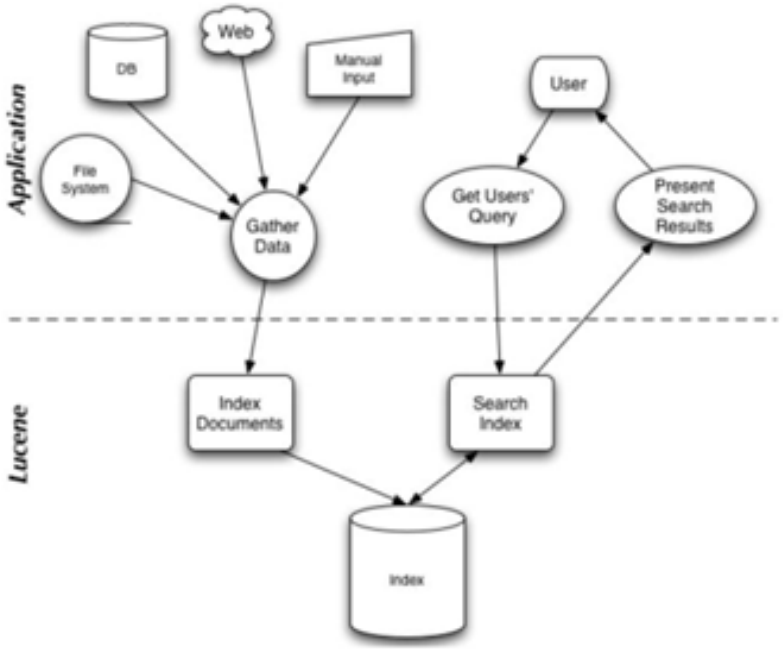
2.1.10 Neo4j Graph Database

Neo4j adalah native graph database yang mampu melakukan transaksi yang cepat dan memenuhi sifat *Atomicity, Consistency, Isolation, Durability* (ACID). Dengan transaksi yang disediakan oleh Neo4j, pengembang dapat memastikan bahwa kegagalan transaksi membuat keadaan basis data tidak berubah untuk memastikan keasliannya. Setiap perubahan pada basis data tidak merusak data, memastikan konsistensi. Data yang dimodifikasi oleh suatu transaksi diisolasi dari transaksi lain sampai transaksi telah selesai dilakukan. Neo4j termasuk graph database yang persisten atau menyimpan data pada perangkat penyimpanan, hasil dari transaksi yang dilakukan selalu dapat diambil, sehingga membuatnya tahan lama (Lal 2015).

Neo4j adalah graph database yang paling matang dan dibuat sejak tahun 2003. Neo4j bersifat open source dengan memiliki komunitas yang besar. Tim pengembangan Neo4j sangat aktif terlibat dengan komunitas itu sehingga fitur dan bug ditangani dengan cepat. Neo4j menyimpan data pada perangkat penyimpanan sehingga penyimpanan transaksional dan pengambilan data dapat dilakukan secara cepat. Dalam melakukan query, Neo4j menggunakan bahasa yang mirip SQL yaitu bahasa *Cypher Query Language* (CQL).

2.1.11 Apache Lucene

Apache Lucene adalah library mesin pencari teks yang memiliki kinerja tinggi dengan fitur lengkap yang seluruhnya ditulis dengan bahasa Java. Teknologi ini cocok untuk hampir semua aplikasi yang memerlukan pencarian teks lengkap, terutama cross-platform. Skema Lucene dengan aplikasi dijelaskan pada Gambar 2.8



Gambar 2.8: Layer aplikasi dengan Lucene

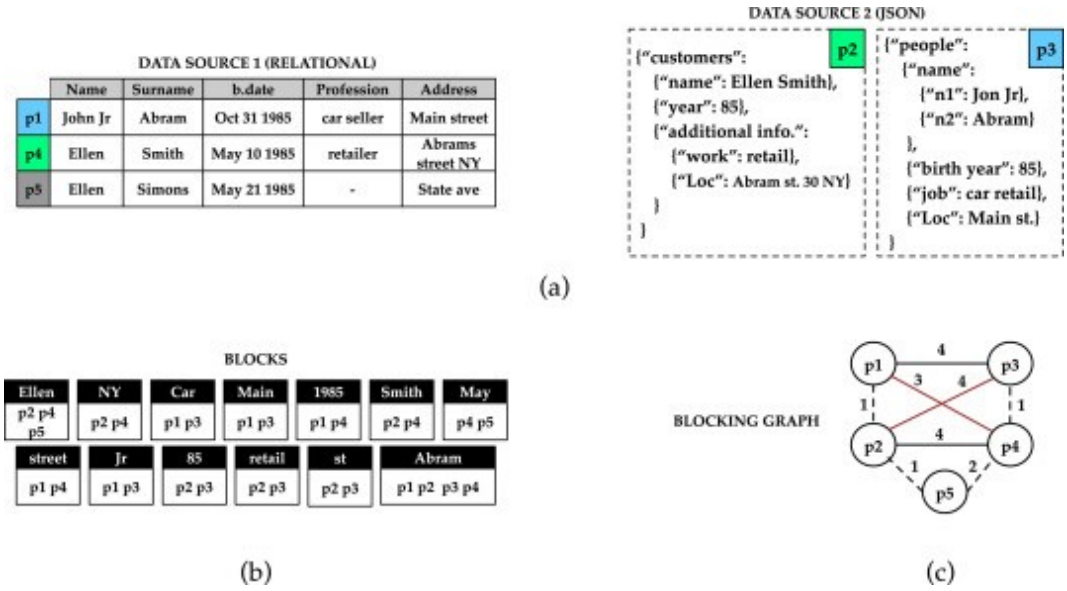
2.2 Kajian Penelitian Terdahulu

2.2.1 Blocking Technique: Schema-agnostic

Teknik blocking berarti mengelompokkan term atau token yang diambil dari nilai sebuah atribut entitas, pengambilan token didasarkan pada teknik bags-of-words. Teknik ini tidak memperhatikan bagaimana schema dari entitas (schema-agnostic). Blocking ini berisi blok yang mewaliki token. Kemudian, setiap entitas yang memiliki kesamaan akan berada dalam blok yang sama. Setelah terkelompokkan pada blok-blok, dilakukan penghitungan similarity antar entitas dengan nilai batas tertentu untuk dilakukan resolusi entitas. Teknik blocking dilakukan dalam rangka memperkecil jumlah perbandingan dan jumlah matching yang keliru.

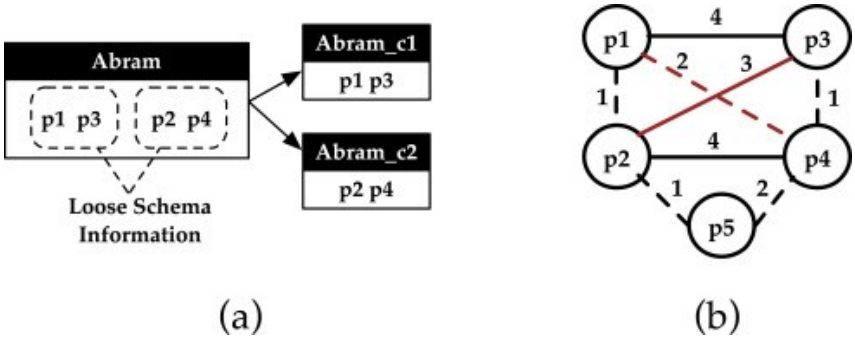
Pendekatan Token blocking (Papadakis et al. 2013) menggunakan teknik blocking dengan mempertimbangkan setiap token sebagai blocking key. Dengan kata lain, setiap token setidaknya memiliki masing pasangan yang akan diresolusi seperti pada Gambar 2.9 (b). Dengan meletakkan setiap entitas ke dalam blok-blok, di satu sisi akan mengurangi kemungkinan kesalahan pada resolusi entitas, namun di sisi lain, akan meningkatkan kemungkinan peletakan entitas yang sebenarnya berbeda namun terletak pada blok yang sama. Hal ini mengakibatkan tingginya recall namun rendah pada precision.

Untuk meningkatkan presisi dari pendekatan Token blocking, maka diusulkan Meta blocking (Papadakis et al. 2014). Meta blocking merupakan sebuah pendekatan yang melakukan pengelompokan kembali terhadap koleksi blok dengan memodelkannya sebagai graf yang memiliki bobot atau blocking graph, dimana setiap entitas berperan sebagai node dan jumlah banyaknya blok berperan sebagai edge seperti ilustrasi Gambar 2.10 (c). Dalam meningkatkan presisi, setiap node akan dihitung rata-rata bobot dari edge yang terkoneksi dengannya, kemudian dilakukan penghilangan node yang nilainya dibawah rata-rata seperti garis putus-putus.

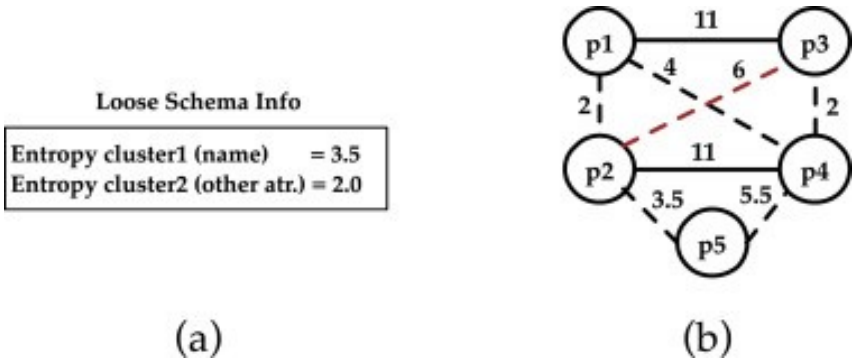


Gambar 2.9: (a) Koleksi entity profiles dengan format yang berbeda (b) Koleksi blok dengan teknik Token Blocking (c) Graph Blocking menggunakan Teknik Meta Blocking (Simonini et al. 2019)

Namun, pada pendekatan Meta blocking sering terjadi pembobotan yang berlebihan karena masih terdapat entitas yang terkumpulkan pada blok yang sama. Oleh karena itu pendekatan ini dikembangkan kembali oleh Simonini (Simonini et al. 2019) dengan istilah Blast (Blocking with Loosely-Aware Schema Techniques). Blast bertujuan untuk meningkatkan kualitas dari blok-blok yang dihasilkan, hal ini berdampak pada akurasi dan presisi. Pendekatan Blast dilakukan dengan cara memecah blok menjadi blok yang lebih kecil seperti Gambar 2.10 (a) dan menambahkan koefisien (attribute entropy) yang menggandakan bobot edge yang terkoneksi antar node. Dari perubahan bobot akan mempengaruhi rata-rata bobot node dan akan mengeliminasi egde yang memiliki bobot dibawah rata-rata seperti ilustrasi Gambar 2.10 (b) dan Gambar 2.11 (b). Dari hasil eliminasi ini koneksi antar node menjadi lebih sedikit dan menjadikan akurasi dan presisi menjadi semakin baik.



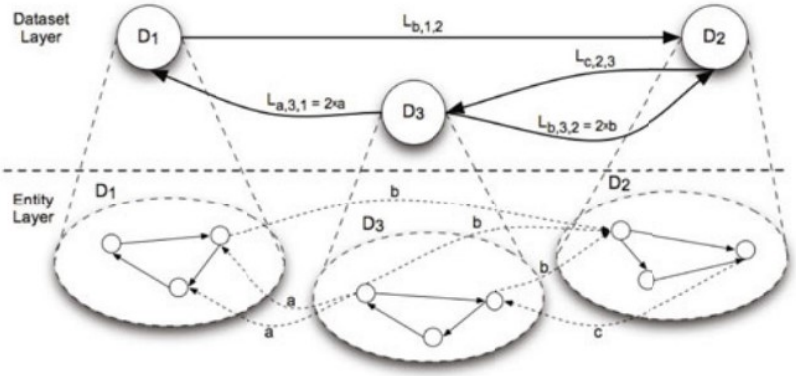
Gambar 2.10: (a) Blok yang memiliki abiguitas dari kata Abram (b) Terjadi perubahan bobot edge (Simonini et al. 2019)



Gambar 2.11: (a) Attribute entropy dan efeknya (b) Eliminasi pada graph (Simonini et al. 2019)

2.2.2 Hierarchical Link Analysis for Ranking Web Data

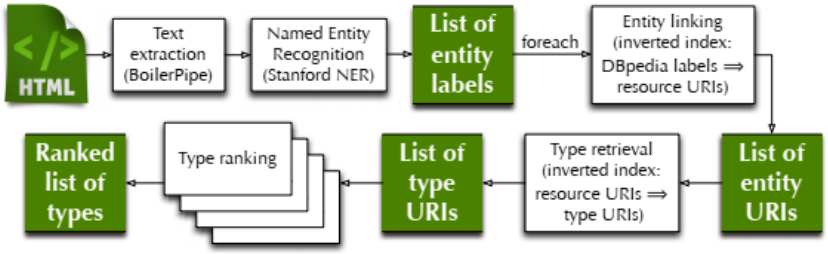
Hirarchical Link Analysis dilakukan oleh Delbru (Delbru, Rakhmawati & Tummarello 2010, Delbru, Toupikov, Catasta, Tummarello & Decker 2010) dengan memberikan nilai bobot ranking PageRank pada lapisan dataset kemudian memberikan nilai bobot pada lapisan entitas. Masing-masing bobot akan dikombinasikan kemudian menghasilkan bobot final yang akan menentukan rangking entitas seperti Gambar 2.12.



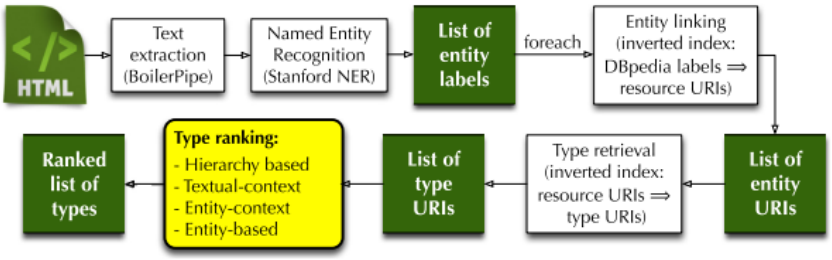
Gambar 2.12: Model dua layer (Delbru, Toupikov, Catasta, Tummarello & Decker 2010)

2.2.3 Entity Type Ranking

Pendekatan perangkingan tipe entitias diusulkan oleh Tonon (Tonon et al. 2013) dengan metode yang disebut TRank. Metode ini merangking tipe entitas (entity types). Hal ini dimotivasi oleh entitas yang seringkali memiliki berbagai jenis tipe entitas sehingga diperlukan rangking tipe entitas untuk menampilkan tipe entitas yang paling relevan dari hasil pencarian terhadap sebuah kueri. Metode TRank dikembangkan kembali oleh Tonon (Tonon et al. 2016) menjadi TRank++ dengan menambahkan pendekatan yang berbeda pada saat melakukan rangking tipe entitas yaitu hierarchy based, textual-context, entity-context, entity-based. Arsitektur perangkingan tipe entitas diilustrasikan seperti Gambar 2.13 dan Gambar 2.14.



Gambar 2.13: Arsitektur TRank (Tonon et al. 2013)



Gambar 2.14: Arsitektur TRank++ (Tonon et al. 2016)

2.2.4 Effective searching of RDF knowledge graphs

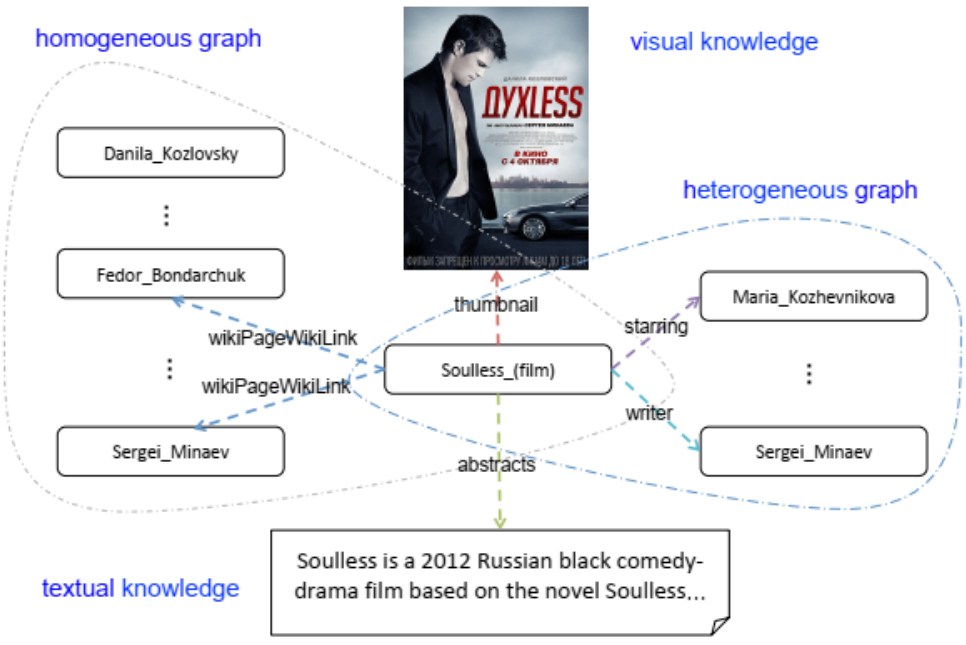
Arnaout (Arnaout & Elbassuoni 2018) mengusulkan pendekatan berupa framework umum yang dapat digunakan mencari entitas pada dataset. Framework ini melakukan ekstensi pada RDF dengan menambahkan bobot dan keyword beserta bobot pada setiap entitas seperti Gambar 2.15. Bobot didapatkan dari irisan jumlah relasi yang mengarah pada entitas dan entitas keyword tersebut. Bobot dan keyword ini akan dijadikan acuan dalam merangking entitas. Metode perangkingan dilakukan dengan memanfaatkan tripple-pattern queries with dengan menambahkan keywords diakhir query untuk memperluas hasil pencarian.

Subject	Predicate	Object	Weight	Keyword:Weight
Annie_Hall	director	Woody_Allen	2032	oscar:1, comedy:251, york:1449, romance:13, …
Annie_Hall	starring	Diane_Keaton	758	psychoanalysis:8, york:567, love:494, sex:8, jew:72, …
Match_Point	writer	Woody_Allen	1866	thriller:121, greed:9, lust:32, money:195, allen:2063, …
Match_Point	starring	Emily_Mortimer	665	london:216, pregnancy:3, thriller:53, love:365, lust:20, …
Match_Point	producer	Letty_Aronson	138	thriller:12, money:20, london:67, crime:14, sister:1, …
Unstrung_Heroes	director	Diane_Keaton	517	1995:115, comic:30, memoir:9, journalist:47, …

Gambar 2.15: Extended RDF (Arnaout & Elbassuoni 2018)

2.2.5 A study of the similarities of entity embeddings learned from different aspects of a knowledge base for item recommendations

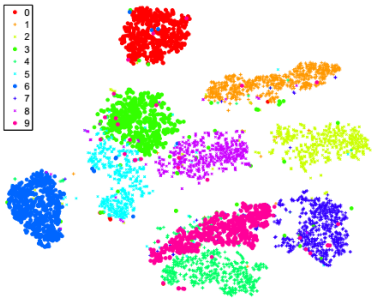
Piao (Piao & Breslin 2018) menggunakan graph embedding untuk menemukan kemiripan entitas pada domain yang spesifik yaitu Music dan Book. Pemanfaatan graph embedding ini selanjutnya dimanfaatkan untuk rekomentasi item dari sebuah entitas. Uji coba dilakukan pada dataset last.fm dan dbbook. Pendekatan yang dilakukan pada penelitian ini meggunakan beberapa aspek untuk mempelajari graf homogen seperti ilustrasi Gambar 2.16, diantaranya hasil terbaik menggunakan properti dbo:wikiPageWikiLink. Metode graph embedding yang digunakan yaitu Doc2vec, Node2vec dan TransE. Berdasarkan perhitungan menggunakan cousine similarity, hasil terbaik dari uji coba yang dilakukan menunjukkan bahwa Node2vec memberikan hasil yang lebih baik dibandingkan dengan yang lainnya. Pada penelitian ini tidak ada penambahan metode perangkingan untuk merangking ulang entitas dalam rangka meningkatkan relevansi rekomendasi item. Relevansi hanya didasarkan pada hasil embedding dari masing-masing metode graph embedding.



Gambar 2.16: Pengetahuan Film Sousless dilihat dari berbagai aspek dari Dbpedia (Piao & Breslin 2018)

2.2.6 t-SNE (Stochastic Neighbor Embedding)

SNE (Stochastic Neighbor Embedding) merupakan sebuah metode yang digunakan untuk mengkonversi vektor *high-dimensional* pada dataset menjadi *low-dimensional* dengan tetap mempertahankan informasi lingkungan sekitar vektor dari kumpulan vektor tersebut (Hinton & Roweis 2003). SNE melakukan konversi dengan cara melakukan proses similarity antar vektor dengan menggunakan *Euclidian distance* kemudian mengelompokkannya ke dalam sebuah map berdasarkan distribusi normal *Gaussian*. Namun, metode ini memiliki kekurangan dalam pengelompokkan, yaitu berupa titik ramai *crowd-points*, berupa ketidakmampuan dalam mengelompokkan beberapa titik yang mirip namun sebenarnya terletak pada kluster yang berbeda. Sedangkan t-SNE merupakan pengembangan dari SNE dengan cara mengurangi *crowd-points* pada titik tengah dari map dengan cara menggunakan *Students distribution* atau yang dikenal sebagai *t-distribution* (Maaten & Hinton 2008). Pada penelitian ini, visualisasi dari embedding akan menggunakan metode t-SNE. Contoh hasil visualisasi t-SNE pada dataset MNIST pada penelitian (Maaten & Hinton 2008) seperti Gambar 2.17.



Gambar 2.17: Visualisasi t-SNE pada Dataset MNIST (Maaten & Hinton 2008)

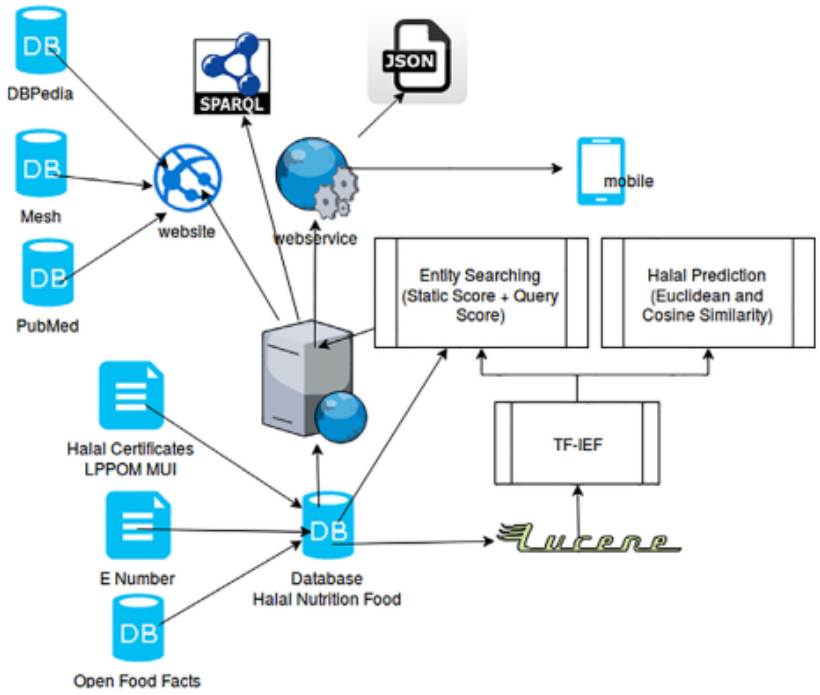
2.2.7 Linked Open Data for Halal Food Products

Rakhmawati (Rakhmawati et al. 2019) mengembangkan sebuah situs web Halal Nutrition Food http://halal.addi.is.its.ac.id yang memanfaatkan teknologi Linked Data untuk menyajikan informasi produk halal dilengkapi dengan informasi nomor sertifikat, manufaktur, dan komposisi produk secara mendetail yang diperoleh dari berbagai dataset. Situs web ini mengintegrasikan dataset dari LPPOM

MUI, OpenFoodFacts, Dbpedia, Mesh, Pubchem, Chebi, E-number dan Crowdsourcing Halal Nutrition Food dengan menggunakan halal food vocabulary seperti Gambar 2.1 dan menyimpannya dalam dokumen berbentuk RDF dengan sintaks .ttl (Turtle).

Halal Nutrition Food juga menyediakan fitur pencarian dan prediksi halal. Fitur pencarian memanfaatkan pendekatan query-independent dan query-dependent. Pendekatan query-independent dihitung dengan algoritma LinkCount yang menghitung jumlah relasi yang masuk pada sebuah entitas. Sedangkan pendekatan query-dependent dihitung dengan menggunakan term frequency-inverse entity frequency (TF-IEF). Kedua pendekatan ini dikombinasikan untuk menghasilkan skor final yang akan dijadikan sebagai acuan proses perangkingan entitas pada hasil pencarian. Fitur prediksi halal yang memanfaatkan entity similarity dengan algoritma cousine similarity pada komposisi produk. Entitas yang belum memiliki sertifikat halal akan diasumsikan memiliki status "halal" jika memiliki kesamaan komposisi produk dengan skor cousine similarity lebih dari batas threshold yang telah ditentukan.

Halal Nutrition Food juga menyediakan akses data berupa Application Programming Interface (API) yang dapat digunakan untuk klien aplikasi lainnya seperti aplikasi Android. Arsitektur sistem dari sistem ini diilustrasikan seperti Gambar 2.18.



Gambar 2.18: Arsitektur Halal Nutrition Food (Rakhmawati et al. 2019)

2.3 Ringkasan Kajian Penelitian Terdahulu dan Penelitian yang Diajukan

Ringkasan dan perbandingan penelitian terdahulu dengan penelitian yang diajukan diilustrasikan seperti Tabel 2.2.

Tabel 2.2: Perbandingan Penelitian Terdahulu dengan Penelitian yang Diajukan

Penelitian	Menggunakan Entity Resolution	Menggunakan Entity Ranking	Menggunakan Graph Embedding
Token Blocking (Papadakis et al. 2013)	Ya	Tidak	Tidak
Meta blocking (Papadakis et al. 2014)	Ya	Tidak	Tidak
Blast (Simonini et al. 2019)	Ya	Tidak	Tidak
Hierarchichal Link Analysis (Delbru, Rakhmawati & Tummarello 2010, Delbru, Toupikov, Catasta, Tummarello & Decker 2010)	Tidak	Ya	Tidak
Entity Type Ranking (Tonon et al. 2016)	Tidak	Ya	Tidak
Effective search of RDF (Arnaout & Elbassuoni 2018)	Tidak	Ya	Tidak
Similarities of entity embeddings for item recommendations (Piao & Breslin 2018)	Tidak	Tidak	Ya
LOD Halal (Rakhmawati et al. 2019)	Tidak	Ya	Tidak
Penelitian ini	Ya	Ya	Ya

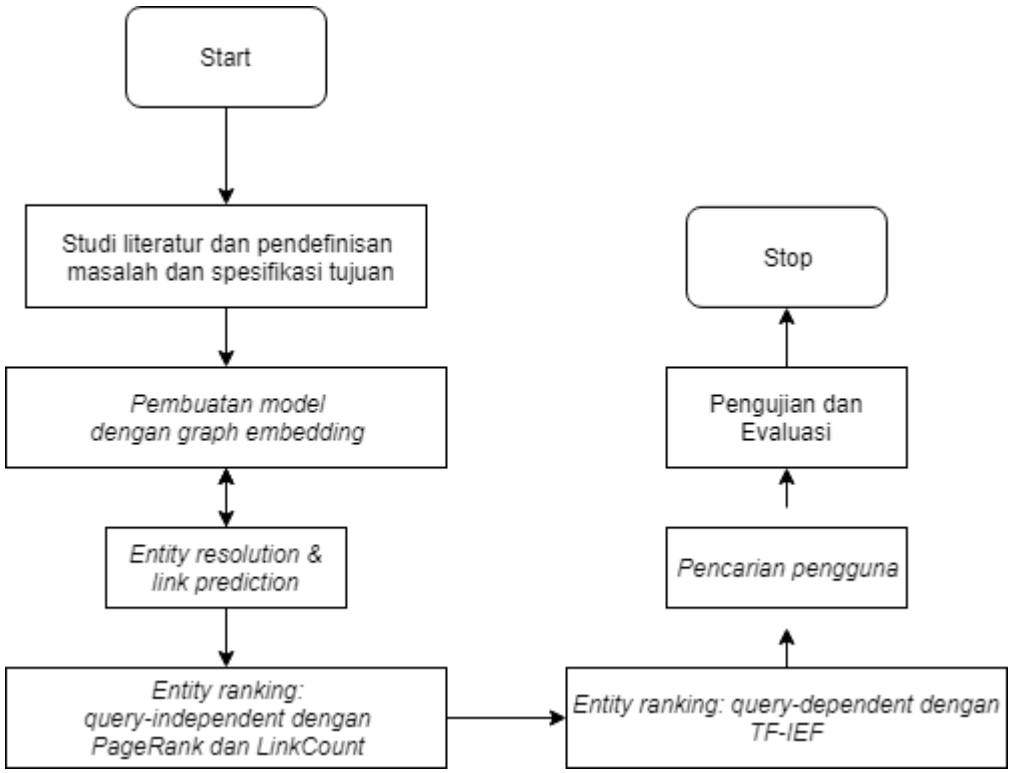
BAB 3

METODOLOGI

Pada bab metodologi akan dijelaskan terkait dengan metodologi penelitian yang dilakukan dalam penelitian tesis mencakup usulan arsitektur sistem, metode penelitian yang digunakan beserta penjelasan tahapan-tahapan yang dilakukan, metode evaluasi, sumber data, dan rencana penelitian.

3.1 Metode Penelitian dan Pengembangan

Berdasarkan usulan arsitektur sistem, diperlukan sebuah metode penelitian untuk mengembangkan sistem Halal Nutrition Food yang akan menyelesaikan rumusan masalah. Berikut merupakan metode penelitian Gambar 3.1 yang digunakan dalam penelitian ini.



Gambar 3.1: Metode penelitian

3.2 Tahapan Penelitian

Berdasarkan metode penelitian yang telah dipaparkan sebelumnya pada Gambar 3.1 maka terdapat tujuh urutan tahapan utama yang dilakukan yaitu:

1. Studi literatur dan pendefinisan masalah dan spesifikasi tujuan
2. Pembuatan model dengan graph embedding
3. Entity resolution & link prediction
4. Entity ranking: query-independent dengan PageRank dan LinkCount
5. Entity ranking: query-dependent dengan TF-IEF dan BM25
6. Pencarian produk
7. Pengujian dan Evaluasi

3.2.1 Studi Literatur dan Definisi Permasalahan dan Spesifikasi Tujuan

Studi literatur dilakukan untuk menemukan research gap dengan mengidentifikasi permasalahan berdasarkan penelitian-penelitian sebelumnya. Research gap akan menentukan posisi dari sebuah penelitian. Identifikasi permasalahan dilakukan untuk menginvestigasi fenomena apa yang harus ditingkatkan dan kenapa perlu ditingkatkan. Identifikasi masalah dilakukan dengan mendefinisikan rumusan permasalahan (research problem) dan spesifikasi tujuan hingga rumusan solusi yang digunakan dapat menyelesaikan permasalahan. Ringkasan studi literatur dari penelitian ini diilustrasikan pada Tabel 2.2.

3.2.2 Pembuatan model dengan graph embedding

Pembuatan model dengan graph embedding terdiri dari dua tahap, yaitu :

1. Menjadikan file dataset RDF Turtle (.ttl) ke dalam bentuk graph. Dataset yang masih terdapat duplikasi diimport ke dalam graph database Neo4j. Kemudian dilakukan ekstraksi node dan relasinya yang akan menghasilkan pasangan node dari setiap node.
2. Melakukan embedding pada node-node dengan graph embedding. Pada tahap ini dilakukan proses embedding yang menggunakan tool library Node2vec yang akan menghasilkan sebuah model embedding yang merepresentasikan

graf dari dataset seperti kemiripan setiap node dan hubungan antar node. Luaran model ini berupa daftar node dengan pasangannya berupa kumpulan vektor yang merepresentasikan node . Vektor ini digunakan untuk mengetahui kemiripan dan hubungan node lain dengan dengan node sumber. Kemiripan node digunakan untuk menemukan kandidat node atau entitas yang akan dilakukan resolusi pada tahap entity resolution. Pada tahap graph embedding akan dilakukan menggunakan pendekatan Node2vec. Masing-masing model embedding akan dilakukan komparasi untuk melihat performa dari masing-masing pendekatan.

3.2.3 Entity Resolution dan Link Prediction

Setelah model embedding dihasilkan pada tahap sebelumnya. Proses entity resolution dilakukan dengan mengukur kemiripan antar node dari kandidat node dari setiap pasangan node berdasarkan input model embedding.

Proses pengukuran kemiripan dilakukan dengan vektor similarity yaitu cousine similarity menggunakan library Gensim for Python. Selanjutnya untuk menghitung kemiripan nilai dari masing-masing atribut dilakukan dengan string similarity yaitu *Jaro-Winkler*, *Cosine* dan *Jaccard similarity* menggunakan library *textdistance* for Python untuk nilai atribut yang terdiri dari beberapa kata misalnya seperti rdfs:label, gr:hasManufacture sedangkan nilai atribut yang terdiri dari banyak kata misalnya seperti food:ingredientsListAsText dilakukan dengan perhitungan frekuensi token yang sering muncul. Masing-masing pengukuran kemiripan akan dibobotkan berdasarkan konfigurasi yang telah ditentukan seperti Tabel 3.1. Kemudian ditentukan nilai threshold untuk menentukan apakah sebuah entitas sumber dan entitas target merupakan entitas yang sama dengan nilai 0,87 pada similarity *Jaro-Winkler* dan *Jaccard*, 0,8 pada similarity *Cosine*.

Untuk setiap pasangan entitas dengan nilai diatas threshold akan ditambahkan properti owl:sameAs yang merujuk kepada pasangan entitas masing-masing. Sebaliknya, akan dilakukan prediksi link dengan menambahkan properti rdfs:seeAlso pada masing-masing entitas. Luaran dari tahapan ini akan menghasilkan dataset

Tabel 3.1: Daftar Konfigurasi Bobot pada Proses Similarity *Entity Resolution* dan *Link Prediction*

No	Nama Produk	Nama Manufaktur	Nama Ingredient
1	0.8	0.2	-
2	0.5	0.2	0.3

yang telah teresolusi dan siap digunakan pada tahap berikutnya. Konfigurasi bobot akan dipilih salah satu dengan hasil resolusi terbaik.

3.2.4 Entity ranking: query-independent dengan PageRank dan LinkCount

Proses perangkingan entitas dilakukan secara independen setelah didapatkan dataset yang telah ditambahkan properti melalui proses entity resolution & link prediction. Setiap entitas akan dihitung nilai PageRank dan LinkCount-nya dan disimpan pada atribut node terkait menggunakan graph database Neo4j. Skor ini disebut dengan *static score*. Setelah pembobotan selesai, dataset akan digunakan kembali pada tahap menghasilkan skor final bersama dengan pendekatan query-dependent.

3.2.5 Entity ranking: query-dependent dengan TF-IEF dan BM25

Pada tahap ini dilakukan proses indexing dari dataset Halal Nutrition Food atau pembuatan model untuk menghasilkan entitas pada tahap pencarian produk. Proses ini menghitung nilai kemiripan entitas berdasarkan token atau kata kunci tertentu dengan metode TF-IEF menggunakan Apache Lucene seperti pada Persamaan 2.3. Proses ini akan memberi bobot untuk setiap entitas dengan kombinasi token-token tertentu. Dari model ini akan digunakan sebagai acuan dalam menghasilkan hasil pencarian berupa dokumen entitas-entitas yang memiliki kemiripan berdasarkan kata kunci tertentu dan proses perangkingan dilakukan dengan menggunakan metode perangkingan BM25. Selanjutnya, skor ini disebut dengan query score.

3.2.6 Pencarian Produk

Pada tahap ini, pencarian produk dilakukan dengan kata kunci tertentu. Skor perangkingan entitas dihasilkan dari penghitungan similarity menggunakan *BM5 Similarity* antar term yang terdapat pada query dengan model indexing berdasarkan perhitungan TF-IEF pada tahap sebelumnya. Kemudian dilakukan penambagan skor dari hasil proses entity resolution berdasarkan model graph embedding berupa PageRank dan LinkCount (*static score*). Kedua skor ini akan dikombinasikan dan dikalkulasi ulang untuk menghasilkan skor final. Skor final akan menjadi bobot akhir dari setiap entitas dan menjadi acuan dalam perangkingan. Skor final (Sf) didapatkan dari *Static score* (Ss) yang terdiri dari kombinasi pembobotan PageRank, LinkCount, serta sertifikat yang terkait pada entitas. *Query Score* (Sq) terdiri dari skor perhitungan similarity antar term query. Sq dilakukan normalisasi menggunakan logaritma log(Sq) dan Ss menggunakan fungsi sigmoid (Craswell et al. 2005) dengan parameter w = 1.8, k = 1 dan a = 0.6 seperti Persamaan 3.1.

$$S_f = log(S_q) + w * \frac{S_s^a}{k^a + S_s^a} \tag{3.1}$$

3.2.7 Pengujian

Setelah sebuah solusi telah didesain dan dikembangkan, maka tahapan berikutnya adalah demonstrasi atau pengujian untuk membuktikkan apakah solusi dapat bekerja dengan baik. Dalam tahapan ini, peneliti mendemonstrasikan penggunaan solusi melalui eksperimen seperti Tabel 3.2.

Setelah pengujian selesai akan dilakukan evaluasi yang menggunakan beberapa pendekatan seperti Precision dan Recall.

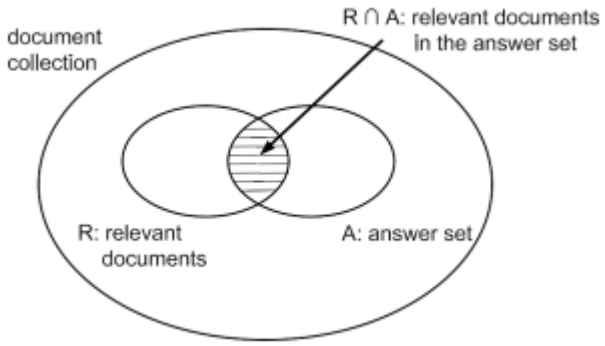
3.2.8 Evaluasi

Diberikan sebuah hasil pencarian dokumen dengan kata kunci tertentu dan didapatkan hasil pencarian seperti Gambar 3.2. Dimana R merupakan dokumen yang relevan, A merupakan hasil pencarian yang diberikan oleh sistem. Sedangkan $R \cap A$ merupakan dokumen relevan yang terdapat pada hasil pencarian. Dapat dii-

Tabel 3.2: Daftar Skenario Pengujian

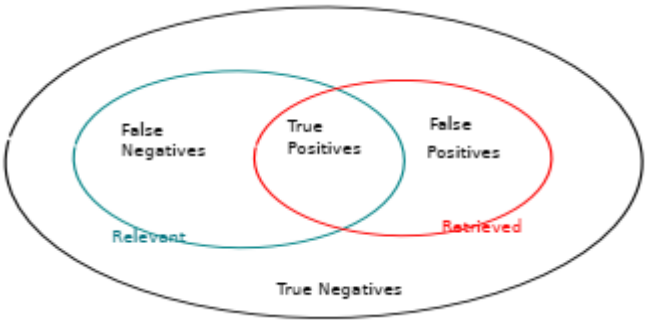
Faktor	Detail	Jumlah	Evaluasi
Relevansi	Pengujian dilakukan menggunakan kata kunci yang: 1. ditambah dengan term halal 2. ditambah dengan term haram 3. tanpa ditambah term halal atau haram	9 term x 3 = 27 kueri	Precision (P@k) Recall (R@k),

lustrasikan bahwa R merupakan False Negative (FN), A merupakan False Positive (FP), dan $R \cap A$ merupakan True Positive (TP) seperti Gambar 3.3 dan Gambar 3.4.



Gambar 3.2: Ilustrasi hasil pencarian (Zezula & Sedmidubsky n.d.)

Diberikan sebuah hasil pencarian dari sebuah query q menghasilkan sejumlah 15 dokumen dengan urutan 1,2..k dan memiliki 5 dokumen yang relevan ditandai dengan tanda bulat seperti diilustrasikan Gambar 3.5. Dapat dihitung nilai Precision dan Recall untuk setiap posisi k dengan Persamaan 3.2 dan 3.3 dan diilustrasikan seperti Gambar 3.6. Dalam kasus perangkingan metode evaluasi yang umum digunakan yaitu P/R@k dengan k 5&10, P/R@5 dan P/R@10. Dengan maksud menghitung masing-masing Precision dan Recall pada saat k ke 5 dan ke 10. Hal ini dimaksudkan karena hasil pencarian yang ditampilkan kepada pengguna pada halaman pertama biasanya berjumlah 10 item dokumen.



Gambar 3.3: Ilustrasi hasil pencarian dengan nilai Tabel Kontingensi (Teufel n.d.)

		THE TRUTH	
WHAT THE SYSTEM THINKS		Relevant	Nonrelevant
	Retrieved	true positives (TP)	false positives (FP)
	Not retrieved	false negatives (FN)	true negatives (TN)

Gambar 3.4: Tabel Kontingensi (Teufel n.d.)

Precision

Berdasarkan ilustrasi Gambar 3.2 dan Gambar 3.3 maka dapat diketahui rumus Precision seperti Persamaan 3.2 atau 3.3 sebagai berikut :

$$Precision = \frac{|R \cap A|}{|A|} \tag{3.2}$$

atau

$$Precision = \frac{TP}{TP + FP} \tag{3.3}$$

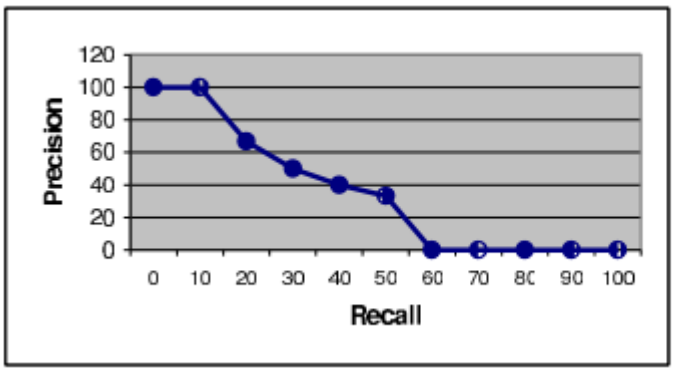
Recall

Berdasarkan ilustrasi Gambar 3.2 dan Gambar 3.3 maka dapat diketahui rumus Recall seperti Persamaan 3.4 atau 3.5 sebagai berikut :

$$Recall = \frac{|R \cap A|}{|R|} \tag{3.4}$$

01. d_{123} •	06. d_9 •	11. d_{38}
02. d_{84}	07. d_{511}	12. d_{48}
03. d_{56} •	08. d_{129}	13. d_{250}
04. d_6	09. d_{187}	14. d_{113}
05. d_8	10. d_{25} •	15. d_3 •

Gambar 3.5: Ilustrasi hasil pencarian dokumen kueri q (Zezula & Sedmidubsky n.d.)



Recall	Precision
0	100
10	100
20	66.6
30	50
40	40
50	33.3
60	0
70	0
80	0
90	0
100	0

Gambar 3.6: Grafik Recall vs Precision Zezula & Sedmidubsky (n.d.)

atau

$$Recall = \frac{TP}{TP + FN} \tag{3.5}$$

3.3 Rangkuman Tahapan Penelitian

Berikut ini merupakan rangkuman dari tahapan penelitian berupa input, tools yang digunakan, dan output dari setiap proses seperti Tabel 3.3.

Tabel 3.3: Rangkuman Tahapan Penelitian

Tahap	Input	Tool	Output
1. Pembuatan model	File dataset Turtle	-Neo4j -Python -Libraryode2vec	-Graph -Relasi node -Embedding model
2. Entity Resolution dan Link Prediction	Embedding model	-Library Gensim for Python -Library Textdistance for Python	-Kandidat entitas -Entitas yang telah diresolusi
3. Entity Ranking: query-independent	Entitas yang telah diresolusi	Neo4j	-Entitas dengan penambahan atribut skor PageRank dan LinkCount -Skor dengan konfigurasi tambahan (sertifikat halal)
4. Entity Ranking: query-dependent	File dataset Turtle	Apache Lucene	-Skor dokumen dengan token tertentu
5. Pencarian produk	-Skor PageRank dan LinkCount -Skor dokumen	Halal Ranker Library	-Dokumen yang telah dirangking berdasarkan kueri tertentu
6. Evaluasi	-Dokumen yang telah dirangking berdasarkan kueri tertentu	-Presisi -Recall	Skor evaluasi

Halaman ini sengaja dikosongkan

BAB 4

HASIL DAN PEMBAHASAN

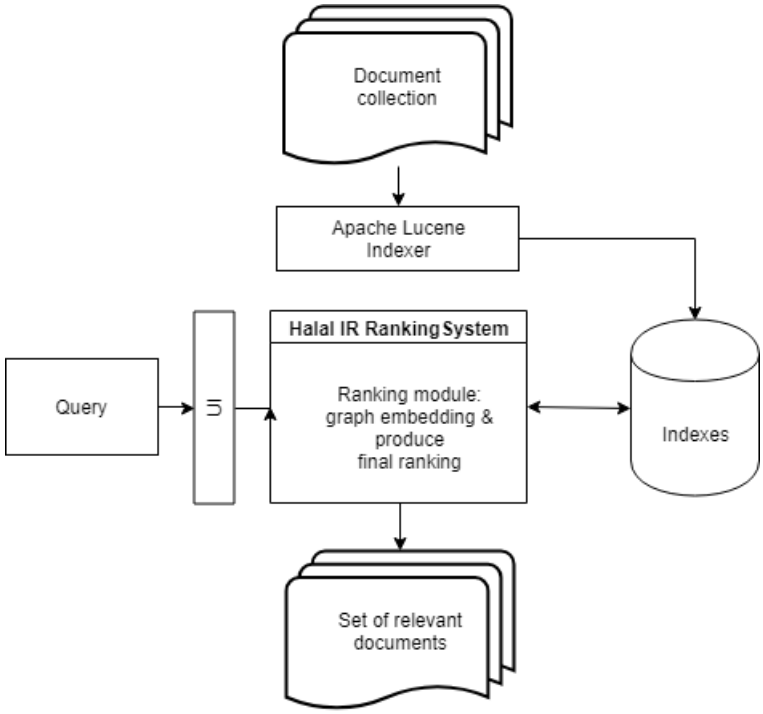
Pada bab ini akan diuraikan secara rinci terkait semua kegiatan yang telah dijabarkan pada metode penelitian hingga diperoleh hasil penelitian. Hasil penelitian tersebut selanjutnya dianalisis dan dibahas lebih lanjut dalam sub bab analisis dan pembahasan.

4.1 Arsitektur Sistem

Pada penelitian ini, terdapat dua bagian proses penting yaitu indexing dan ranking. Pada proses indexing dilakukan dengan pendekatan query-dependent. Proses indexing dilakukan oleh Apache Lucene indexer yang akan menyimpan dokumen-dokumen kedalam sebuah model berbentuk *indexes*. Dari *indexes* ini akan disimpan token-token yang terdapat pada dokumen dengan konfigurasi bobot token pada dokumen masing-masing dengan bobot tertentu sesuai dengan kemunculan token berdasarkan perhitungan *TF-IEF*, model ini akan menjadi awal skor perangkingan dan mempengaruhi perangkingan dokumen berdasarkan kueri pengguna, dimana perangkingan berdasarkan query dilakukan dengan metode rangking *BM25 Similarity*.

Selanjutnya, pada proses ranking dilakukan dengan pendekatan graph embedding Node2vec. Pada proses perangkingan, embedding dari Node2vec akan menghasilkan vektor yang merepresentasikan node sumber. Vektor ini digunakan untuk mendapatkan entitas yang memiliki kemiripan dengan perhitungan *cosine similarity* yang akan berperan sebagai kandidat untuk diresolusi atau prediksi link. Node yang memiliki kemiripan diatas threshold akan diresolusi dengan penambahan properti owl:sameAs, sebaliknya node yang mrmiliki kemiripan dibawah threshold akan ditambahkan properti rdfs:seeAlso. Penambahan properti ini akan mempengaruhi bobot PageRank dan LinkCount yang juga berdampak pada perangkingan entitas. Setelah model graph embedding terbentuk, setiap dokumen yang dimunculkan oleh sistem akan dilakukan kueri lanjutan pada model graph embedding untuk mengetahui infromasi bobot tambahan berupa PageRank dan LinkCount. Masing-masing

bobot pada tahap indexing dan ranking akan dikombinasikan sehingga menghasilkan skor final yang akan menentukan bobot akhir dan rangking dari setiap dokumen yang dimunculkan oleh sistem. Arsitektur pada penelitian ini diilustrasikan seperti Gambar 4.1.



Gambar 4.1: Arsitektur Sistem

4.2 Hasil Penelitian

Pada bagian berikut ini akan dijelaskan mengenai hasil yang didapatkan pada saat penelitian dilakukan dan pembahasan mengenai hasil tersebut.

4.2.1 Pembuatan Model Graph Embedding

Tahap pertama pembuatan model graph embedding yaitu analisa skema dataset, penambahan relasi, dan proses embedding dengan berbagai konfigurasi untuk mendapatkan model dengan hasil terbaik.

Ringkasan Dataset

Dataset Halal Nutrition Food telah diintegrasikan dengan beberapa dataset institusi halal lainnya seperti berikut:

- Jabatan Kemajuan Islam Malaysia (JAKIM)
- Emirates Authority for Standardization and Metrology (EASM)
- Muslim Consumer Group (MCG)
- Supreme Islamic Council of Halal Meat in Australia Inc (SICHMA)
- Taiwan Halal Integrity Development Association (THIDA)
- South Afrika National Halaal Authority (SANHA)
- The Islamic Food and Nutrition Council of Canada (IFANCC).

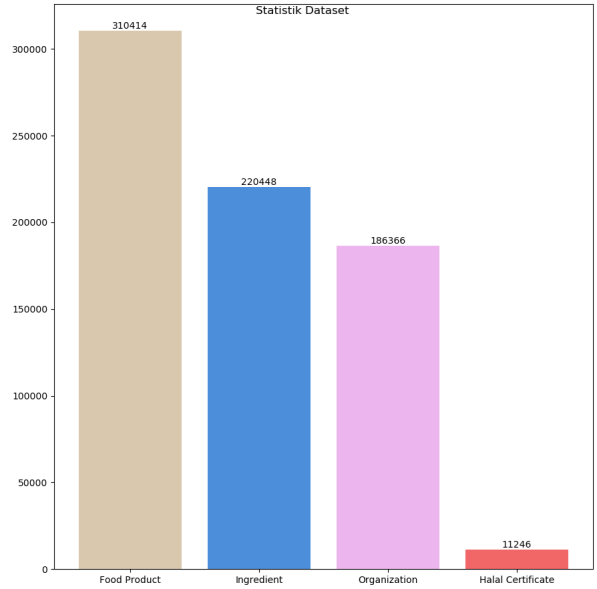
Jumlah setiap dataset yang berhasil diintegrasikan disajikan seperti Tabel 4.1. Jumlah terbanyak diperoleh dari institusi MUI sebanyak 42.004, IFANCC sebanyak 5.855, JAKIM sebanyak 5.603, MCG sebanyak 1.792, SANHA sebanyak 1.785, EASM sebanyak 27, SICHMA sebanyak 23, dan THIDA sebanyak 12. Sementara jumlah masing-masing jenis entitas pada dataset digambarkan seperti Grafik 4.2. Entitas produk berjumlah 310.414, komposisi berjumlah 220.448, organisasi yang terdiri dari manufaktur dan institusi halal berjumlah 186.368 dan 8, dan sertifikat halal berjumlah 11.246. Sementara perbandingan jumlah produk tersertifikasi halal dan jumlah produk belum tersertifikasi yaitu 15.5% berbanding 85.5% diilustrasikan seperti Grafik 4.3. Produk yang belum tersertifikasi mayoritas didapatkan dari dataset *Openfoodfacts*. Setiap entitas berbentuk dokumen file turtle RDF (.ttl), kemudian diimpor kedalam Graph Database Neo4j untuk dilakukan proses analisa, penambahan relasi, dan perangkingan entitas pada tahap berikutnya.

Penambahan Relasi

Penambahan relasi bertujuan untuk memperbanyak jumlah node unik yang akan ditangkap dan memperbesar kemungkinan node lebih banyak terjangkau pada saat proses embedding, khususnya pada proses *RandomWalk*. Pada proses *RandomWalk* akan dilakukan perjalanan secara acak untuk menangkap representasi node sekitar dari sebuah node. Sehingga, node-node yang memiliki kesamaan (similari-

Tabel 4.1: Jumlah Produk per Institusi Halal

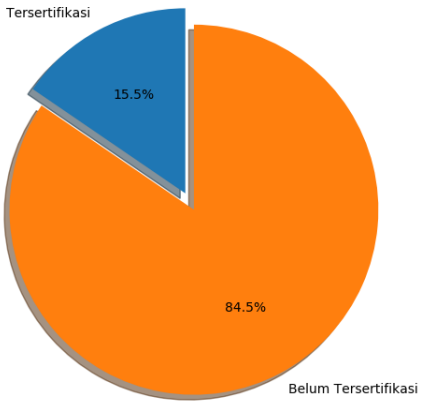
Institusi Halal	Jumlah
Majelis Ulama Indonesia	42.004
The Islamic Food and Nutrition Council of Canada	5.855
Jabatan Kemajuan Islam Malaysia	5.603
Muslim Consumer Group	1.792
The South African National Halaal Authority	1.785
Emirates Authority for Standardization and Metrology	27
Supreme Islamic Council of Halal Meatin Australia	23
Taiwan Halal Integrity Development Association	12



Gambar 4.2: Statistik Dataset Berdasarkan Tipe Entitas

ty) akan direpresentasikan pada lokasi embedding yang berdekatan. Penambahan dilakukan dengan cara melihat skema dataset, dalam tesis ini khususnya relasi yang berkaitan dengan entitas produk. Relasi yang terbentuk pada entitas produk dengan entitas lainnya akan mempengaruhi dan menentukan hasil proses embedding.

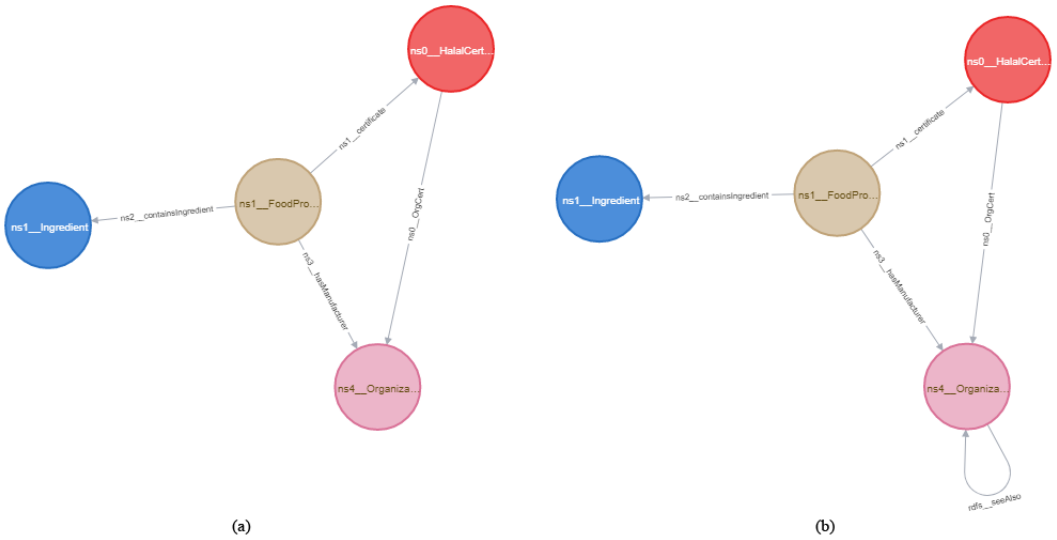
Dari ilustrasi skema pada Gambar 4.4 (a), relasi produk (ns1_FoodProduct) terdiri dari ns1_certificate yang menghubungkan entitas ns0_HalalCertificate, dan ns3_hasManufacturer yang menghubungkan entitas ns4_Organization. Sehingga kemungkinan kombinasi node yang tertangkap ketika melakukan *RandomWalk* dari node ns1_FoodProduct yaitu ns1_FoodProduct dengan ns0_HalalCertificate



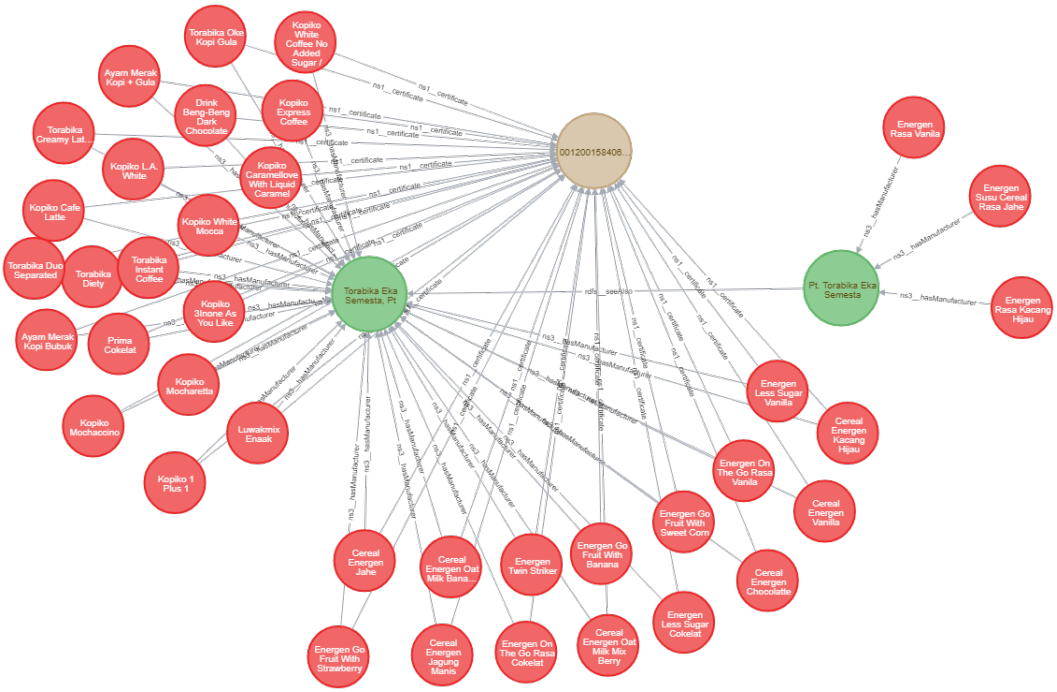
Gambar 4.3: Perbandingan Jumlah Produk Tersertifikasi

dan ns1_FoodProduct dengan ns4_Organization. Namun, berdasarkan kedua relasi tersebut, perjalanan yang menangkap node ns1_FoodProduct ke node ns1_FoodProduct tidak dapat dilakukan karena tidak ada relasi yang menghubungkan salah satu node untuk kembali lagi ke node ns1_FoodProduct setelah melakukan perjalanan, sehingga penambahan relasi antar ns4_Organization dengan ns4_Organization perlu dilakukan agar proses *RandomWalk* dapat menangkap node ns1_FoodProduct yang terhubung dengan node ns1_FoodProduct lainnya melalui perantara node ns4_Organization seperti Gambar 4.4 (b). Relasi tersebut berupa rdfs_seeAlso yang menggambarkan bahwa entitas ns4_Organization memiliki keterkaitan dengan entitas ns4_Organization lainnya seperti Gambar 4.5. Sebagai contoh, pada Gambar 4.5 dapat dilihat bahwa relasi rdfs_seeAlso menghubungkan manufaktur (node warna hijau) ”Pt. Torabika Eka Semesta” yang berasal dari dataset HalalCrowsourcing dengan ”Torabika Eka Semesta, Pt” yang berasal dari dataset Halal MUI yang ditunjukkan dengan adanya node sertifikat (warna coklat) yang dituju oleh node (merah) produk dari manufaktur ”Torabika Eka Semesta, Pt”.

Proses similarity antar manufaktur dilakukan dengan melihat term yang sama pada nama manufaktur asal menggunakan fungsi *get_close_matches* pada class *difflib* dengan metode *Ratcliff-Obershelp algorithm*. Fungsi ini akan memberikan luaran berupa list manufaktur yang memiliki term yang sama dengan nama manufaktur asal. Hasil luaran ini kemudian dilakukan similarity ulang menggunakan



Gambar 4.4: Skema database (a) Sebelum penambahan relasi rdfs_seeAlso, (b) Sesudah penambahan relasi rdfs_seeAlso



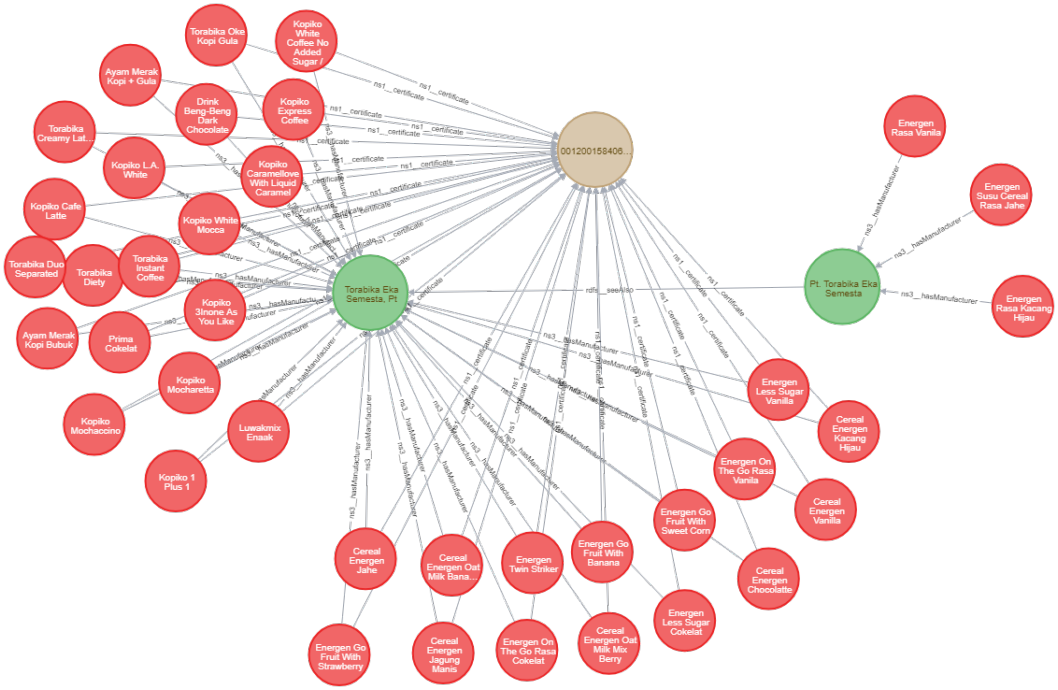
Gambar 4.5: Contoh Penambahan Relasi rdfs_seeAlso antar Manufaktur

similarity Jaro-Winkler, Cosine, dan Jaccard dengan threshold masing-masing 0.9. Keseluruhan proses similarity diilustrasikan seperti Kode 4.1. Hasil dari similarity jumlah relasi rdfs_seeAlso yang berhasil dibuat yaitu 942/56.996 (1,65%) dari total

keseluruhan manufaktur yang memiliki produk dengan sertifikasi halal/haram/mus-booh.

```
candidates = difflib.get_close_matches(manufacture,
 listallmanufactures)
threshold = 0.9
for candidate in candidates:
  similarityJaro = getSimilarityJaro(manufacture, candidate)
  similariyCosine = getSimilarityCosine(manufacture, candidate)
  similarityJaccard = getSimilarityJaccard(manufacture,
 candidate)
  if similarityJaro > threshold or similarityCosine > threshold
 or similarityJaccard > threshold:
      createRdfsSeeAlsoRelationship(manufactureId, candidateId)
```

Kode 4.1: Pseudocode Proses Seleksi dan Simiality Antar Manufaktur untuk membentuk Relasi rdfs:seeAlso



Gambar 4.6: Perbandingan Hasil Relasi rdfs:seeAlso antara Manufaktur dengan Produk Bersertifikasi Halal dengan Keseluruhan Manufaktur

Embedding dengan Metode Node2vec

Permbuatan model dengan metode Node2vec menggunakan beberapa kombinasi parameter untuk mendapatkan luaran model yang terbaik. Metode ini memiliki beberapa parameter yaitu (*d*, *l*, *r*, *p*, dan *q*). Parameter *d* menunjukkan jumlah dimensi embedding, *l* menunjukkan panjang maksimum sebuah perjalanan (*walk*) dilakukan dari sebuah node ke node lainnya, dan *r* menunjukkan jumlah berapa kali perjalanan (*walk*) dilakukan. Sementara parameter lainnya berupa *hyperparameter* yaitu return parameter (*p*) dan inout parameter (*q*). Parameter *p* menunjukkan kemungkinan node akan kembali mengunjungi node sebelumnya setelah melakukan sebuah perjalanan, sementara *q* menunjukkan kemungkinan node akan pergi ke node yang lebih jauh dari node sebelumnya. Penggunaan parameter yang digunakan pada tesis ini yaitu menggunakan kombinasi antara *l*, *p*, dan *q*. Hal ini bertujuan untuk memanipulasi proses RandomWalk yang akan mempengaruhi hasil embedding. Parameter *l* terdiri dari 4 kombinasi yaitu *30, 50, 70, dan 90* yang bertujuan untuk mempengaruhi hasil jumlah tangkapan node dari proses RandomWalk. Sementara parameter *p* dan *q* yaitu *(0,5 dan 0,5)* bertujuan untuk memberikan peluang yang sama apakah akan kembali ke node atau melangkah lebih jauh, sedangkan *(0,4 dan 0,6)* bertujuan untuk memberikan peluang untuk meneruskan langkah lebih jauh setelah melakukan perjalanan daripada kembali ke node sebelumnya. Jumlah kombinasi parameter yang digunakan dalam pembuatan model menggunakan metode Node2vec yaitu berjumlah 24 kombinasi. Masing-masing kombinasi parameter disajikan seperti Tabel 4.2.

Penentuan baik tidaknya hasil luaran model ditentukan oleh berapa jumlah luaran entitas yang didapatkan dari proses *similarity* pada sebuah entitas tertentu (*entity source*), dimana luaran entitas diidentifikasi memiliki kemiripan atau mengacu pada entitas yang sama. Proses similarity ditentukan oleh tingkat kesamaan vektor entitas pada embedding dengan pendekatan *Cosine-similarity* untuk menghasilkan *top-k* entitas, selanjutnya entitas-entitas tersebut dilakukan seleksi kembali dengan pendekatan string similarity pada atribut nama (rdfs_label) dari *entitas source* dengan *entity target* untuk memilih entitas yang memiliki kemiripan atau

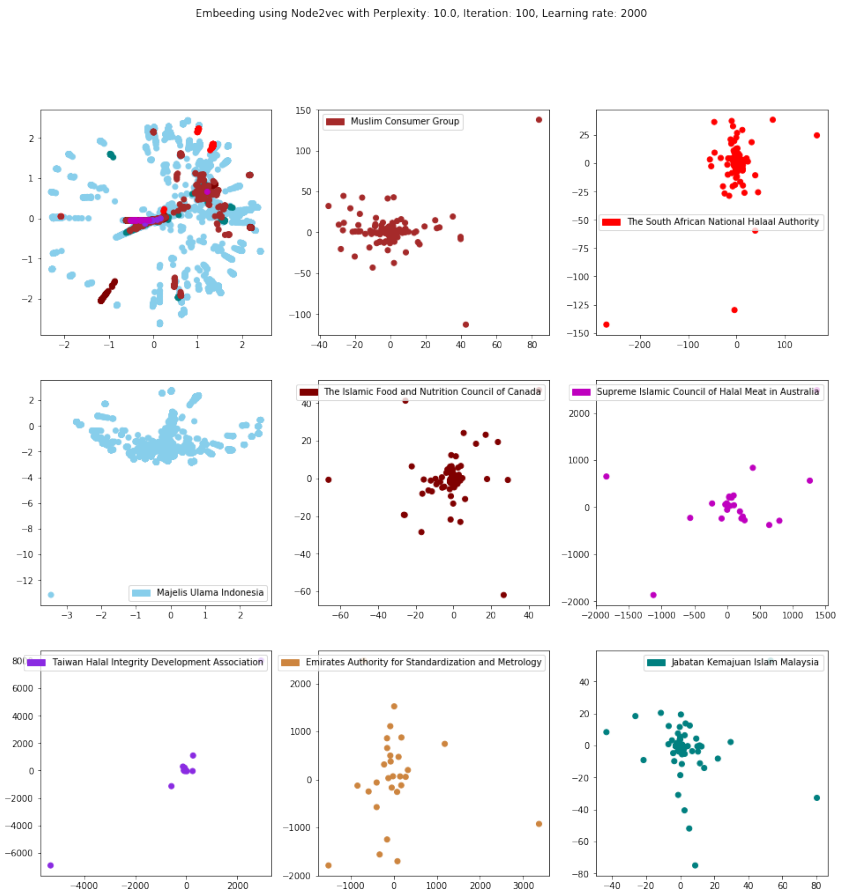
Tabel 4.2: Konfigurasi Parameter dalam Pembuatan Model

Relasi Entitas	***kode***	***d***	***l***	***r***	***p* & *q* (node2vec)**
A. · Produk → Manufaktur	1	128	80	10	(0,5 & 0,5)
· Manufaktur → Manufaktur	2		30		(0,5 & 0,5)
	3		50		(0,5 & 0,5)
	4		70		(0,5 & 0,5)
	5		90		(0,6 & 0,4)
	6		30		(0,6 & 0,4)
	7		50		(0,6 & 0,4)
	8		70		(0,6 & 0,4)
			90		
B. · Produk → Manufaktur	1	128	80	10	(0,5 & 0,5)
· Manufaktur → Manufaktur	2		30		(0,5 & 0,5)
· Produk → Sertifikat	3		50		(0,5 & 0,5)
	4		70		(0,5 & 0,5)
	5		90		(0,6 & 0,4)
	6		30		(0,6 & 0,4)
	7		50		(0,6 & 0,4)
	8		70		(0,6 & 0,4)
			90		
C. · Produk → Manufaktur	1	128	80	10	(0,5 & 0,5)
· Manufaktur → Manufaktur	2		30		(0,5 & 0,5)
· Produk → Sertifikat	3		50		(0,5 & 0,5)
· Produk → Komposisi	4		70		(0,5 & 0,5)
	5		90		(0,6 & 0,4)
	6		30		(0,6 & 0,4)
	7		50		(0,6 & 0,4)
	8		70		(0,6 & 0,4)
			90		

mengacu pada entitas yang sama. Proses seleksi *top-k* diilustrasikan seperti Kode 4.2. Semakin banyak jumlah hasil entitas yang memiliki kemiripan atau kesamaan tersebut, maka model tersebut semakin baik. Hal tersebut menunjukkan bahwa koleksi entitas tersebut terletak pada embedding vektor yang berdekatan. Hal ini diilustrasikan seperti visualisasi dari hasil embedding model pada Gambar 4.7. Dari Gambar 4.7 dapat dilihat bahwa terdapat cluster-cluster yang dikategorikan berdasarkan Institusi Halal dengan warna, selain itu juga terdapat cluster yang terdiri dari dua warna atau lebih. Cluster dengan lebih dari satu warna menunjukkan adanya kemiripan entitas yang berasal dari Institusi Halal yang berbeda.

```
listTotalMatch = []
listSimilarEntities = getSimilarEntitiesByCosineSimilarity(
  entitySource)
threshold = 0.87
similarityJaro = getSimilarityJaro(eSourceName, eTargetName)
similariyCosine = getSimilarityCosine(eSourceName, eTargetName)
similarityJaccard = getSimilarityJaccard(eSourceName,
  eTargetName)
if similarityJaro > threshold or similarityCosine > threshold or
    similarityJaccard > threshold:
    listTotalMatch.append(entityTarget)
```

Kode 4.2: Pseudocode Proses Seleksi Hasil Model Embedding



Gambar 4.7: Visualisasi Embedding Menggunakan Node2vec

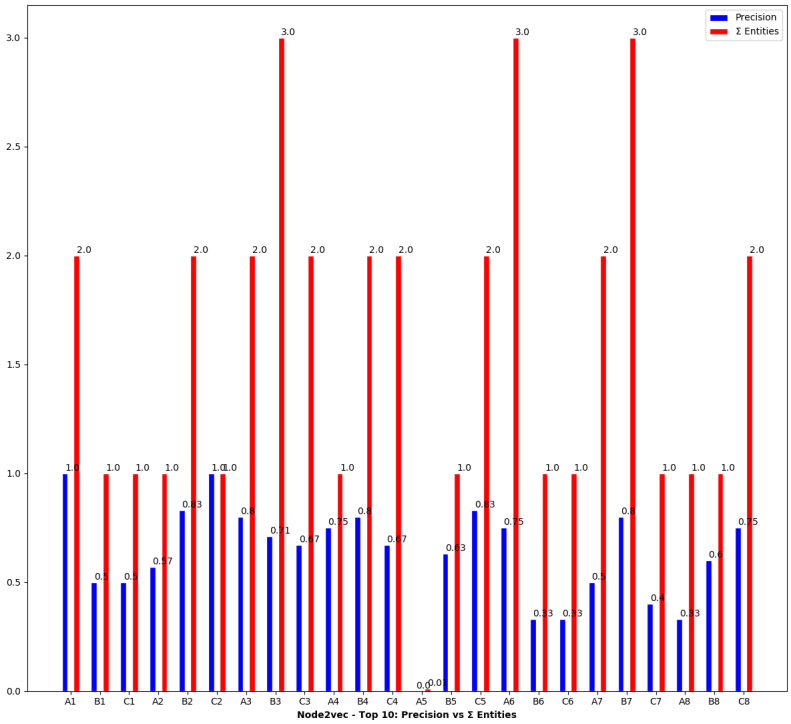
Sebagai contoh proses similarity untuk menentukan model terbaik dilakuk-

an proses similarity pada 7 entitas asal (*entity source*) yang telah ditentukan sebagai entitas yang memiliki kemiripan dengan entitas lain yang sama pada dataset yaitu *Energen rasa kacang hijau*, *Energen rasa vanila*, dan *Nissin Wafers Coklat*. Entitas tersebut dipaparkan seperti Tabel 4.3. Hasil pemilihan model yang terbaik ditentukan oleh jumlah luaran entitas yang sama, dibandingkan dengan jumlah luaran entitas dan presisi model dalam melakukan klasifikasi seperti ilustrasi pada Gambar 4.8 dan Gambar 4.9. Dari kedua gambar tersebut, maka didapatkan model terbaik yaitu model dengan kode B7 pada Top 10 dengan nilai *Precision* sebesar 0,8 dan jumlah entitas yang sama sejumlah 3.

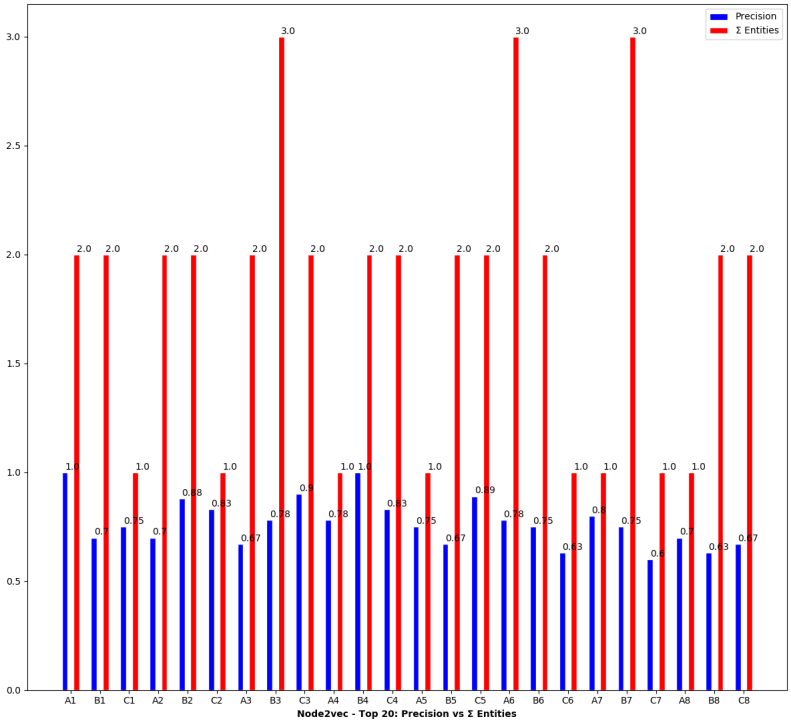
Tabel 4.3: Daftar Contoh Entity Similarity dengan Threshold 0.87

No	Entity Source	Luaran Entitas	$\sum$ Entitas	$\sum$ Entitas Sama
1	Cereal energen kacang hijau	Cereal Energen Jagung Manis, Cereal Energen Jahe, *Energen rasa kacang hija*u	3	1
2	Energen rasa kacang hijau	Cereal Energen Vanilla, Energen Less Sugar Vanilla	3	0
3	Energen rasa vanila	Energen Rasa Kacang Hijau, Energen Less Sugar Cokelat	2	0
4	Energen less sugar vanilla	Cereal Energen Jahe, *Cereal Energen Vanilla*	2	1
5	Nissin Wafers Coklat	*Nissin Wafer Krim Coklat*	1	1
6	Nissin Wafer Krim Coklat	*Nissin Wafers Coklat*	1	1
7	Nissin wafer malt cokelat	*Nissin Wafer Krim Coklat*	1	1
Total			13	5

Kemudian, hasil luaran entitas dari model ini diseleksi kembali hingga ditemukan *top-k* entitas yang memiliki kemiripan (*similarity*) tertinggi. Entitas-entitas tersebut selanjutnya disebut sebagai *entity profiles*. *Entity profiles* akan dilakukan proses *entity resolution* pada tahap selanjutnya dengan cara melakukan seleksi



Gambar 4.8: Performa Model Luaran Top-10 Luaran: Node2vec



Gambar 4.9: Performa Model Luaran Top-20 Luaran: Node2vec

kembali terhadap *top-k* dengan menghitung skor string similarity antara nama entitas asal *entity source* dengan nama entitas target *entity target*. Proses seleksi *top-k* diilustrasikan seperti Kode 4.3.

```
listEntityProfilesCandidate = []
threshold = 0.87
similarity = getSimilarityJaro(eSourceName, eTargetName)
if similarity > threshold:
  listEntityProfilesCandidate.append(entityTarget)
```

Kode 4.3: Pseudocode Proses Seleksi *top-k* Entity Profiles

4.2.2 Entity Resolution dan Link Prediction

Entity resolution memiliki tujuan untuk mengetahui beberapa entitas yang sebenarnya mengacu pada entitas yang sama di dunia nyata. Proses entity resolution dilakukan dengan cara melakukan string similarity antara entitas asal dengan entitas target yang berasal dari *entity profiles*. Atribut yang digunakan untuk melakukan *entity resolution* pada dataset Halal Nutrition Food yaitu atribut nama produk (*product name*), nama manufaktur (*manufacturer name*) dan nama komposisi produk (*ingredients name*). Pada dataset terdapat 253.582 dari 310.414 atau 81% produk memiliki komposisi produk, sedangkan sisanya tidak memiliki produk. Sehingga proses *string similarity* dilakukan dengan menggunakan beberapa kombinasi seperti Tabel 3.1, yaitu 0.8 dan 0.2 untuk perbandingan produk yang tidak memiliki *ingredients*, sedangkan 0.5, 0.2, dan 0.3 untuk perbandingan produk yang memiliki *ingredients*.

Hasil dari proses similarity selanjutnya akan menentukan entitas-entitas yang akan dilakukan resolusi. Jika hasil similarity melebihi thresold, entitas-entitas tersebut dianggap sebagai pasangan entitas yang merujuk kepada entitas yang sama atau memiliki kemiripan yang sangat mirip dan dihubungkan dengan relasi *owl:sameAs*. Sedangkan entitas-entitas yang memiliki hasil similarity kurang dari threshold, maka pasangan entitas tersebut dianggap sebagai pasangan entitas yang memiliki kore-

lasi namun tidak merujuk kepada entitas yang sama atau memiliki kemiripan yang tidak terlalu mirip. Setelah proses similarity, didapatkan jumlah relasi *owl:sameAs* dengan *rdfs:seeAlso* pada pendekatan Node2vec seperti Tabel 4.4.

Tabel 4.4: Jumlah Masing-masing Relasi pada Metode Node2vec

Metode	owl:sameAs	rdfs:seeAlso
Node2vec	36.200	640

4.2.3 Entity Ranking

Pada tahap ini dokumen file turtle RDF (.ttl) dari setiap entitas pada dataset akan dilakukan proses indexing, perangkingan secara independen dan dependen. Proses indexing dilakukan dengan cara membaca dokumen file .ttl pada atribut *rdfs:label*, dan *food:ingredientsListAsText*. Atribut *rdfs:label* akan memuat informasi berupa nama produk, nama manufaktur, dan nama komposisi produk. Sedangkan *food:ingredientsListAsText* akan memuat informasi komposisi produk yang langsung termuat dalam dokumen .ttl entitas produk.

Independent Ranking

Proses perangkingan independent ranking ditentukan oleh relasi *owl:sameAs* dan *rdfs:seeAlso* yang terbentuk pada setiap masing-masing entitas. Relasi-relasi ini akan dihitung nilai skor *PageRank*-nya dan menjadi acuan dalam perangkingan entitas secara independen. Semakin banyak relasi yang mengarah pada sebuah entitas, maka semakin besar kemungkinan nilai skor *PageRank*-nya. Hal ini menunjukkan bahwa entitas tersebut populer pada sebuah dataset. Nilai skor *PageRank* yang digunakan yaitu *PageRank* pada relasi *owl:sameAs* dan *rdfs:seeAlso* dengan penggunaan bobot pada masing-masing nilai skor PageRank. Query untuk melakukan perhitungan *PageRank* pada relasi *owl:sameAs* dan *rdfs:seeAlso* yaitu dengan melakukan eksekusi Kode 4.4 dan Kode 4.5 pada database Neo4j.

```
1  CALL algo.pageRank('ns1__FoodProducut', 'owl__sameAs',
```

```
  {iterations:20, dampingFactor:0.85, write: true,writeProperty:"
    pagerankOwlSameAs"})
YIELD nodes, iterations, loadMillis, computeMillis, writeMillis,
    dampingFactor, write, writeProperty
```

Kode 4.4: Query Cypher untuk Menghitung Nilai *PageRank* pada Relasi owl:sameAs

```
  CALL algo.pageRank('ns1__FoodProducut', 'rdfs__seeAlso',
  {iterations:20, dampingFactor:0.85, write: true,writeProperty:"
    pagerankRdfsSeeAlso"})
YIELD nodes, iterations, loadMillis, computeMillis, writeMillis,
    dampingFactor, write, writeProperty
```

Kode 4.5: Query Cypher untuk Menghitung Nilai *PageRank* pada Relasi rdfs:seeAlso

Berdasarkan hasil perhitungan nilai *PageRank* pada masing-masing relasi. Didapatkan hasil statistik *PagerRank* pada metode *Node2vec* dari masing-masing relasi seperti Tabel 4.5 dan Gambar 4.10. Perhitungan statistik *PageRank* dilakukan dengan mengunakan query seperti Kode 4.6 dan Kode 4.7. Berdasarkan Gambar 4.10, terdapat banyak outlier diatas quartil 2, hal ini disebabkan oleh nilai PageRank entitas lain yang memiliki nilai *PageRank* yang besar, sehingga menyebabkan nilai *PageRank* entitas lain juga semakin besar. Hal ini diilustrasikan seperti Gambar 4.11, banyak entitas lain yang memiliki skor PageRank yang tinggi dan memberikan pengaruh pada entitas yang dituju, sehingga mempengaruhi pada entitas dengan skor PageRank terbesar seperti pada entitas ”Sarimi - Gelas Mi Instan Rasa Kari Ayam” dan ”Ovaltine Slim”.

```
  MATCH(p:FoodProduct)
WHERE p.pagerankSameAs > 0.15000000000000002
RETURN avg(p.pagerankSameAs) aS avg, min(p.pagerankSameAs) as min,
    max(p.pagerankSameAs) as max, count(p) as count
```

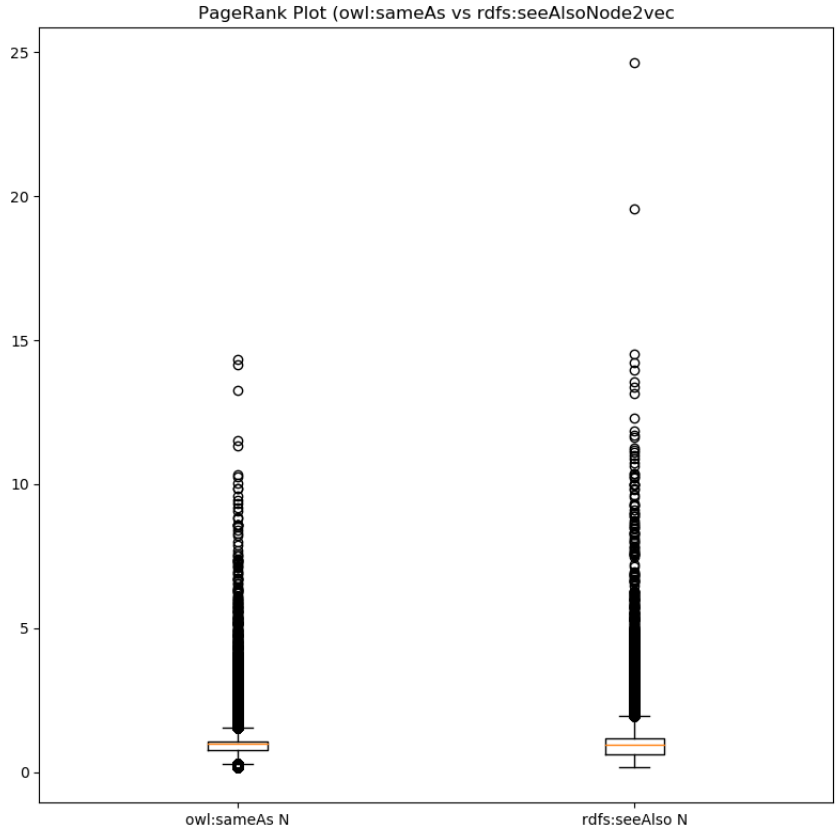
Kode 4.6: Query Cypher untuk Menghitung Data Statistik owl:sameAs

```
  MATCH(p:FoodProduct)
WHERE p.pagerankSameAs > 0.15000000000000002
RETURN avg(p.pagerankSameAs) aS avg, min(p.pagerankSameAs) as min,
    max(p.pagerankSameAs) as max, count(p) as count
```

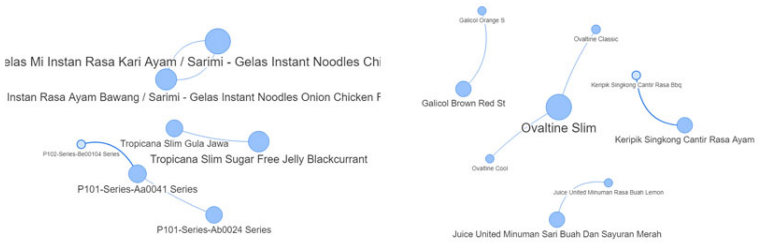
Kode 4.7: Query Cypher untuk Menghitung Data Statistik PageRank rdfs:seeAlso

Tabel 4.5: Statistik PageRank Masing-masing Relasi

Statistik	owl:sameAs Node2vec	rdfs:seeAlso Node2vec
Max	11.333	2.131
Min	0.159	0.213
Average	0.808	0.416
Count	7.107	296



Gambar 4.10: Visualisasi Boxplot Nilai PageRank Node2vec



(a) (b)

Gambar 4.11: Ilustrasi Entitas yang memiliki nilai PageRank Tertinggi pada Masing-masing Relasi (a) Relasi owl:sameAs, (b) Relasi rdfs:seeAlso

Dependent Ranking

Proses perangkingan dependent ranking dilakukan ketika proses indexing dengan melihat jumlah *term* atau kata yang muncul pada setiap dokumen file turtle (.ttl). Setiap dokumen akan memiliki skor untuk setiap term yang terdapat pada keseluruhan dokumen. Kemudian, perangkingan akhir akan ditentukan berdasarkan query pencarian oleh pengguna menggunakan metode perangkingan BM25. Selain itu, pada tahap ini juga dilakukan indexing pada skor *PageRank* pada masing-masing relasi dan nilai *LinkCount*. Hal ini dilakukan untuk menghemat resource komputasi dengan cara tidak melakukan query langsung ke database dan mempercepat proses pencarian. Field yang diindex pada setiap dokumen yaitu seperti Tabel 4.6. Query ke database Neo4j pada proses indexing dipaparkan seperti Kode 4.8, dimana *label* merupakan nama produk dari entitas.

```
1 "MATCH (m:FoodProduct)-[:certificate]->(c:HalalCertificate)-[:
     halalStatus]->(r:Resource)\n" +
2 "MATCH(m)-[:hasManufacturer]->(o:Organization)\n" +
3 "MATCH(c)-[:OrgCert]->(o1:Organization)\n" +
4 "WHERE m.label contains '" + label + "'\n" +
5 "RETURN id(m) as id, m.pagerankSameAs as pagerankSameAs, m.
     pagerankSeeAlso as pagerankSeeAlso, m.label as name, m.
     linkCount as linkCount, o.label as manufacture," +
```

Tabel 4.6: Field yang Diindex pada Proses Indexing

Field	Kode Field
Nama Produk	label
Kode Makanan	code
Manufaktur	hasManufacturer
Komposisi Produk	ingredientsListAsText
Sertifikat Halal	-certCode
	-certExp
	-certOrg
PageRank owl:sameAs	pagerankSameAs
PageRank rdfs:seeAlso	pagerankSeeAlso
LinkCount	linkCount
Id Entitas Neo4j	idNeo4j
Entitas dan sertifikat terkait dengan relasi owl:sameAs	owlSameAs
	certSameAs
Entitas dan sertifikat terkait dengan relasi rdfs:seeAlso	certrdfsSeeAlso
	certSeeAlso

```
6 "m.code as code, c.halalCode as statusCode, c.halalExp as
     statusExp, o1.label as statusOrg, r.uri as statusSource
```

Kode 4.8: Potongan Kode Java untuk Melakukan Query Data Entitas pada Proses Indexing

4.2.4 Pencarian Produk

Pada tahapan ini, sistem menampilkan hasil pencarian berdasarkan query pengguna dengan mengurutkan dokumen berdasarkan skor final yang merupakan gabungan dari skor independen dan dependen. Data yang ditampilkan berupa detail nama produk, manufaktur, komposisi produk, sertifikat dan institusi halal, ringkasan institusi halal, dan informasi lainnya seperti pada Tabel 4.6. Sebagai contoh, pada dilakukan pencarian dengan query ”medicine halal”, didapatkan hasil seperti Tabel 4.7 dengan *Rs* yaitu *raw score*, *Fs* yaitu *final score*, *Qs* yaitu query score (BM25Similarity Lucene), dan *Ss* yaitu static score (PageRank dan LinkCount) yang dihasilkan dari proses *entity resolution* dan *link prediction*. Penggunaan static score terdiri dari beberapa sub-skor yaitu PageRank owl:sameAs, PageRank rdfs:seeAlso, LinkCount, dan jumlah sertifikat pada produk dan yang terhubung

dengannya. Masing-masing sub-skor dari *static score* ini dilakukan pembobotan dan total keseluruhan dinormalisasi menggunakan fungsi *log* yang selanjutnya akan mewakili skor static score. Selanjutnya *static score* dihitung dan dikombinasikan dengan *query score* untuk menghasilkan *final score* yang akan menjadi penentu rangking dari setiap dokumen.

Penggunaan pembobotan dan normalisasi dilakukan agar skor dari *static score* tidak melebihi nilai skor dari *query score*. Nilai *static score* yang terlalu tinggi akan mengacaukan final score dan berakibat pada relevansi hasil pencarian yang buruk. Pembobotan pada masing-masing sub-skor berkontribusi terhadap besarnya hasil akhir static score, bobot dari masing-masing sub-skor yaitu (0.25*pagerankOwlSameAs+0.1*pagerankRdfsSeeAlso+0.1*linkCount+0.55*certificateCount)+1.0, penambahan angka 1 dilakukan untuk menangani jika tidak ada sub-skor yang berkontribusi atau bernilai 0, sehingga nilai hasil normalisasi *static score* akan menjadi *Infinity*. Angka 1 dapat digantikan dengan angka > 0, agar nilai hasil normalisasi *static score* tidak bernilai *Infinity*. Selain itu, penghitungan nilai jumlah sertifikat (*certificateCount*) dilakukan penambahan nilai pada sertifikat yang memiliki label ”Haram” dan ”Halal” dengan besar penambahan nilai secara berurutan atau ”Haram” ¿ ”Halal”. Sebagai contoh, jika produk memiliki sertifikat ”Haram” maka akan ditambahkan penambahan nilai skor sebesar 1.7, sedangkan ”Halal” sebesar 1.3, ”Musbooh” sebesar 1.0, 0 jika tidak memiliki sertifikat. Hal ini bertujuan agar produk yang memiliki sertifikat ”Haram” memiliki rangking yang lebih tinggi. Sebagai dampaknya, pengguna akan lebih sadar terhadap produk yang memiliki label sertifikat ”Haram”.

Sebagai perbandingan, untuk melihat pengaruh dari masing-masing komponen *scoring* yaitu *Qs* dan *Ss*, dilakukan percobaan dengan melihat skor Final dan Posisi rangking dari setiap dokumen pada sebuah query. Percobaan pertama dengan hanya melibatkan *Qs* tanpa *Ss*, dan melibatkan keduanya. Hasil uji coba keduanya dipaparkan seperti Tabel 4.7, dimana *Fs1* dan *R1* merupakan rangking tanpa *Ss*, *Fs2* dan *R2* merupakan hasil rangking dengan melibatkan *Ss*, sedangkan Δ merupakan perubahan rangking pada saat *Fs1* menjadi *Fs2*, *r0* merupakan penilaian relevansi

pengguna pada *Fs1*, sedangkan *r2* pada *Fs2* dan bernilai relevan (1), tidak relevan (0).

Tabel 4.7: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”medicine halal”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
Cf Cough Medicine	Halal	[]	39.64	1.67	2.57	1	1	0	1	1
Ny Quil Liquid Cold Medicine For Children	Halal	[39865]	29.81	1.54	2.51	3	2	+1	1	1
Original, Cold Medicine, Effervescent Cold Running Nose	Halal	[]	27.54	1.51	2.41	5	3	+2	1	1
Labeesity Labeesity 125Mg Capsules (Labisia Pumila Standardised Extract-Traditional Medicine For Health And Weight Loss)	Halal	[]	18.91	1.35	2.25	8	4	+4	1	1
Labeesity Labeesity 60Mg Capsules ((Labisia Pumila Standardised Extract-Traditional Medicine For Health And Weight Loss)	Halal	[]	18.91	1.35	2.25	9	5	+4	1	1

Berdasarkan Tabel 4.7, diketahui bahwa terdapat beberapa perubahan rangking pada dokumen seperti yang ditunjukkan pada kolom Δ. Hal tersebut dipengaruhi oleh adanya skor statis yang berasal dari skor *PageRank* owl:sameAs dan rdfs:seeAlso serta LinkCount dan sertifikat yang terhubung pada relasi owl:sameAs.

4.3 Pengujian dan Evaluasi

Pengujian dan evaluasi dilakukan dengan cara menilai relevansi hasil pencarian berdasarkan query pengguna. Penilaian dilakukan dengan menilai setiap do-

kumen berdasarkan rangking dokumen dengan pelabelan ”relevan” / ”tidak relevan” yang ditentukan dengan penggunaan *Ground truth* dari lembaga institusi halal Muslim Consumer Group, dimana lembaga ini memiliki sertifikat label ”Halal” dan ”Haram” dengan tetap memperhatikan korelasi dari *query* pengguna dengan term yang terdapat pada dokumen. Pengujian penilaian relevansi mengikuti metode penilaian seperti Gambar 3.3 dan Gambar **??**. Proses pengujian dilakukan dengan skenario seperti Tabel 3.2. Pengujian ini melibatkan 24 query/kata kunci berupa nama produk random yang masing-masing produk memiliki sertifikat halal atau haram dari institusi halal seperti Tabel 4.8.

Tabel 4.8: Daftar Query/kata kunci Pengujian Hasil Pencarian

No	Query
1	Chicken Halal
	Chicken Haram
	Chicken
2	Cookies Halal
	Cookies Haram
	Cookies
3	Ice Cream Halal
	Ice Cream Haram
	Ice Cream
4	Medicine Halal
	Medicine Haram
	Medicine
5	Milk Tea Halal
	Milk Tea Haram
	Milk Tea
6	Sauce Halal
	Sauce Haram
	Sauce
7	Snack Halal
	Snack Haram
	Snack
8	Tobacco Halal
	Tobacco Haram
	Tobacco

Hasil perangkingan metode Node2vec dipaparkan seperti pada Tabel Tabel 4.9 hingga Tabel 4.32. Berdasarkan Tabel tersebut, dapat dilihat bahwa ter-

dapat beberapa perubahan pada *final score* ditandai dengan tanda (+) dan (-) yang mengakibatkan perubahan posisi rangking dari dokumen. Pada Tabel Tabel 4.9, Tabel 4.12, Tabel 4.13, Tabel 4.14, Tabel 4.21, Tabel 4.23, Tabel 4.27, Tabel 4.29, dapat dilihat bahwa pengaruh dari *static score* berhasil mempengaruhi rangking dari dokumen dengan menaikkan peringkat dokuman menjadi peringkat teratas dan ditunjukkan dengan tanda (+) pada kolom delta. Kemudian pada Tabel 4.18, Tabel 4.19, Tabel 4.25 *static score* juga berhasil menaikkan rangking dokumen meskipun pada posisi tidak teratas. Pada Tabel 4.20 dapat dilihat bahwa posisi dokumen dengan label ”Haram” memiliki rangking yang lebih tinggi dibandingkan dengan label ”Halal”. Dokumen yang tidak memiliki informasi lebih, seperti Tabel 4.11 dengan sertifikat ”null”/tidak diketahui akan menempati posisi rangking yang paling bawah.

Evaluasi: *Precision dan Recall (P@k)*

Pada bagian ini, dilakukan pengujian evaluasi berdasarkan Precision dan Recall dari setiap dokumen pada setiap kasus. Setiap kasus dihitung nilai Precision dan Recall berdasarkan perangkingan pada pendekatan Node2vec. Berdasarkan Gambar 4.12 dan Gambar 4.13, didapatkan bahwa sistem mampu menghasilkan hasil yang sangat bagus dilihat dari nilai Precision yang bernilai 1.0 pada recal 0 hingga 1.

Tabel 4.9: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”chicken halal”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
Chicken Barbeque	Halal	[21966]	20.21	1.38	2.33	10	1	+9	1	1
FROZEN CHICKEN - CHICKEN CHOP	Halal	[]	22.12	1.42	2.32	1	2	-1	1	1
FROZEN CHICKEN - CHICKEN RICE	Halal	[]	22.12	1.42	2.32	2	3	-1	1	1
FROZEN CHICKEN - CHICKEN SPAGHETTI	Halal	[]	22.12	1.42	2.32	3	4	-1	1	1
Figo Chicken Ball (With Chicken Slice)	Halal	[]	20.72	1.39	2.29	4	5	-1	1	1
FROZEN CHICKEN - POP CORN CHICKEN	Halal	[]	20.72	1.39	2.29	5	6	-1	1	1
Cannelloni Chicken	Halal	[]	20.21	1.38	2.28	9	7	+2	1	1
BOLOGNESE CHICKEN	Halal	[]	20.21	1.38	2.28	6	8	-2	1	1
BRAISED CHICKEN	Halal	[]	20.21	1.38	2.28	7	9	-2	1	1
BUTTER CHICKEN	Halal	[]	20.21	1.38	2.28	8	10	-2	1	1

Tabel 4.10: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”chicken haram”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
Chicken In The Baskit Original Crackers	Haram	[]	25.31	1.48	2.4	1	1	0	1	1
Kick In Chicken Taco Potato Crisp	Haram	[]	23.85	1.45	2.37	2	2	0	1	1

Tabel 4.11: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”chicken”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
FROZEN CHICKEN - CHICKEN CHOP	Halal	[]	5.49	1.03	1.93	1	1	0	1	1
FROZEN CHICKEN - CHICKEN RICE	Halal	[]	5.49	1.03	1.93	2	2	0	1	1
FROZEN CHICKEN - CHICKEN SPAGHETTI	Halal	[]	5.49	1.03	1.93	3	3	0	1	1
Figo Chicken Ball (With Chicken Slice)	Halal	[]	5.14	1.01	1.91	4	4	0	1	1
FROZEN CHICKEN - POP CORN CHICKEN	Halal	[]	5.14	1.01	1.91	5	5	0	1	1
BOLOGNESE CHICKEN	Halal	[]	5.02	0.99	1.89	7	6	+1	1	1
BRAISED CHICKEN	Halal	[]	5.02	0.99	1.89	8	7	+1	1	1
BUTTER CHICKEN	Halal	[]	5.02	0.99	1.89	9	8	+1	1	1
Cannelloni Chicken	Halal	[]	5.02	0.99	1.89	10	9	+1	1	1
Chicken Roast	null	null	5.02	0.99	1.8	6	10	-4	1	1

Tabel 4.12: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”cookies halal”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
Tender Cookies Cappuccino Cookies	Halal	[6640]	26.34	1.49	2.45	3	1	+2	1	1
Tender Cookies Hugging Cookies	Halal	[6352]	26.34	1.49	2.44	4	2	+2	1	1
COOKIES - BUTTER COOKIES	Halal	[]	28.25	1.52	2.42	1	3	-2	1	1
Cookies	Halal	[]	27.46	1.51	2.41	2	4	-2	1	1
Aice Cookies	Halal	[]	24.07	1.46	2.36	5	5	0	1	1
Alcapone Cookies	Halal	[]	24.07	1.46	2.36	6	6	0	1	1
Assorted Cookies	Halal	[]	24.07	1.46	2.36	8	7	+1	1	1
Banana Cookies	Halal	[]	24.07	1.46	2.36	9	8	+1	1	1
ALMOND COOKIES	Halal	[]	24.07	1.45	2.35	7	9	-2	1	1
Bordeaux Cookies	Halal	[]	24.07	1.45	2.35	10	10	0	1	1

Tabel 4.13: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”cookies haram”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
Chocolate Chip Cookies	Haram	[39741]	30.51	1.55	2.53	5	1	+4	1	1
Brussels Cookies	Haram	[]	33.16	1.59	2.51	1	2	-1	1	1
Tahoe Cookies	Haram	[]	33.16	1.59	2.51	2	3	-1	1	1
Animal Cookies Frosted	Haram	[]	30.51	1.56	2.48	3	4	-1	1	1
Caramel Apple Cookies	Haram	[]	30.51	1.56	2.48	4	5	-1	1	1
Double Chocolate Cookies	Haram	[]	30.51	1.56	2.48	6	6	0	1	1
Frosted Lemon Cookies	Haram	[]	30.51	1.56	2.48	7	7	0	1	1
Golden Oreo Cookies	Haram	[]	30.51	1.56	2.48	8	8	0	1	1
Iced Oatmeal Cookies	Haram	[]	30.51	1.56	2.48	9	9	0	1	1
Oatmeal Raisin Cookies	Haram	[]	30.51	1.56	2.48	10	10	0	1	1

Tabel 4.14: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”cookies”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
Tender Cookies Cappuccino Cookies	Halal	[6640]	6.55	1.1	2.06	3	1	+2	1	1
Tender Cookies Hugging Cookies	Halal	[6352]	6.55	1.1	2.06	4	2	+2	1	1
COOKIES - BUTTER COOKIES	Halal	[]	7.03	1.14	2.04	1	3	-2	1	1
Cookies	Halal	[]	6.83	1.13	2.03	2	4	-2	1	1
Aice Cookies	Halal	[]	5.98	1.07	1.97	5	5	0	1	1
Alcapone Cookies	Halal	[]	5.98	1.07	1.97	6	6	0	1	1
ALMOND COOKIES	Halal	[]	5.98	1.07	1.97	7	7	0	1	1
Assorted Cookies	Halal	[]	5.98	1.07	1.97	8	8	0	1	1
Banana Cookies	Halal	[]	5.98	1.07	1.97	9	9	0	1	1
Bordeaux Cookies	Halal	[]	5.98	1.07	1.97	10	10	0	1	1

Tabel 4.15: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”ice cream halal”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
Ice Cream Cookies And Cream	Halal	[]	42.85	1.68	2.58	1	1	0	1	1
Strawberries & Cream Ice Cream	Halal	[]	42.85	1.68	2.58	2	2	0	1	1
Banana Ice Cream	Halal	[]	40.76	1.66	2.56	3	3	0	1	1
Date Ice Cream	Halal	[]	40.76	1.66	2.56	6	4	+2	1	1
Ice Cream Alpukat	Halal	[]	40.76	1.66	2.56	7	5	+2	1	1
Ice Cream Avocado	Halal	[]	40.76	1.66	2.56	8	6	+2	1	1
Ice Cream Cheese	Halal	[]	40.76	1.66	2.56	9	7	+2	1	1
Ice Cream Chocolate	Halal	[]	40.76	1.66	2.56	10	8	+2	1	1
Chocolate Ice Cream	Halal	[]	40.76	1.66	2.56	4	9	-5	1	1
Coffee Ice Cream	Halal	[]	40.76	1.66	2.56	5	10	-5	1	1

Tabel 4.16: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”ice cream haram”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
Birthday Cake Ice Cream	Haram	[]	45.81	1.71	2.63	1	1	0	1	1

Tabel 4.17: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”ice cream”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
Ice Cream Cookies And Cream	Halal	[]	10.68	1.2	2.1	1	1	0	1	1
Strawberries & Cream Ice Cream	Halal	[]	10.68	1.2	2.1	2	2	0	1	1
Banana Ice Cream	Halal	[]	10.16	1.18	2.08	3	3	0	1	1
Chocolate Ice Cream	Halal	[]	10.16	1.18	2.08	4	4	0	1	1
Coffee Ice Cream	Halal	[]	10.16	1.18	2.08	5	5	0	1	1
Date Ice Cream	Halal	[]	10.16	1.18	2.08	6	6	0	1	1
Ice Cream Alpukat	Halal	[]	10.16	1.18	2.08	7	7	0	1	1
Ice Cream Avocado	Halal	[]	10.16	1.18	2.08	8	8	0	1	1
Ice Cream Cheese	Halal	[]	10.16	1.18	2.08	9	9	0	1	1
Ice Cream Chocolate	Halal	[]	10.16	1.18	2.08	10	10	0	1	1

Tabel 4.18: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”medicine halal”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
Cf Cough Medicine	Halal	[]	39.64	1.67	2.57	1	1	0	1	1
Ny Quil Liquid Cold Medicine For Children	Halal	[39865]	29.81	1.54	2.51	3	2	+1	1	1
Original, Cold Medicine, Effervescent Cold Running Nose	Halal	[]	27.54	1.51	2.41	5	3	+2	1	1
Labeesity Labeesity 125Mg Capsules (Labisia Pumila Standardised Extract-Traditional Medicine For Health And Weight Loss)	Halal	[]	18.91	1.35	2.25	8	4	+4	1	1
Labeesity Labeesity 60Mg Capsules ((Labisia Pumila Standardised Extract-Traditional Medicine For Health And Weight Loss)	Halal	[]	18.91	1.35	2.25	9	5	+4	1	1

Tabel 4.19: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”medicine haram”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
Soar Throat Medicine	Haram	[]	48.73	1.76	2.68	1	1	0	1	1
Ny Quil Cold Liquid Medicine For Adult	Haram	[40646]	38.9	1.66	2.64	4	2	+2	1	1

Tabel 4.20: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”medicine”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
Soar Throat Medicine	Haram	[]	9.88	1.29	2.21	2	1	+1	1	1
Cf Cough Medicine	Halal	[]	9.88	1.29	2.19	1	2	-1	1	1
Ny Quil Cold Liquid Medicine For Adult	Haram	[40646]	7.42	1.16	2.14	3	3	0	1	1
Ny Quil Liquid Cold Medicine For Children	Halal	[39865]	7.42	1.16	2.13	4	4	0	1	1
Original, Cold Medicine, Effervescent Cold Running Nose	Halal	[]	6.85	1.13	2.03	5	5	0	1	1
Medicine Benzonatate 100 Mg Capsules Generic Name For Tessalon	Mushbooh	[]	6.36	1.1	1.98	6	6	0	1	1
Blood Thinner Medicine Lovenox Injection Is Made From Pig Intestinal Mucosa And It Is Haram	Haram	[]	5.24	1.01	1.93	7	7	0	1	1
Labeesity Labeesity 125Mg Capsules (Labisia Pumila Standardised Extract-Traditional Medicine For Health And Weight Loss)	Halal	[]	4.69	0.97	1.87	8	8	0	1	1
Labeesity Labeesity 60Mg Capsules ((Labisia Pumila Standardised Extract-Traditional Medicine For Health And Weight Loss)	Halal	[]	4.69	0.97	1.87	9	9	0	1	1

Tabel 4.21: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”milk tea halal”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
Milk Tea Kopi	Halal	[13971]	37.75	1.62	2.58	8	1	+7	1	1
Milk Tea	Halal	[]	42.42	1.68	2.58	1	2	-1	1	1
Milk Tea Green Tea	Halal	[]	40.34	1.65	2.55	2	3	-1	1	1
Lychee Milk Tea	Halal	[]	37.75	1.63	2.52	3	4	-1	1	1
Milk Tea Alpukat	Halal	[]	37.75	1.63	2.52	4	5	-1	1	1
Milk Tea Bluberry	Halal	[]	37.75	1.63	2.52	5	6	-1	1	1
Milk Tea Durian	Halal	[]	37.75	1.63	2.52	6	7	-1	1	1
Milk Tea Jagung	Halal	[]	37.75	1.63	2.52	7	8	-1	1	1
Milk Tea Melon	Halal	[]	37.75	1.63	2.52	9	9	0	1	1
Milk Tea Mocca	Halal	[]	37.75	1.63	2.52	10	10	0	1	1

Tabel 4.22: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”milk tea haram”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
-	-	-	-	-	-	-	-	-	-	-

Tabel 4.23: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”milk tea”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
Milk Tea Kopi	Halal	[13971]	9.4	1.14	2.1	8	1	+7	1	1
Milk Tea	Halal	[]	10.57	1.2	2.1	1	2	-1	1	1
Milk Tea Green Tea	Halal	[]	10.05	1.17	2.07	2	3	-1	1	1
Lychee Milk Tea	Halal	[]	9.4	1.14	2.04	3	4	-1	1	1
Milk Tea Alpukat	Halal	[]	9.4	1.14	2.04	4	5	-1	1	1
Milk Tea Bluberry	Halal	[]	9.4	1.14	2.04	5	6	-1	1	1
Milk Tea Durian	Halal	[]	9.4	1.14	2.04	6	7	-1	1	1
Milk Tea Jagung	Halal	[]	9.4	1.14	2.04	7	8	-1	1	1
Milk Tea Melon	Halal	[]	9.4	1.14	2.04	9	9	0	1	1
Milk Tea Mocca	Halal	[]	9.4	1.14	2.04	10	10	0	1	1

Tabel 4.24: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”sauce halal”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
SAUCE - BBQ SAUCE	Halal	[]	23.89	1.45	2.35	1	1	0	1	1
SAUCE - BUTTERMILK SAUCE	Halal	[]	23.89	1.45	2.35	2	2	0	1	1
SAUCE - MAKHAM SAUCE	Halal	[]	23.89	1.45	2.35	3	3	0	1	1
SAUCE - MARINATION SAUCE	Halal	[]	23.89	1.45	2.35	4	4	0	1	1
SAUCE - SIGNATURE SAUCE	Halal	[]	23.89	1.45	2.35	5	5	0	1	1
SAUCE - SPAGHETTI SAUCE	Halal	[]	23.89	1.45	2.35	6	6	0	1	1
SAUCE - SUKI SAUCE	Halal	[]	23.89	1.45	2.35	7	7	0	1	1
SAUCE - YUM SAUCE	Halal	[]	23.89	1.45	2.35	8	8	0	1	1
Hoisin Sauce (Sweet Sauce)	Halal	[]	22.27	1.42	2.32	9	9	0	1	1
SAUCE - MUSHROOM SAUCE (MIX)	Halal	[]	22.27	1.42	2.32	10	10	0	1	1

Tabel 4.25: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”sauce haram”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
Tartar Sauce	Haram	[]	29.44	1.54	2.46	1	1	0	1	1
Teriyaki Sauce	Haram	[]	29.44	1.54	2.46	2	2	0	1	1
Bbq Sauce Hot & Sassy	Haram	[40084]	25.42	1.47	2.45	6	3	+3	1	1
Bbq Sauce Rich & Sassy	Haram	[41432]	25.42	1.47	2.45	7	4	+3	1	1
Carolina Barbecue Sauce	Haram	[]	27.21	1.51	2.43	3	5	-2	1	1
Memphis Bbq Sauce	Haram	[]	27.21	1.51	2.43	4	6	-2	1	1
Zesty Cocktail Sauce	Haram	[]	27.21	1.51	2.43	5	7	-2	1	1
Honey Smoke House Barbecue Sauce	Haram	[]	23.95	1.45	2.37	8	8	0	1	1
Original No. 7 Reciepe Bbq Sauce	Haram	[]	23.95	1.45	2.37	9	9	0	1	1
Thick & Tangy Original Bbq Sauce	Haram	[]	23.95	1.45	2.37	10	10	0	1	1

Tabel 4.26: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”sauce”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
SAUCE - BBQ SAUCE	Halal	[]	5.94	1.07	1.97	1	1	0	1	1
SAUCE - BUTTERMILK SAUCE	Halal	[]	5.94	1.07	1.97	2	2	0	1	1
SAUCE - MAKHAM SAUCE	Halal	[]	5.94	1.07	1.97	3	3	0	1	1
SAUCE - MARINATION SAUCE	Halal	[]	5.94	1.07	1.97	4	4	0	1	1
SAUCE - SIGNATURE SAUCE	Halal	[]	5.94	1.07	1.97	5	5	0	1	1
SAUCE - SPAGHETTI SAUCE	Halal	[]	5.94	1.07	1.97	6	6	0	1	1
SAUCE - SUKI SAUCE	Halal	[]	5.94	1.07	1.97	7	7	0	1	1
SAUCE - YUM SAUCE	Halal	[]	5.94	1.07	1.97	8	8	0	1	1
Hoisin Sauce (Sweet Sauce)	Halal	[]	5.53	1.04	1.94	9	9	0	1	1
SAUCE - MUSHROOM SAUCE (MIX)	Halal	[]	5.53	1.04	1.94	10	10	0	1	1

Tabel 4.27: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”snack halal”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
Caca Snack Keciput	Halal	[68799]	25.66	1.48	2.44	5	1	+4	1	1
Serena Snack	Halal	[]	28.83	1.54	2.44	2	2	0	1	1
Crispy Snack	Halal	[]	28.83	1.53	2.43	1	3	-2	1	1
Snack Cracker	Halal	[]	28.83	1.53	2.43	3	4	-1	1	1
Wheat Snack	Halal	[]	28.83	1.53	2.43	4	5	-1	1	1
Delfi Milky Snack	Halal	[]	25.66	1.48	2.38	6	6	0	1	1
Dodol Peter Snack	Halal	[]	25.66	1.48	2.38	7	7	0	1	1
Fish Skin Snack	Halal	[]	25.66	1.48	2.38	8	8	0	1	1
Nastar Ocy Snack	Halal	[]	25.66	1.48	2.38	9	9	0	1	1
Roasted Vegetable Snack	Halal	[]	25.66	1.48	2.38	10	10	0	1	1

Tabel 4.28: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”snack haram”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
Graham Snack Cinnamon	Haram	[]	34.75	1.61	2.53	1	1	0	1	1
Disney Freeze Graham Snack	Haram	[]	32.21	1.58	2.5	2	2	0	1	1
Disney Finding Dory Graham Snack	Haram	[]	30.13	1.55	2.47	3	3	0	1	1
New Snack Mix Sriracha Crackers	Haram	[]	30.13	1.55	2.47	4	4	0	1	1
Snack Mix Double Cheese Crackers	Haram	[]	30.13	1.55	2.47	5	5	0	1	1
Snack Saks Barnum’s Animal Crackers	Haram	[]	30.13	1.55	2.47	6	6	0	1	1

Tabel 4.29: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”snack”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
Caca Snack Keciput	Halal	[68799]	6.38	1.09	2.05	5	1	+4	1	1
Serena Snack	Halal	[]	7.17	1.15	2.05	2	2	0	1	1
Crispy Snack	Halal	[]	7.17	1.15	2.05	1	3	-2	1	1
Snack Cracker	Halal	[]	7.17	1.15	2.05	3	4	-1	1	1
Wheat Snack	Halal	[]	7.17	1.15	2.05	4	5	-1	1	1
Graham Snack Cinnamon	Haram	[]	6.38	1.1	2.02	9	6	+3	1	1
Delfi Milky Snack	Halal	[]	6.38	1.1	2.0	6	7	-1	1	1
Dodol Peter Snack	Halal	[]	6.38	1.1	2.0	7	8	-1	1	1
Fish Skin Snack	Halal	[]	6.38	1.1	2.0	8	9	-1	1	1
Nastar Ocy Snack	Halal	[]	6.38	1.09	1.99	10	10	0	1r	1

Tabel 4.30: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”tobacco halal”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

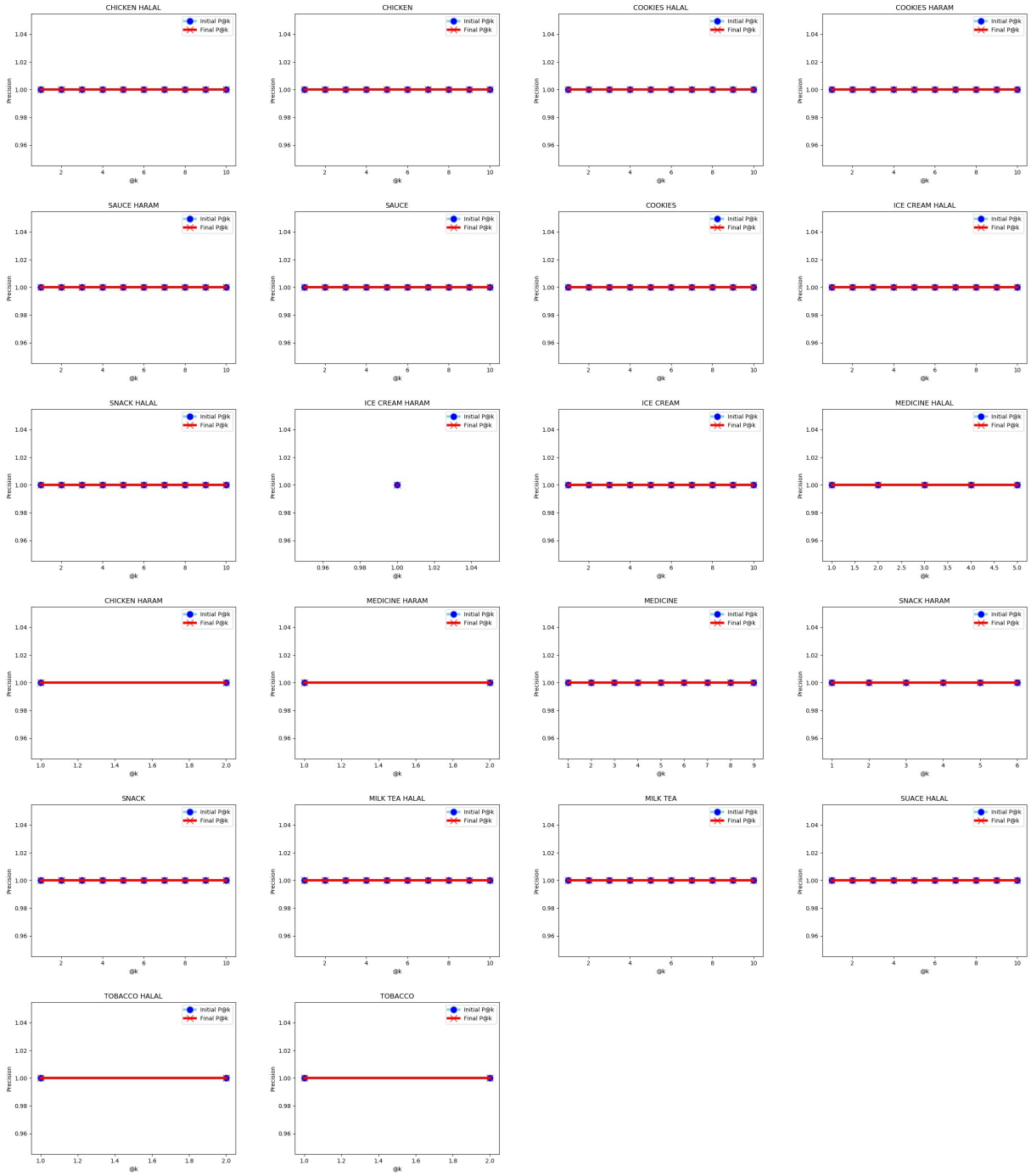
Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
Tobacco Oriental Extract	Halal	[]	45.68	1.73	2.63	1	1	0	1	1
Tobacco Rag Extract	Halal	[]	45.68	1.73	2.63	2	2	0	1	1

Tabel 4.31: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”tobacco haram”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

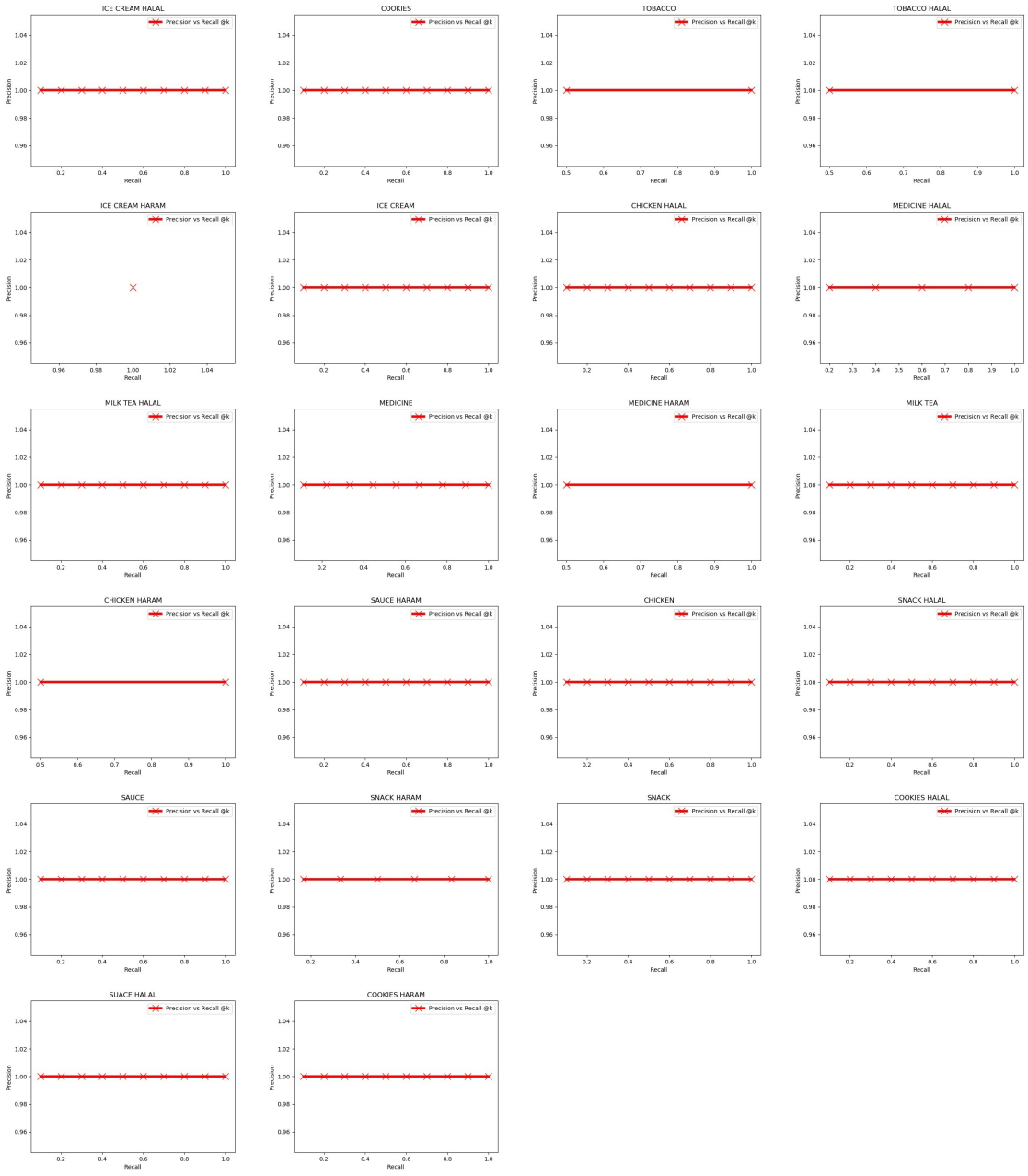
Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
-	-	-	-	-	-	-	-	-	-	-

Tabel 4.32: Perbandingan Pengaruh *Ss* dan *Qs* pada query ”tobacco”, dengan nilai *Ss* berasal dari kombinasi skor Sertifikat produk, PageRank, Linkcount, dan Jumlah Sertifikat

Nama	Status	sameAs	Fs0	Fs1	Fs2	R0	R1	Δ	r1	r0
Tobacco Oriental Extract	Halal	[]	11.39	1.35	2.25	1	1	0	1	1
Tobacco Rag Extract	Halal	[]	11.39	1.35	2.25	2	2	0	1	1



Gambar 4.12: Precision @k pada Masing-masing kasus Node2vec



Gambar 4.13: Precision vs Recal @k pada Masing-masing kasus Node2vec

Halaman ini sengaja dikosongkan

BAB 5

KESIMPULAN DAN SARAN

Pada bab ini akan disimpulkan seluruh kegiatan penelitian hingga diperoleh hasil dan pembahasan penelitian serta diberikan saran penelitian mendatang terkait dengan penelitian ini.

5.1 Kesimpulan

1. Perangkingan menggunakan pendekatan resolusi entitas ditentukan oleh beberapa faktor yang paling dominan yaitu nilai skor dependen *Qs*, kemudian diikuti oleh skor independen *Ss*.
2. Semakin besar nilai skor dependen maka semakin besar peluang muncul sebuah dokumen untuk ditampilkan pada sistem dan menjadi hasil pencarian yang relevan.
3. Semakin besar nilai skor independen akan mempengaruhi peringkat rangking dari dokumen pada hasil pencarian.
4. Penggunaan entity resolution untuk perangkingan memiliki peran yang penting dalam mempengaruhi perubahan rangking pada setiap dokumen melalui relasi-relasi yang terbentuk (owl:sameAs dan rdfs:seeAlso) pada proses *graph embedding* melalui similarity pada hasil embedding dan koten dokumen.
5. Entity resolution memberikan pengaruh yang positif untuk meningkatkan relevansi pencarian entitas pada hasil pencarian dari sistem.

5.2 Saran

1. Peningkatan relevansi perlu melibatkan berkas log data dari interaksi pengguna seperti jumlah klik dokumen pada pencarian dengan query tertentu, penggunaan cache, durasi pengguna membuka halaman dalam sebuah dokumen, dan feedback secara langsung dari pengguna.
2. Perlu ditambahkan implementasi *n-gram model* untuk mengenali komposisi dari term pada query, sehingga komposisi query menjadi semakin spesifik dan hasil pencarian semakin baik.

Halaman ini sengaja dikosongkan

DAFTAR PUSTAKA

Arnaout, H. & Elbassuoni, S. (2018), 'Effective searching of rdf knowledge graphs', *Journal of Web Semantics* **48**, 66 – 84.
URL: *http://www.sciencedirect.com/science/article/pii/S1570826817300677*

Christophides, V., Efthymiou, V. & Stefanidis, K. (2015), 'Entity resolution in the web of data', *Synthesis Lectures on the Semantic Web* **5**(3), 1–122.

Cohen, E. (n.d.), 'node2vec: Embeddings for graph data – towards data science'.
URL: *https://towardsdatascience.com/node2vec-embeddings-for-graph-data-32a866340fef*

Craswell, N., Robertson, S., Zaragoza, H. & Taylor, M. (2005), Relevance weighting for query independent evidence, *in* 'Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval', ACM, pp. 416–423.

Delbru, R., Rakhmawati, N. A. & Tummarello, G. (2010), Sindice at semsearch 2010, *in* 'Proceedings of the 19th International World Wide Web Conference, Raleigh, North Carolina, USA', Citeseer.

Delbru, R., Toupikov, N., Catasta, M., Tummarello, G. & Decker, S. (2010), Hierarchical link analysis for ranking web data, *in* 'Extended Semantic Web Conference', Springer, pp. 225–239.

Grover, A. & Leskovec, J. (2016), node2vec: Scalable feature learning for networks, *in* 'Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining', ACM, pp. 855–864.

Hinton, G. E. & Roweis, S. T. (2003), Stochastic neighbor embedding, *in* 'Advances in neural information processing systems', pp. 857–864.

Hogan, A., Decker, S. & Harth, A. (2006), 'Reconrank: A scalable ranking method for semantic web data with context'.

Indrawan, A. (2015), 'Inilah 10 negara dengan populasi muslim terbesar di dunia'.

Jaafar, H. S., Endut, I. R., Faisol, N. & Omar, E. N. (2011), 'Innovation in logistics services–halal logistics'.

Kettani, H. (2010), 2010 world muslim population, *in* 'proceedings of the 8th Hawaii Internafional Conference on Arts and Humanifies', pp. 12–16.

Lal, M. (2015), *Neo4j graph data modeling*, Packt Publishing Ltd.

Latifi, S. & Nematbakhsh, M. (2014), Query-independent learning to rank rdf entity results of sparql queries, *in* '2014 4th International Conference on Computer and Knowledge Engineering (ICCKE)', pp. 297–301.

Lembaga Pengkajian Pangan Obat-obatan dan Kosmetika MUI (2019).
URL: *https://www.halalmui.org/mui14*

Maaten, L. v. d. & Hinton, G. (2008), 'Visualizing data using t-sne', *Journal of machine learning research* **9**(Nov), 2579–2605.

Page, L., Brin, S., Motwani, R. & Winograd, T. (1999), The pagerank citation ranking: Bringing order to the web., Technical report, Stanford InfoLab.

Papadakis, G., Ioannou, E., Palpanas, T., Niederee, C. & Nejdl, W. (2013), 'A blocking framework for entity resolution in highly heterogeneous information spaces', *IEEE Transactions on Knowledge and Data Engineering* **25**(12), 2665–2682.

Papadakis, G., Koutrika, G., Palpanas, T. & Nejdl, W. (2014), 'Meta-blocking: Taking entity resolutionto the next level', *IEEE Transactions on Knowledge and Data Engineering* **26**(8), 1946–1960.

Piao, G. & Breslin, J. G. (2018), A study of the similarities of entity embeddings learned from different aspects of a knowledge base for item recommendations, *in* 'European Semantic Web Conference', Springer, pp. 345–359.

Produk Bersertifikasi Halal di Indonesia Baru 20 Persen, Malaysia Sudah 90 Persen (2018).
URL: *http://www.tribunnews.com/nasional/2014/03/07/produk-bersertifikasi-halal-di-indonesia-baru-20-persen-malaysia-sudah-90-persen*

Rakhmawati, N. A., Fatawi, J., Najib, A. C. & Firmansyah, A. A. (2019), 'Linked open data for halal food products', *Journal of King Saud University - Computer and Information Sciences* .
URL: *http://www.sciencedirect.com/science/article/pii/S1319157818312680*

Rezai, G., Mohamed, Z. & Shamsudin, M. N. (2012), 'Assessment of consumers' confidence on halal labelled manufactured food in malaysia', *Pertanika Journal of Social Science & Humanity* **20**(1), 33–42.

Robertson, S., Zaragoza, H. et al. (2009), 'The probabilistic relevance framework: Bm25 and beyond', *Foundations and Trends® in Information Retrieval* **3**(4), 333–389.

Schreiber, G., VU University Amsterdam, Raimond, Y. & BBC (2014), 'RDF 1.1 primer'.
URL: *https://www.w3.org/TR/rdf11-primer/#section-Introduction*

Sham, R., Rasi, R. Z., Abdamia, N., Mohamed, S. & Bibi, T. T. (2017), 'Halal logistics implementation in malaysia: A practical view', *IOP Conference Series: Materials Science and Engineering* **226**, 012040.
URL: *https://doi.org/10.1088%2F1757-899x%2F226%2F1%2F012040*

Shannon, C. E. (1948), 'A mathematical theory of communication', *Bell system technical journal* **27**(3), 379–423.

Simonini, G., Gagliardelli, L., Bergamaschi, S. & Jagadish, H. (2019), 'Scaling entity resolution: A loosely schema-aware approach', *Information Systems* **83**, 145 – 165.
URL: *http://www.sciencedirect.com/science/article/pii/S0306437918304083*

Teufel, S. (n.d.), 'Lecture 5: Evaluation - information retrieval computer science tripos part ii'.
URL: *https://www.cl.cam.ac.uk/teaching/1415/InfoRtrv/lecture5.pdf*

Tonon, A., Catasta, M., Demartini, G., Cudré-Mauroux, P. & Aberer, K. (2013), Trank: Ranking entity types using the web of data, *in* 'International semantic web conference', Springer, pp. 640–656.

Tonon, A., Catasta, M., Prokofyev, R., Demartini, G., Aberer, K. & Cudré-Mauroux, P. (2016), 'Contextualized ranking of entity types based on knowledge graphs', *Journal of Web Semantics* **37**, 170–183.

Webopedia (2016), 'What is semantic web? webopedia definition'.
URL: *http://www.webopedia.com/TERM/S/Semantic_Web.html*

Zezula, P. & Sedmidubsky, J. (n.d.), 'Advanced search techniques for large scale data analytics'.
URL: *http://disa.fi.muni.cz*

Halaman ini sengaja dikosongkan

BIODATA PENULIS

Penulis lahir di Blitar pada tanggal 9 Desember 1996. Merupakan anak kedua dari 3 bersaudara dan telah menempuh pendidikan formal yaitu; MI Roudlotut Tholibin Ringinanom, MTsN Kunir Wonodadi Blitar, dan SMA Negeri 1 Srengat Blitar dan pendidikan non formal melalui Madrasah Diniyah di kampung halaman.

Pada tahun 2014 melanjutkan pendidikan di Jurusan Sistem Informasi FTIK - Institut Teknologi Sepuluh Nopember (ITS) Surabaya dan terdaftar sebagai mahasiswa dengan NRP 5214100057. Selama menjadi mahasiswa penulis aktif mengikuti kegiatan ormawa Himpunan Mahasiswa Sistem Informasi dan Unit Kegiatan Mahasiswa Cinta Rebana ITS.

Penulis meraih berbagai prestasi baik dalam bidang hardskill maupun softskill, antara lain Finalis Gemastik 9 UI pada cabang Pengembangan Perangkat Lunak, Juara 1 MTQ Cabang Desain Aplikasi Al-Qur'an tingkat ITS dan Penulis pernah menjadi ketua UKM terbesar di ITS yaitu UKM Cinta Rebana ITS periode 2016-2017.

Pada tahun keempat karena penulis tertarik dengan bidang desiminasi informasi, maka penulis mengambil bidang minat Laboratorium Akuisisi Data dan Diseminasi Informasi (ADDI), lebih tepatnya lagi yaitu pada konsentrasi *Information Retrieval*. Selanjutnya, penulis melanjutkan studi untuk mendapatkan gelar Magister di Departemen Sistem Informasi, Institut Teknologi Sepuluh Nopember dengan NRP 05211850010006 dan mengambil bidang minat dan konsentrasi yang sama.

Selain itu, penulis juga menekuni profesi sebagai Internet Marketer lebih tepatnya lagi menjadi Android Publisher. Penulis memiliki sebuah private project yaitu Ahmad Studio yang fokus mengembangkan berbagai aplikasi berbasis web dan android. Penulis memiliki website portofolio yang dapat diakses di http://ahmadchoirunnajib.github.io dan dapat dihubungi melalui email ahmadchoirunnajib@gmail.com.