

**SISTEM TANYA JAWAB MENGGUNAKAN METODE
DEEP GRAPH CONVOLUTINALNEURAL
NETWORK (DGCNN)**

SKRIPSI

**Oleh:
DINDA OCKTA NOORYAWATI
NIM. 14650036**



**JURUSAN TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2019**

HALAMAN PENGAJUAN

**SISTEM TANYA JAWAB MENGGUNAKAN METODE
DEEP GRAPH CONVOLUTINALNEURAL
NETWORK (DGCNN)**

SKRIPSI

**Diajukan kepada:
Fakultas Sains dan Teknologi
Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang
Untuk Memenuhi Salah Satu Persyaratan Dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)**

**Oleh:
DINDA OCKTA NOORYAWAYI
NIM. 14650036**

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2019**

LEMBAR PERSETUJUAN

**SISTEM TANYA JAWAB MENGGUNAKAN METODE
DEEP GRAPH CONVOLUTINAL NEURAL
NETWORK (DGCNN)**

SKRIPSI

Oleh :
DINDA OCKTA NOORYAWATI
NIM. 14650036

Telah Diperiksa dan Disetujui untuk Diuji

Tanggal : 29 November 2019

Dosen Pembimbing I

Dosen Pembimbing II



Fatchurrohman, M.Kom
NIP. 19700731 200501 1 002

A'la Syauqi, M.Kom
NIP. 19771201 200801 1 007

Mengetahui,

Ketua Jurusan Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Sahyo Crysdian
NIP. 19740424 200901 1 008

LEMBAR PENGESAHAN

SISTEM TANYA JAWAB MENGGUNAKAN METODE
DEEP GRAPH CONVOLUTINAL NEURAL
NETWORK (DGCNN)

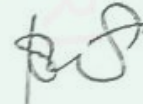
SKRIPSI

Oleh :
DINDA OKTA NOORYAWTI
NIM. 14650036

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Pada Tanggal Desember 2019

Susunan Dewan Penguji

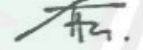
Penguji Utama : Prof. Dr. Suhartono, M.Kom
NIP. 19680519 200312 1 001

()

Ketua Penguji : Fajar Rohman Hariri, M.Kom
NIP. 19890515 201801 1 001

()

Sekretaris Penguji : Fatchurrochman, M.Kom
NIP. 19700731 200501 1 002

()

Anggota Penguji : A'la Syauqi, M.Kom
NIP. 19771201 200801 1 007

()

Mengetahui,

Ketua Jurusan Teknik Informatika

Fakultas Sains dan Teknologi

Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Cahyo Crysdian

NIP. 19740424 200901 1 008

HALAMAN PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini,

Nama : Dinda Ockta N

NIM : 14650036

Jurusan : Teknik Informatika

Fakultas : Sains dan Teknologi

Menyatakan dengan ini bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya, bukan merupakan pengambilan tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber-sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 29 November 2019

Yang membuat pernyataan,



Dinda Ockta N
NIM. 14650036

MOTTO

Never be afraid to get out of your comfort zone

Because

وُسْعَهَا إِلَّا نَفْسًا اللَّهُ يُكَلِّفُ لَا

Allah does not burden anyone except according to his/her ability

Al Baqarah: 286



KATA PENGANTAR

Segala puji dan syukur kehadirat Allah SWT atas Berkah, Rahmat dan Hidayah-Nya yang senantiasa dilimpahkan kepada penulis, sehingga bisa menyelesaikan skripsi sebagai syarat untuk menyelesaikan Program Sarjana (S1) pada Program Sarjana Fakultas Sains dan Teknologi Jurusan Teknik Informatik Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Dalam penyusunan skripsi ini banyak hambatan serta rintangan yang penulis hadapi namun pada akhirnya dapat melaluinya berkat adanya bimbingan dan bantuan dari berbagai pihak baik secara moral maupun spiritual. Untuk itu pada kesempatan ini penulis menyampaikan ucapan terimakasih kepada:

1. Orang tua saya Bapak Hadarsono, Ibu Ummi Robithoh, Bapak Achmad Nur, Ibu Rianah telah memberikan doa dan dukungan selama proses pembuatan skripsi.
2. Bapak Fatchurrochman, M.Kom dan Bapak A'la Syauqi, M.Kom selaku Dosen Pembimbing yang telah bersedia meluangkan waktu untuk memberikan arahan selama penyusunan skripsi.
3. Seluruh jajaran Dosen dan Staf jurusan Teknik Informatika yang telah mendidik dan memberikan arahan kepada saya.
4. Sahabat saya yaitu Sifa, Fitri, Putri, Yuli, Mia yang selalu menemani suka dan duka semasa kuliah saya.
5. Teman-teman dari Alim Indev dan jurusan Teknik Informatika yang selalu memberikan ilmu tentang teknologi.
6. Inside of Art yang selalu menginspirasi saya.
7. Semua pihak yang tidak dapat disebutkan satu persatu yang telah membantu memberikan dukungan.

Penulis mohon maaf atas segala kesalahan yang pernah dilakukan. Semoga skripsi ini dapat memberikan manfaat untuk mendorong penelitianpenelitian selanjutnya.

Malang, Desember 2019

Dinda Ockta N

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
LEMBAR PERSETUJUAN	iii
LEMBAR PENGESAHAN	iv
HALAMAN PERNYATAAN KEASLIAN TULISAN.....	v
MOTTO	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL	xi
ABSTRAK	xiii
ABSTRACT	xiv
ملخص.....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian.....	4
1.4 Batasan Masalah.....	4
1.5 Manfaat Penelitian.....	5
1.6 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Sistem Tanya Jawab	6
2.2 EAT (<i>Expected Answer Type</i>)	6
2.3 Basis Pengetahuan	7
2.3.1 Wikipedia	7
2.3.2 Dbpedia	7
2.3.3 Babelnet.....	8

2.3.4	OpenEphyra.....	8
2.3.5	TREC	8
2.4	Deep Graph Convolutional Neural Networks (DGCNN)	8
2.4.1	Proses Graph	8
2.4.2	Feature Learning	9
2.4.3	Classification.....	12
2.5	Cosine Similarity	13
2.6	Tentang Penelitian Sebelumnya	13
BAB III DESAIN DAN PERENCANAAN SISTEM		15
3.1	Deskripsi Sistem.....	15
3.2	Perancangan Sistem.....	16
3.2.1	Preprocessing	16
3.2.2	Klasifikasi Pertanyaan.....	18
3.2.3	Ekstraksi Judul Halaman.....	24
3.2.4	Ekstraksi Jawaban dengan SPARQL dari DBpedia.....	29
3.2.5	Ekstraksi Jawaban dengan Cosine Similarity dari Wikipedia.....	31
BAB IV UJI COBA DAN PEMBAHASAN		36
4.1	Platform yang digunakan.....	36
4.2	Implementasi Sistem	37
4.3	Uji Coba Sistem.....	41
4.4	Analisa Hasil dan Pembahasan.....	64
BAB IV KESIMPULAN.....		69
5.1	Kesimpulan.....	69
5.2	Saran.....	69
REFERENSI.....		70

DAFTAR GAMBAR

Gambar 2.1 Word Graph.....	9
Gambar 2.2 Operasi <i>Convolution</i>	10
Gambar 2.3 <i>Rectified Linear Unit (ReLU)</i>	11
Gambar 2.4 <i>Average</i> dan <i>Max Polling</i>	11
Gambar 2.5 <i>Flatten</i>	12
Gambar 3.1 Diagram Sistem.....	15
Gambar 3.2 Diagram <i>Preprocessing</i> Pertanyaan.....	16
Gambar 3.3 Hasil dari Proses <i>Part of Speech (POS)</i>	17
Gambar 3.4 Flowchart Ekstraksi data di Babelnet.....	19
Gambar 3.5 Proses CNN.....	21
Gambar 3.6 Perhitungan Matriks <i>Convolutional</i> tahap 1.....	22
Gambar 3.7 Perhitungan Matriks <i>Convolutional</i> tahap 2.....	23
Gambar 3.8 Flowchart Ekstraksi Jawaban dengan Cosine Similarity	31
Gambar 3.9 Flowchart pencarian jawaban.....	32
Gambar 4.1 <i>Form</i> pertanyaan	37
Gambar 4.2 Tampilan proses <i>Preprocessing</i>	38
Gambar 4.3 <i>Form</i> untuk memasukan topik.....	38
Gambar 4.4 Hasil dari proses N-Gram.....	39
Gambar 4.5 <i>Form option</i> untuk memilih topik.....	39
Gambar 4.6 Tampilan Hasil <i>Word Graph</i>	40
Gambar 4.7 Tampilan hasil proses <i>Convolutional Neural Network (CNN)</i>	40
Gambar 4.8 Hasil Jawaban.....	41
Gambar 4.9 Grafik waktu eksekusi.....	62
Gambar 4.10 Grafik rata-rata waktu eksekusi.....	62
Gambar 4.11 Grafik rata-rata jumlah key dari Babelnet.....	63
Gambar 4.12 Grafik rata-rata jumlah kalimat dari Wikipedia.....	63

DAFTAR TABEL

Tabel 3.1 Hasil dari Proses <i>Tokenizing</i>	16
Tabel 3.2 Hasil dari Proses <i>Stop Words</i>	17
Tabel 3.3 Hasil dari proses chunked	18
Tabel 3.4 Hasil dari Proses <i>Named Entity Recognition (NER)</i>	18
Tabel 3.5 Hasil dari Proses Lemmatization	18
Tabel 3.6 Hasil Ekstraksi dari Babelnet	20
Tabel 3.7 TF dari <i>Class 37</i>	22
Tabel 3.8 IDF dari <i>Class 37</i>	22
Tabel 3.9 Perhitungan Softmax.....	23
Tabel 3.10 Hasil Pengambilan Judul dari Wikipedia berdasarkan Unigram	24
Tabel 3.11 Hasil Pengambilan Judul dari Wikipedia berdasarkan Bigram	26
Tabel 3.12 Hasil Pengambilan Judul dari Wikipedia berdasarkan Trigram	27
Tabel 3.13 <i>Template Query</i> pada SPARQL.....	29
Tabel 3.14 Hasil SPARQL berdasarkan EAT.....	30
Tabel 3.15 Proses Query SPARQL dengan kata “ <i>Notable</i> ”.....	30
Tabel 3.16 TF pada setiap paragraf.....	33
Tabel 3.17 IDF pada setiap paragraf	33
Tabel 3.18 TF setiap kalimat pada paragraf 1	33
Tabel 3.19 IDF setiap kalimat pada paragraf 1	34
Tabel 3.20 <i>Scalar</i> dan Panjang Vektor Kalimat Pada Paragraf 5	35
Tabel 3.21 Perhitungan <i>Cosine Similarity</i>	35
Tabel 4.1 Library yang digunakan	37
Tabel 4.2 Uji Coba Pertanyaan menggunakan kata tanya “ <i>How</i> ”	44
Tabel 4.3 Analisa kata tanya “ How ”	46
Tabel 4.4 Uji Coba Pertanyaan menggunakan kata tanya “ <i>What</i> ”	47
Tabel 4.5 Analisa kata tanya “ What ”.....	49
Tabel 4.6 Uji Coba Pertanyaan menggunakan kata tanya “ <i>When</i> ”.....	50
Tabel 4.7 Analisa kata tanya “ When ”	52
Tabel 4.8 Uji Coba Pertanyaan menggunakan kata tanya “ <i>Where</i> ”.....	53
Tabel 4.9 Analisa kata tanya “ Where ”.....	55

Tabel 4.10 Uji Coba Pertanyaan menggunakan kata tanya “ <i>Who</i> ”.....	56
Tabel 4.11 Analisa kata tanya “ <i>Who</i> ”	58
Tabel 4.12 Uji Coba Pertanyaan “ <i>Yes or No</i> ”.....	59
Tabel 4.13 Analisa kata tanya “ <i>Yes or No</i> ”.....	61
Tabel 4.14 Hasil <i>Precision</i> , <i>Recall</i> , dan <i>Accuracy</i>	64



ABSTRAK

Ockta N, Dinda. 2019. **Sistem tanya jawab menggunakan metode Deep Graph Convolutinal Neural Network (DGCNN)**. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing : (I) Fatchurrohman, M.Kom, (II) A'la Syauqi, M.Kom

Kata kunci: *Sistem Tanya Jawab, Deep Graph Convolutional Neural Network.*

Sejarah dibidang matematika merupakan sumber pengetahuan tentang kerja keras para matematikawan dalam menemukan, mengembangkan suatu konsep atau memecahkan sesuatu masalah. Agar pengguna dapat mengetahui sejarah matematika di masa lampau di buat sebuah sistem tanya bertujuan agar memberikan informasi atas pertanyaan pengguna. Sistem tanya jawab memiliki 3 tugas utama. Pertama proses menganalisa pertanyaan dengan menggunakan tahapan preprocessing dan Deep Graph Convolutional Neural Network (DGCNN). DGCNN merupakan algoritma jaringan saraf yang digunakan untuk pengklasifikasian pertanyaan dengan menggunakan graf untuk memperluas term dan mendapatkan class atau topik dari pertanyaan user. Kedua pengambilan dokumen menggunakan basis pengetahuan Wikipedia. Ketiga dilanjutkan dengan proses pencarian jawaban dengan menggunakan Cosine Similarity untuk pencari kemiripan antara dokumen dengan pertanyaan. Sistem tanya jawab pada penelitian ini berfokus pada tema Sejarah dibidang matematika yang memberikan informasi terkait perkembangan di masa lampau. Sehingga hasil dari penelitian ini didapatkan akurasi kebenaran jawaban sebesar 87,76% dengan jumlah percobaan sebanyak 60 uji coba.

ABSTRACT

Ockta N, Dinda. 2019. The question and answer system uses the Deep Graph Convolutinal Neural Network (DGCNN) method. Undergraduate Thesis. Department of Informatics Engineering Faculty of Science and Technology University Islamic State Maulana Malik Ibrahim Malang. Adviser: (I) Fatchurrochman, M. Kom (II) A'la Syauqi, M. Kom

Keywords: Question and answer system, Deep Graph Convolutional Neural Network.

History of mathematics is a source of knowledge about the hard work of mathematicians in discovering, developing a concept or solving a problem. In order for users to know the history of mathematics in the past, the question and answer system is intended to provide information on questions asked by users. The question and answer system has 3 main processes. First, analyze the questions using preprocessing stages and Deep Graph Convolutional Neural Network (DGCNN). DGCNN is a neural network algorithm used to classify questions by using graphs to expand terms and get classes or topics from user inquiries. Second Document retrieval using the Wikipedia Knowledge Base. Third is continued with the process of searching for answers by using Cosine Similarity to search similarities between documents with questions. The question and answer system in this research focuses on the mathematical history theme which provides information relating to the past developments. So the results of this study obtained the correctness accuracy of 87.76% with a trial number of 60 trials.

ملخص

اوكتا ، ديندا. 2019. نظام الاسئلة والاجوبه يستخدم أسلوب الشبكة العصبية كونفولوتينال الرسم البياني العميق. اطروحه. قسم هندسه المعلوماتية كلية العلوم والتكنولوجيا جامعه الدولة الاسلاميه مولانا مالك إبراهيم مالانغ.
المدرّب: (1) فاتشوروتشمان ، م. كوم (2) علاء سيقى ، م. كوم

الكلمات الرئيسية: نظام الاسئلة والاجوبه, الرسم البياني العميق التلافيفي الشبكة العصبية
تاريخ الرياضيات هو مصدر للمعرفة حول العمل الشاق للرياضيين في اكتشاف ، وتطوير مفهوم أو حل مشكله. من أجل ان يعرف المستخدمون تاريخ الرياضيات في الماضي ، فان نظام الاسئلة والاجوبه يهدف إلى توفير معلومات عن الاسئلة التي يطرحها المستخدمون. نظام الاسئلة والاجوبه لديه 3 عمليات رئيسيه. أولاً ، تحليل الاسئلة باستخدام مراحل المعالجة المسبقة والشبكة العصبية التلافيفي الرسم البياني العميق. الخوارزميه هي خوارزميه الشبكة العصبية المستخدمة لتصنيف الاسئلة باستخدام الرسوم البيانية لتوسيع المصطلحات والحصول علي فئات أو مواضيع من استفسارات المستخدم. يستخدم كل من استرداد المستند "قاعده المعارف في ويكيبيديا". ويستمر الثالث مع عمليه البحث عن إجابات باستخدام تشابه جيب التمام للبحث أوجه التشابه بين الوثائق مع الاسئلة. يركز نظام الاسئلة والاجوبه في هذا البحث علي موضوع التاريخ الرياضي الذي يوفر المعلومات المتعلقة بالتطورات السابقة. التالي فان نتائج هذه الدراسة الحصول علي دقه صحة 87.76 % مع عدد محاكمه من 60 المحاكمات

BAB I

PENDAHULUAN

1.1 Latar Belakang

Belajar ilmu tentang sejarah merupakan bagian penting yang tidak mungkin dipisahkan dari kehidupan dari masa ke masa. Dengan memahami sejarah dengan baik dan benar, dapat mengambil banyak pelajaran dan membenahi kekurangan atau kesalahan. Sebagaimana salah satu ayat di dalam Al-Quran yaitu:

وَلَكِنْ يُفْتَرَىٰ حَدِيثًا كَانَ آَمًا ۖ الْأَلْبَابِ لِأُولِي عِبْرَةٍ قَصَصِهِمْ فِي كَانَ لَقَدْ -
يُؤْمِنُونَ لِقَوْمٍ وَرَحْمَةً وَهُدًى شَيْءٍ كُلِّ وَتَفْصِيلَ يَدِيهِ بَيْنَ الَّذِي تَصَدِّقَ

Artinya: “Dalam sejarah mereka sesungguhnya ada pelajaran bagi orang-orang yang memahami. Ini bukan cerita yang dibuat-buat, tetapi membenarkan (Kitab-kitab) yang sebelumnya, menjelaskan segala sesuatu, dan rahmat bagi orang-orang yang beriman.” (Q.S Yusuf:111)

Menurut Tafsir al-Jalalain oleh Jalaluddin al-Mahalli & Jalaluddin as-Suyuthi, (Sesungguhnya pada kisah mereka itu terdapat) yang dimaksud adalah kisah-kisah para rasul (pengajaran bagi orang-orang yang mempunyai akal) orang-orang yang berakal (Ini bukanlah) Alquran ini bukanlah (cerita yang dibuat-buat) sengaja dibuat-buat (akan tetapi) tetapi (membenarkan kitab-kitab yang sebelumnya) kitab-kitab yang diturunkan sebelum Alquran (dan menjelaskan) menerangkan (segala sesuatu) yang diperlukan dalam agama (dan sebagai petunjuk) dari kesesatan (dan rahmat bagi kaum yang beriman) mereka disebutkan secara khusus dalam ayat ini mengingat hanya mereka sajalah yang dapat mengambil manfaat Alquran bukan orang-orang selain mereka.

Manusia akan menjadi berilmu manakala mampu memanfaatkan pendengaran, penglihatan, dan hatinya dengan baik. Ia harus mencari ilmu dan terus mencari, karena ilmu tidak akan pernah datang menghampirinya. Allah menjanjikan kepada manusia yang beriman dan mempunyai ilmu yaitu memudahkan jalannya menuju ke surga, diinari hati hingga menjadi terang, tenteram dan akan diangkat derajatnya. Adapun salah satu hadits yang

diriwayatkan oleh Imam Muslim dalam shahihnya, dari hadits Abu Hurairah Radhiyallahu ‘Anhu. Sesungguhnya Nabi shallallahu ‘alaihi wa sallam bersabda:

لِي رِيقًا إِهَ طِهَ بِهَلَّ اللهُ لَنَا سِرِيْقًا يَلْتَمِسُ فِيهِ عَ وَمَنْ سَلَكَ ط
تَلُوْنَ كِتَابَ يُوْتِ اللهُ يَنْ بِيْتِ مَيَّ بِوْمٍ فَمَا اجْتَمَعَ قَ الْجَنَّةِ، و
عَشِيَّتُهُمُ الرَّحْمَةُ لِيَهُمُ السَّكِيْنَةُ، وَزَلَّتْ عَ لَانِ يَنْهَمُ إِتْدَارِ سُوْنَهُ بَ اللهُ و
ذَكَرَهُمُ اللهُ فَيَمِنُ عِنْدَهُ ةٌ، وَوَحَفَّتُهُمُ الْمَلَائِكُ ،

“Barangsiapa yang menempuh suatu perjalanan dalam rangka untuk menuntut ilmu maka Allah akan memudahkan baginya jalan ke surga. Tidaklah berkumpul suatu kaum disalah satu masjid diantara masjid-masjid Allah, mereka membaca Kitabullah serta saling mempelajarinya kecuali akan turun kepada mereka ketenangan dan rahmat serta diliputi oleh para malaikat. Allah menyebut-nyebut mereka dihadapan para malaikat”.

Oleh sebab itu manusia tidak boleh bermalas-malasan agar tidak mengalami kesesatan, seperti yang dikatakan oleh Imam Al Ghazali, *“Bersungguh-sungguhlah engkau dalam menuntut ilmu, jauhilah kemalasan dan kebosanan karena jika tidak demikian engkau akan berada dalam bahaya kesesatan”.*

Sejarah merupakan sumber pengetahuan yang memberikan informasi terkait perkembangan di masa lampau untuk kemajuan di masa sekarang. Dalam hal ini, sejarah matematika juga memberikan pengetahuan bagaimana konsep matematika berkembang. Melalui sejarah matematika, kerja keras para matematikawan dalam menemukan dan mengembangkan suatu konsep atau memecahkan sesuatu bisa menjadi kisah inspiratif. (Wahyu dan Mahfudy 2016)

Dengan teknologi tersebut yang sudah berkembang yaitu internet, ilmu pengetahuan jadi semakin mudah untuk diperoleh (Breck dkk. 2000). Namun, eksplorasi data yang besar membuat pencarian informasi yang lebih spesifik menjadi rumit dan memakan waktu yang banyak. Kesulitan ini memotivasi peneliti untuk melakukan pengembangan baru, seperti Sistem Tanya Jawab (Bouziane dkk. 2015). Sistem yang memberikan suatu informasi yang spesifik sesuai permintaan pengguna. Yang mana sistem tersebut tidak seperti *Search Engine* yang memberikan satu set dokumen yang relavan (Hirschman dan Gaizauskas 2001).

Sistem tanya jawab merupakan sebuah sistem yang sering ditemui pada berbagai bidang seperti *speech recognition* (Ho, Mawardi, dan Dharmawan 2017), *computer vision* (Noh, Seo, dan Han 2015), kedokteran (Skeith dkk. 2017), pemerintahan (Rodrigo dkk. 2013), komunitas masyarakat (Zhou dkk. 2016), agama (Novreni 2012), berita (Zheng 2003), pembelajaran (Xu 2017), dll.

Sistem tersebut memiliki 3 proses inti yaitu menganalisa pertanyaan, pengambilan dokumen, dan pencari jawaban (Bouziane dkk. 2015). Untuk menganalisa sebuah pertanyaan dibutuhkan metode pengklasifikasian pertanyaan agar dapat menemukan sebuah kata kunci. Pengklasifikasian metode dapat dilakukan dengan berbagai metode. Namun pada penelitian Peng dkk (2018), mereka membandingkan beberapa metode yang ada, dengan metode terbaru yang dipublikasikan oleh Kipf dan Welling pada ICLR 2017 yaitu *Deep Graph Convolutional Neural Network* (DGCNN). Hasilnya adalah metode DGCNN mempunyai nilai mikro dan makro presisi yang lebih baik daripada metode lainnya.

Selanjutnya untuk pengambilan dokumen atau basis pengetahuan sebagai sumber jawaban, bisa didapatkan di berbagai situs web, buku, dll. Situs web yang sering digunakan pada penelitian yaitu Wikipedia (Yang dkk. 2015) sebagai sumber ilmu pembelajaran, *The New York Times* (Peng dkk. 2018) sebagai sumber berita, Imdb (Setiawan dan Bendi 2014) sebagai sumber tentang dunia perfilman, dll.

Sehingga di dalam penelitian ini diangkat judul Sistem Tanya Jawab dengan tema berbasis pembelajaran tentang ilmu pengetahuan sejarah matematika, pada basis pengetahuan Wikipedia. Dan metode yang digunakan adalah *Deep Graph Convolutional Neural Network* (DGCNN) yang mempunyai kelebihan signifikan daripada metode sebelumnya dan komputasi yang lebih efisien (Kipf dan Welling 2016).

1.2 Rumusan Masalah

Berdasarkan penjelasan pada latar belakang, maka rumusan masalah pada penelitian ini adalah

- a. Memaparkan pembuatan Sistem Tanya Jawab yang dapat berjalan di versi website.
- b. Untuk mendapatkan akurasi jawaban yang dihasilkan oleh Sistem Tanya Jawab yang dibangun dengan metode *Deep Graph Convolutional Neural Network* dan *Cosine Similarity*.

1.3 Tujuan Penelitian

Tujuan dari penelitian sistem tanya jawab dengan metode *Deep Graph Convolutional Neural Network* dan *Cosine Similarity* yaitu:

- a. Merancang dan membuat sistem tanya jawab berbasis web untuk memudahkan dalam pencarian jawaban tentang sejarah matematika
- b. Mendapatkan akurasi jawaban yang dihasilkan oleh Sistem Tanya Jawab dengan menghitung precision dan recall.

1.4 Batasan Masalah

Agar penelitian ini tidak menyimpang dari pemasalahan, maka beberapa batasan masalah pada peneltian ini yaitu:

- a. Pertanyaan yang diijinkan menggunakan Bahasa Inggris.
- b. Jenis Kalimat tanya yang diijinkan dalam bentuk tertutup atau *factoid*.
- c. Basis pengetahuan yang digunakan adalah Wikipedia berbahasa Inggris.
- d. Tema pertanyaan yang diijinkan tetang sejarah matematika dan biografi matematikawan.

1.5 Manfaat Penelitian

Dengan adanya penelitian ini, diharapkan dapat memberikan manfaat sebagai berikut :

- a. Membantu atau memberikan kemudahan bagi pengguna untuk mendapat jawaban atas pertanyaan yang dimasukkan tentang sejarah matematika.
- b. Menghasilkan jawaban yang sesuai dengan pertanyaan yang diajukan berdasarkan Text REtrieval Conference (TREC) dan SQUAD

1.6 Sistematika Penulisan

Penulisan skripsi ini disusun dalam tiga bab dengan sistematika penulisan sebagai berikut:

BAB I PENDAHULUAN

Pendahuluan membahas tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian serta sistematika penyusunan tugas akhir.

BAB II TINJAUAN PUSTAKA

Tinjauan pustaka berisi tentang informasi beberapa teori yang digunakan pada penyusunan tugas akhir. Pada bab ini, teori yang dibahas berkaitan tentang Sistem Tanya Jawab.

BAB III ANALISI DAN DESAIN SISTEM

Bab ini mengenai analisa kebutuhan sistem untuk membangun aplikasi meliputi desai dan langkah-langkah pembuatan aplikasi tanya jawab dengan metode *Deep Graph Convolutional Neural Network* dan *Cosine Similarity*.

BAB IV HASIL DAN PEMBAHASAN

Menjelaskan tentang uji coba sistem tanya jawab menggunakan metode *Deep Graph Convolutional Neural Network* dan *Cosine Similarity*.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dan saran

BAB II TINJAUAN PUSTAKA

2.1 Sistem Tanya Jawab

Sistem tanya jawab adalah sistem yang memberikan jawaban atas pertanyaan yang dimasukkan oleh pengguna, yang mana memiliki 3 tugas utama yaitu menganalisis pertanyaan, pengambilan dokumen, dan pencari jawaban (Bouziane dkk. 2015).

Di dalam tahapan penganalisaan pertanyaan, kalimat tanya dikelompokkan menjadi dua yaitu *factoid* dan *non-factoid*. *Factoid* adalah kalimat tanya yang memberikan jawaban dengan ringkas seperti deskripsi suatu entitas, identifikasi orang, tempat, waktu, dll (Iyyer dkk. 2014). Sedangkan jawaban dari *non-factoid* yang diberikan lebih luas daripada *factoid*. Tipe pertanyaan yang digunakan yaitu *why*, *definition*, dan *how*. Hasil dari analisis pertanyaan tersebut menghasilkan kata kunci atau EAT (*Expected Answer Type*).

Pengambilan dokumen adalah mencari dokumen yang relevan sesuai dengan kata kunci. Lalu untuk memperoleh jawaban terbaik komponen pencari jawaban mencocokkan pola kalimat pada dokumen yang diperoleh dengan pola-pola kalimat yang telah ditetapkan sebelumnya (Purwarianti dan Yusliani 2012).

2.2 EAT (*Expected Answer Type*)

EAT merupakan label yang diberikan untuk suatu pertanyaan. Menentukan EAT dari suatu pertanyaan tidaklah mudah karena disebabkan oleh beberapa hal yaitu (Abdiansah dan Sari 2015) :

1. Pertanyaan dapat memiliki EAT lebih dari satu misalnya "*Who made the first airplane?*" dan EAT adalah *person*, *company* dan *organization*.
2. Penentuan EAT berdasarkan wh-group (*what*, *where*, *who*, *when*, *why* dan *how*) dapat menyebabkan ambiguitas, misalnya "*How long would it take to get to Mars?*", kata *how* tidak menyatakan arti tetapi menyatakan waktu/jarak.
3. Suatu pertanyaan memiliki focus yang menentukan semantik dan arah pertanyaan, misalnya "*What is the capitol of Uruguay?*", dimana fokus pertanyaan adalah *capitol* bukan Uruguay.

4. Banyak pertanyaan memiliki satu jawaban, misalnya "*What is the age of the Queen of Holland?*" dan "*How old is the Netherlands' queen?*" yang memiliki jawaban "*since January 1938*".
5. Satu pertanyaan memiliki lebih dari satu jawaban, misalnya "*Who invented the gas laser?*" yang bisa dijawab dengan "*Ali Javan*" atau "*ascientist at MIT*".

2.3 Basis Pengetahuan

Basis pengetahuan adalah jenis basis data yang dipergunakan untuk manajemen pengetahuan. Pengelolaan pengetahuan tersebut harus dapat menjamin proses transformasi dan ekstraksi dari pengetahuan ke data dan sebaliknya dengan benar, tanpa kehilangan informasi yang ada di dalam pengetahuan-pengetahuan. (Yuliadi & Ismail, 2008).

2.3.1 Wikipedia

Wikipedia adalah proyek ensiklopedia multibahasa dalam jaringan yang bebas dan terbuka, yang dijalankan oleh Wikimedia Foundation. Nama Wikipedia berasal dari gabungan kata wiki dan encyclopedia. Wikipedia dirilis pada tahun 2001 oleh Jimmy Wales dan Larry Sanger, dan kini merupakan karya referensi paling besar, cepat berkembang, dan populer di Internet. Wikipedia bertujuan untuk memberikan ilmu pengetahuan manusia. (Jimmy, 2004)

2.3.2 Dbpedia

DBpedia adalah sebuah basis data terstruktur dari Wikipedia. DBpedia memungkinkan pengguna mencari hubungan dan properti yang berkaitan dengan sumber daya Wikipedia, termasuk tautan ke dataset lainnya (Bizer, dkk. 2009). Cara mengambil data Dbpedia dengan SPARQL yang kepanjangan dari *Simple Protocol and RDF (Resource Description Framework) Query Language* (Segaran dkk. 2009).

2.3.3 *Babelnet*

BabelNet merupakan kamus dengan multi bahasa yang berisi leksikalisasi dengan menghubungkan beberapa situs seperti ke *free-license* wordnets, OmegaWiki, Wiktionary dalam bahasa Inggris, Wikidata, FrameNet, VerbNet dan lain-lain. BabelNet meliputi 284 bahasa. BabelNet 4.0 berisi hampir 16 juta synsets dan sekitar 833 juta kata Indra (Navigli dan Ponzetto 2012).

2.3.4 *OpenEphyra*

OpenEphyra adalah sistem penjawab pertanyaan open-source. Sistem ini didasarkan pada Ephyra, yang dikembangkan oleh Nico Schlaefer dan telah berpartisipasi dalam kompetisi menjawab pertanyaan TREC. (Schlaefer, 2006)

2.3.5 *TREC*

Text REtrieval Conference (TREC) adalah serangkaian *workshops* yang sedang berlangsung yang berfokus pada daftar area pencarian informasi yang berbeda (IR). Ini disponsori bersama oleh *National Institute of Standards and Technology* (NIST) dan *Intelligence Advanced Research Projects Activity* (bagian dari kantor Direktur Intelijen Nasional), dan dimulai pada tahun 1992 sebagai bagian dari program Teks TIPSTER.

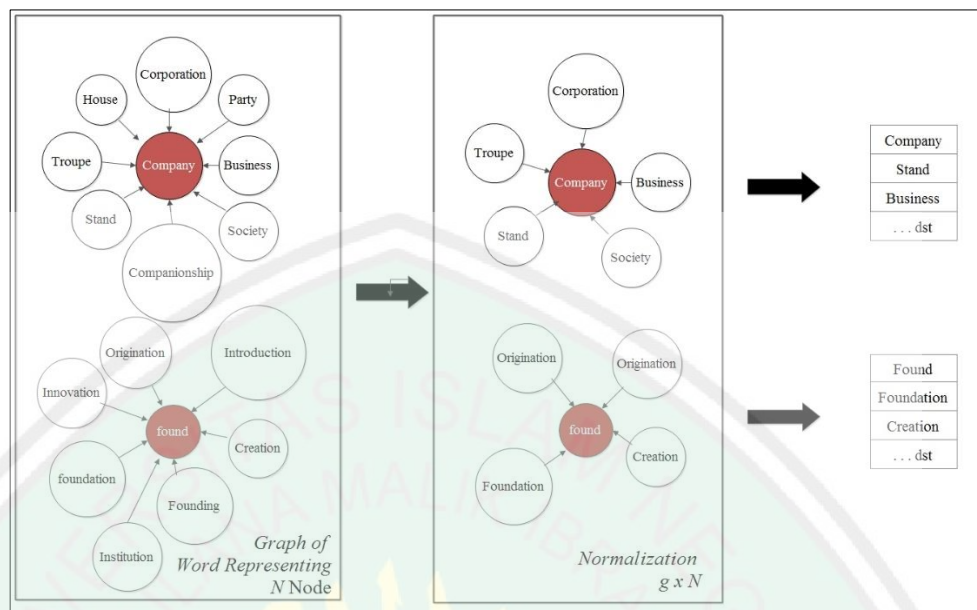
2.4 **Deep Graph Convolutional Neural Networks (DGCNN)**

DGCNN atau bisa disebut *Graph Convolutional Network* GCN adalah jaringan saraf yang beroperasi pada grafik yang diperkenalkan oleh Kipf & Welling pada ICLR 2017, termasuk salah satu algoritma *Semi-supervised* (Kipf dan Welling 2016).

2.4.1 *Proses Graph*

Merupakan proses pencarian *term* yang memiliki keterdekatan dengan *term* input. GCN mengambil sebagai input matriks fitur input $g \times N$ matriks

fitur, X , di mana N adalah jumlah node dan g adalah jumlah fitur input untuk setiap node.



Gambar 2.1 Word Graph

2.4.2 Feature Learning

Lapisan-lapisan yang terdapat dalam Feature Learning berguna untuk mentranslasikan suatu input menjadi menjadi features berdasarkan ciri dari input tersebut yang berbentuk angka-angka dalam vektor. Lapisan ekstraksi fitur ini terdiri dari Convolutional Layer dan Pooling Layer.

2.4.2.1 Convolutional Layer

Merupakan lapisan utama yang paling penting digunakan. Pada lapisan ini adanya operasi aljabar linear yang mengkalikan matriks dari *convolutional filter* atau kernel pada objek yang akan diproses. Sebelum kita menghitung dimensi atau size output dari *convolutional*, maka ada parameter yang nilainya harus ditetapkan terlebih dahulu atau bisa disebut dengan *hyperparameter* yaitu:

- **Padding atau Zero Padding**

Parameter yang menentukan jumlah pixels (berisi nilai 0) yang akan ditambahkan di setiap sisi dari input.

- **Stride (S)**

Parameter yang menentukan berapa jumlah pergeseran filter.

Pada (2.3) merupakan perhitungan untuk mengetahui ukuran yang akan dihasilkan setelah melewati lapisan *convolutional*.

$$Volume\ input = Wide_1 \times Height_1 \times Deep_1 \quad (2.1)$$

$$Volume\ Output = Wide_2 \times Height_2 \times Deep_2$$

$$Wide_2 = (Wide_1 - Filter_x) / Stride + 1$$

$$Height_2 = (Height_2 - Filter_y) / Stride + 1$$

Keterangan:

- $Height_1$: Tinggi dari suatu objek yang di-input-kan
- $Height_2$: Tinggi dari suatu objek yang di-output-kan
- $Wide_1$: Lebar dari suatu objek yang di-input-kan
- $Wide_2$: Lebar dari suatu objek yang di-output-kan
- $Filter_x$: Lebar suatu matriks yang digunakan untuk mengfilter objek
- $Filter_y$: Panjang suatu matriks yang digunakan untuk mengfilter objek
- $Stride$: Nilai yang digunakan untuk menggeser filter matriks ke kanan dan ke bawah.

Pada Gambar 2.2, merupakan cara menghitung matriks *input* dengan matriks *filter*. Matriks *input* digambarkan pada tabel berwarna biru, sedangkan matriks *filter* digambarkan pada warna merah. Hasil dari *convolution* digambarkan pada tabel berwarna ungu.

<table border="1" style="border-collapse: collapse; display: inline-table;"> <tr><td style="padding: 2px 5px;">O₁₁</td><td style="padding: 2px 5px;">O₁₂</td></tr> <tr><td style="padding: 2px 5px;">O₂₁</td><td style="padding: 2px 5px;">O₂₂</td></tr> </table>	O ₁₁	O ₁₂	O ₂₁	O ₂₂	= Convolution	<table border="1" style="border-collapse: collapse; display: inline-table;"> <tr><td style="padding: 2px 5px;">X₁₁</td><td style="padding: 2px 5px;">X₁₂</td><td style="padding: 2px 5px;">X₁₃</td></tr> <tr><td style="padding: 2px 5px;">X₂₁</td><td style="padding: 2px 5px;">X₂₂</td><td style="padding: 2px 5px;">X₂₃</td></tr> <tr><td style="padding: 2px 5px;">X₃₁</td><td style="padding: 2px 5px;">X₃₂</td><td style="padding: 2px 5px;">X₃₃</td></tr> </table>	X ₁₁	X ₁₂	X ₁₃	X ₂₁	X ₂₂	X ₂₃	X ₃₁	X ₃₂	X ₃₃	,	<table border="1" style="border-collapse: collapse; display: inline-table;"> <tr><td style="padding: 2px 5px;">F₁₁</td><td style="padding: 2px 5px;">F₁₂</td></tr> <tr><td style="padding: 2px 5px;">F₂₁</td><td style="padding: 2px 5px;">F₂₂</td></tr> </table>	F ₁₁	F ₁₂	F ₂₁	F ₂₂
O ₁₁	O ₁₂																				
O ₂₁	O ₂₂																				
X ₁₁	X ₁₂	X ₁₃																			
X ₂₁	X ₂₂	X ₂₃																			
X ₃₁	X ₃₂	X ₃₃																			
F ₁₁	F ₁₂																				
F ₂₁	F ₂₂																				

$$O_{11} = F_{11}X_{11} + F_{12}X_{12} + F_{21}X_{21} + F_{22}X_{22}$$

$$O_{12} = F_{11}X_{12} + F_{12}X_{13} + F_{21}X_{22} + F_{22}X_{23}$$

$$O_{21} = F_{11}X_{21} + F_{12}X_{22} + F_{21}X_{31} + F_{22}X_{32}$$

$$O_{22} = F_{11}X_{22} + F_{12}X_{23} + F_{21}X_{32} + F_{22}X_{33}$$

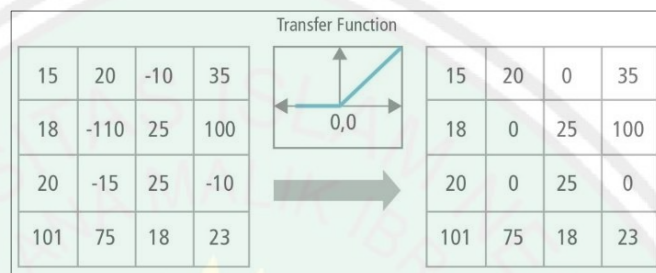
Gambar 2.2 Operasi *Convolution*

2.4.2.2 Rectified Linear Unit (ReLU)

Menerapkan fungsi aktivasi element sebagai $f(x)=\max(0,x)$, aktivasi elemen akan dilakukan saat berada di ambang batas 0. Output 0 ketika $x < 0$. Dan sebaliknya, output fungsi linear ketika $x \geq 0$ (Agarap 2018).

ReLU dapat divisualisasikan seperti

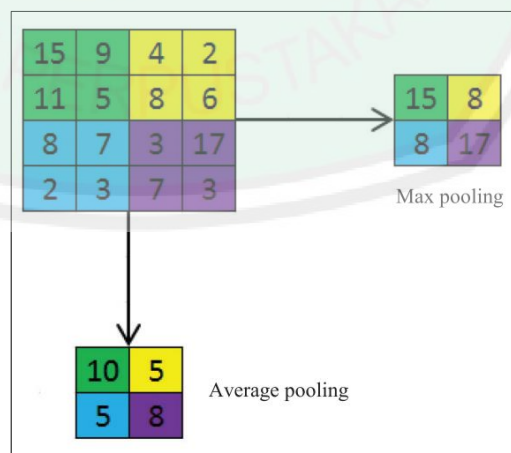
Gambar 2.3.



Gambar 2.3 Rectified Linear Unit (ReLU)

2.4.2.3 Pooling Layer

Pooling layer bertujuan untuk mengurangi dimensi dari feature map (downsampling). Pooling terdiri Max Pooling dan Average Pooling. Seperti Gambar 2.4 Max Pooling untuk menentukan nilai maksimum tiap pergeseran filter, sementara Average Pooling akan menentukan nilai rata-ratanya. (Rawat dan Wang 2017)



Gambar 2.4 Average dan Max Polling

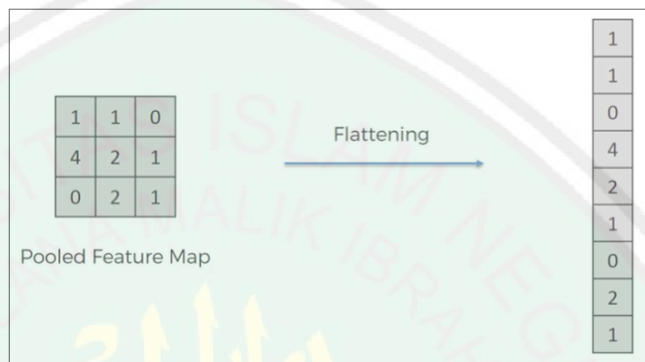
2.4.3 Classification

Lapisan ini berguna untuk mengklasifikasikan tiap neuron yang telah diekstraksi fitur pada sebelumnya. Terdiri dari:

2.4.3.1 Flatten

Membentuk ulang fitur (reshape feature map) menjadi sebuah vector agar bisa kita gunakan sebagai input dari fully-connected layer. Seperti

Gambar 2.5 merupakan bentuk dari *flatten*.



Gambar 2.5 Flatten

2.4.3.2 Fully-connected (FC)

Lapisan FC yaitu penggabungan skor akhir setiap *feature* dengan tumpukan. (Rawat dan Wang 2017)

2.4.3.3 Softmax

Softmax adalah algoritma *Logistic Regression* (Danukusumo 2017). Di bawah ini merupakan rumusnya (Rawat dan Wang 2017).

$$f_j = \frac{e^{z_j}}{\sum_k e^{z_k}} \quad (2.2)$$

Keterangan:

- j: vektor nilai suatu kelas
- k: ($j \in [1, k]$, k adalah jumlah kelas)
- e: eksponensial (e-power)

2.5 Cosine Similarity

Fungsi similarity adalah fungsi yang menerima dua buah objek dan mengembalikan nilai kemiripan (similarity) berupa bilangan riil dengan interval [0...1] (Ana dkk, 2017). Rumus Cosine similarity adalah sebagai berikut:

$$\text{Similarity}(x, y) = \cos(\theta) = \frac{x, y}{\|x\| \cdot \|y\|} \quad (2.3)$$

$$\cos(q, d_i) = \frac{\sum_{t_j} [w(t_j, q)] * [w(t_j, d_i)]}{\sqrt{\sum |w(q)|^2 * \sum |w(d_i)|^2}}$$

Dalam cos persamaan (q, d_i) adalah kosinus antara kueri dan dokumen i , $w(t_j, q)$ adalah bobot istilah t_j TF.IDF pada kueri, dan $w(t_j, d)$ adalah bobot untuk setiap istilah TF.IDF t_j dokumen berdasarkan istilah distribusi di ruang kelas. Sementara $|w(q)|$ dan $|w(d_i)|$ masing-masing panjang vektor q dan panjang dokumen vektor i .

Semakin besar hasil fungsi similarity, maka dianggap semakin mirip, dan sebaliknya. Pada fungsi yang menghasilkan nilai pada jangkauan [0...1], nilai 1 melambangkan kedua objek persis sama, sedangkan nilai 0 melambangkan kedua objek sama sekali berbeda. (Triana, Saptono, dan Sulistyono 2016)

2.6 Tentang Penelitian Sebelumnya

- Metode Deep Graph Convolutional Neural Network (DGCNN).

Lalu penelitian dari Peng, dkk (2018), mereka membandingkan beberapa metode yaitu *Support Vector Machine (SVM)*, *Hierarchical Attention Network (HAN)*, *Recurrent Convolutional Neural Network (RCNN)*, *Deep Convolutional Neural Networks DCNN*, DGCNN dan *Hierarchically Regularized Deep Graph Convolutional Neural Networks (HR-DGCNN)*. Metode yangungguli ada metode pertama HR-DGCNN, kedua DGCNN, dan yang ketiga adalah DCNN.

Menurut penelitian, DGCNN juga sebagai metode performa yang terbaik ketika dibandingkan dengan *propagation*, *semi-supervised embedding*, *skip-gram based graph embeddings*, Planetoid, dan *iterative classification algorithm (ICA)*.

- **Algoritma *Logistic Regression - SOFTMAX***

Dan menurut (Sugiarto, Kristian, & Setyaningsih, 2017) Aktivasi softmax lebih sering digunakan daripada menggunakan ReLU, sigmoid, tanh, ataupun fungsi aktivasi lainnya. Alasannya berguna mengubah output dari lapisan terakhir di neural network menjadi distribusi probabilitas dasarnya.



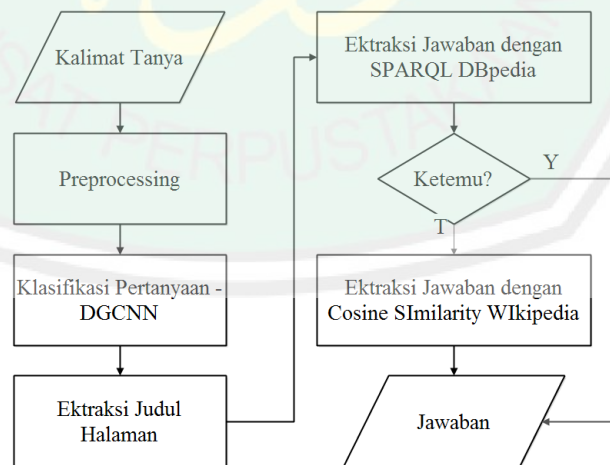
BAB III DESAIN DAN PERENCANAAN SISTEM

3.1 Deskripsi Sistem

Sistem desain pada penelitian ini digambarkan pada Gambar 3.1. ada 4 proses tahapan pada gambar tersebut. Pertama pengguna memasukan kalimat tanya kedalam sistem. Kata tanya digunakan yaitu *what, where, when, who, whom, how old, how many, how far, how fast*. Lalu diolah pada *preprocessing*. Hasil dari *preprocessing* adalah mendapatkan kata kunci dari kalimat tersebut.

Tahapan kedua memasuki metode yaitu pencarian graph atau *embedding word* untuk mendapatkan *similarity* pada kata *leksikal*. Dan, pengklasifikasian kalimat tanya dengan *Convolutional Neural Network*. Bertujuan untuk dapat mengetahui *Expected Answer Type (EAT)*. Tahapan ketiga yaitu ekstraksi judul halaman di Wikipedia. Tahapan yang melakukan penyaringan dokument untuk mencari artikel yang paling sesuai.

Tahapan keempat yaitu pencarian jawaban, dengan dilakukan pencarian judul halaman wikipedia terlebih dahulu yang sesuai. Pencarian jawaban ada 2 cara yaitu pengambilan data dari Dbpedia dengan *Query SPARQL* dan mencari kemiripan pertanyaan dengan *cosine similarity* pada halaman wikipedia.



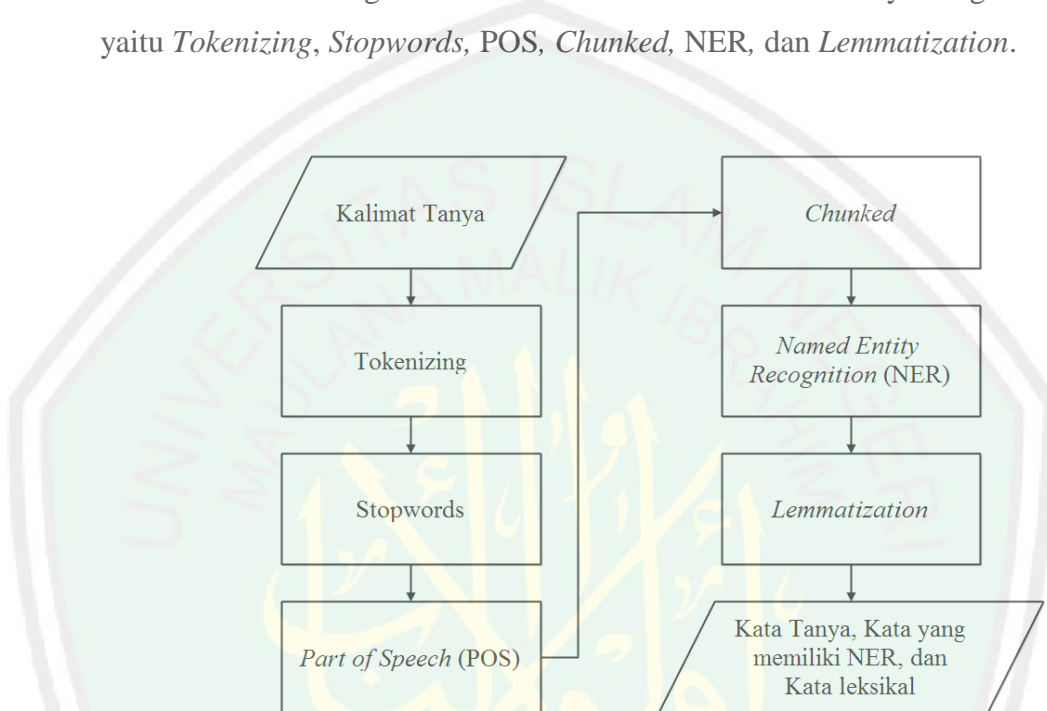
Gambar 3.1 Diagram Sistem

3.2 Perancangan Sistem

Merupakan langkah-langkah lebih detail dari gambaran umum pada deskripsi sistem.

3.2.1 Preprocessing

Proses untuk mengidentifikasi kata kunci dari kalimat tanya dengan 6 tahapan yaitu *Tokenizing*, *Stopwords*, *POS*, *Chunked*, *NER*, dan *Lemmatization*.



Gambar 3.2 Diagram *Preprocessing* Pertanyaan

3.2.1.1 Tokenizing

Tokenizing merupakan proses pemotongan kalimat tanya menjadi satuan kata atau *token*. Dan penghilangan delimiter atau simbol.

Tabel 3.1 Hasil dari Proses *Tokenizing*

Kalimat Tanya	Hasil <i>Tokenizing</i>
<i>What famous works created by Muhammad ibn Musa Al Khwarizmi?</i>	<i>“What”, “famous”, “works”, “created”, “by”, “Muhammad”, “ibn”, “Musa”, “Al”, “Khwarizmi”</i>

3.2.1.2 Stopwords

Stop words adalah merupakan proses penghilangan kata umum (*common words*) atau yang sering muncul dan dianggap kurang penting.

Tabel 3.2 Hasil dari Proses *Stop Words*

Kalimat Tanya	Hasil <i>Stopwords</i>
What famous works created by Muhammad ibn Musa Al Khwarizmi?	“What”, “famous”, “works”, “created”, “Muhammad”, “ibn”, “Musa”, “Al”, “Khwarizmi”

3.2.1.3 Part of Speech (POS)

Proses disambiguasi makna dengan mengidentifikasi arti kata dalam suatu kalimat. Ada 8 kategori POS yaitu *adjective, verb, adverb, noun, pronoun, preposition, conjunction, interjection*. (Archana dkk. 2016)

WDT	JJ	NNS	VBN	IN	NNP	NNP	NNP	NNP	NNP
What	famous	works	created	by	Muhammad	ibn	Musa	Al	Khwarizmi

Gambar 3.3 Hasil dari Proses *Part of Speech (POS)*

Keterangan:

- **WDT** : *Wh-determiner*
- **VBN** : *Verbs*
- **IN** : *Prepositions*
- **NNS** : *Common Nouns (Plural)*
- **NNP** : *Proper Nouns (Singular)*
- **JJ** : *Comparative Adjectives*

3.2.1.4 Chunked

Chunking adalah pengelompokan kata dalam kalimat menjadi frasa pendek yang bermakna. (Casteel 1988). Pemisahan atau pemotongan kata frase di *chunk* ini berdasarkan satu atau lebih kata verbs, prepositions, determiners, atau to.

Tabel 3.3 Hasil dari proses chunked

Kalimat Tanya	Hasil <i>Chuncked</i>
What famous works created by Muhammad ibn Musa Al Khwarizmi?	“What famous works created”, “Muhammad ibn Al Khwarizmi”

3.2.1.5 *Named Entity Recognition (NER)*

Pemberian nama pada entity dengan kategori telah ditentukan seperti *Person*, *Organization*, *Place*, *Date* atau *Time*, dll (Manning dkk. 2014).

Tabel 3.4 Hasil dari Proses *Named Entity Recognition (NER)*

<i>Chunked</i>	Who	<i>famous</i>	<i>works</i>	<i>created</i>	Muhammad ibn Musa Al Khwarizmi
NER	-	-	-		<i>PERSON</i>

3.2.1.6 *Lemmatization*

Lemmatization adalah sebuah proses untuk menemukan bentuk dasar dari sebuah kata (Ingason dkk. 2008).

Tabel 3.5 Hasil dari Proses Lemmatization

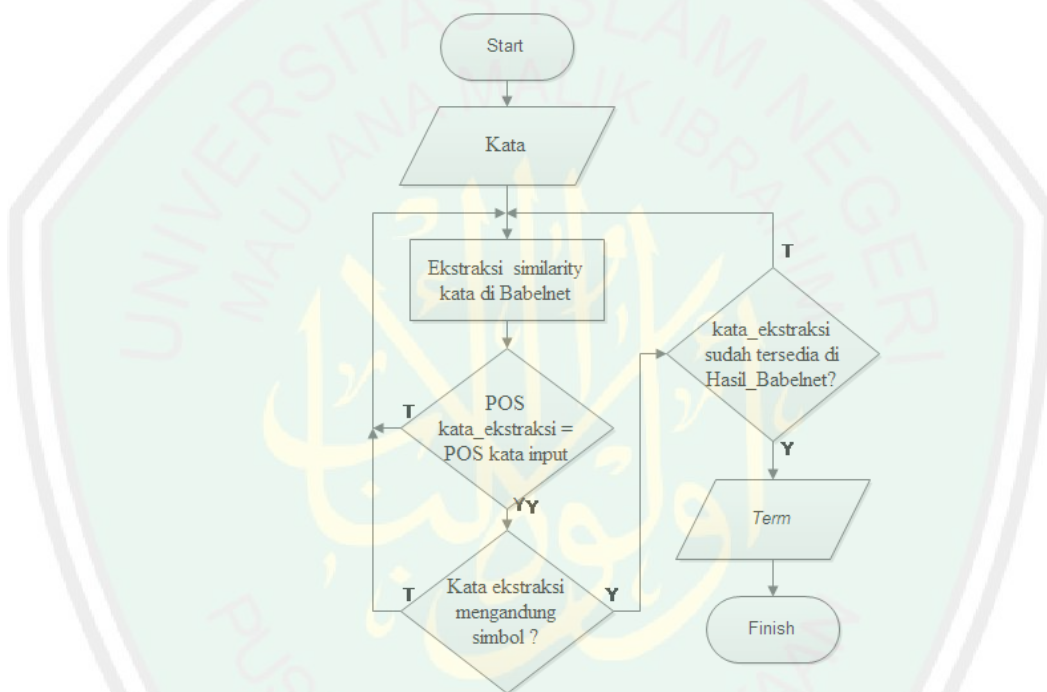
Kata Leksikal	Hasil <i>Lemmatization</i>
<i>famous</i>	<i>famous</i>
<i>works</i>	<i>work</i>
<i>created</i>	<i>create</i>

3.2.2 *Klasifikasi Pertanyaan*

Klasifikasi pertanyaan bertujuan untuk memahami arah pertanyaan yang diajukan (Biswas, Sharan, dan Kumar 2014), untuk menentukan *Expected Answer Type (EAT)* yang berupa *ontology class* yang didapatkan datanya dari OpenEphyra (Ns dan Winarko 2015). Jumlah *class* yang digunakan ada 101 *class*.

3.2.2.1 Proses Graph Word atau Embedding Word

Pencarian Graph atau *embedding word* menggunakan basis pengetahuan Babelnet. Hasilnya berupa kata yang mempunyai *similarity* atau keterdekatan dengan kata tersebut. Setelah itu data perlu di normalisasi. Mulai dari *Part of Speech* (POS) yang sama dengan kata yang dimasukkan, tidak mengandung simbol dan kata yang sudah ada tidak perlu ditambahkan. Jika data yang dihasilkan masih banyak, maka diambil 15 teratas.



Gambar 3.4 Flowchart Ekstraksi data di Babelnet

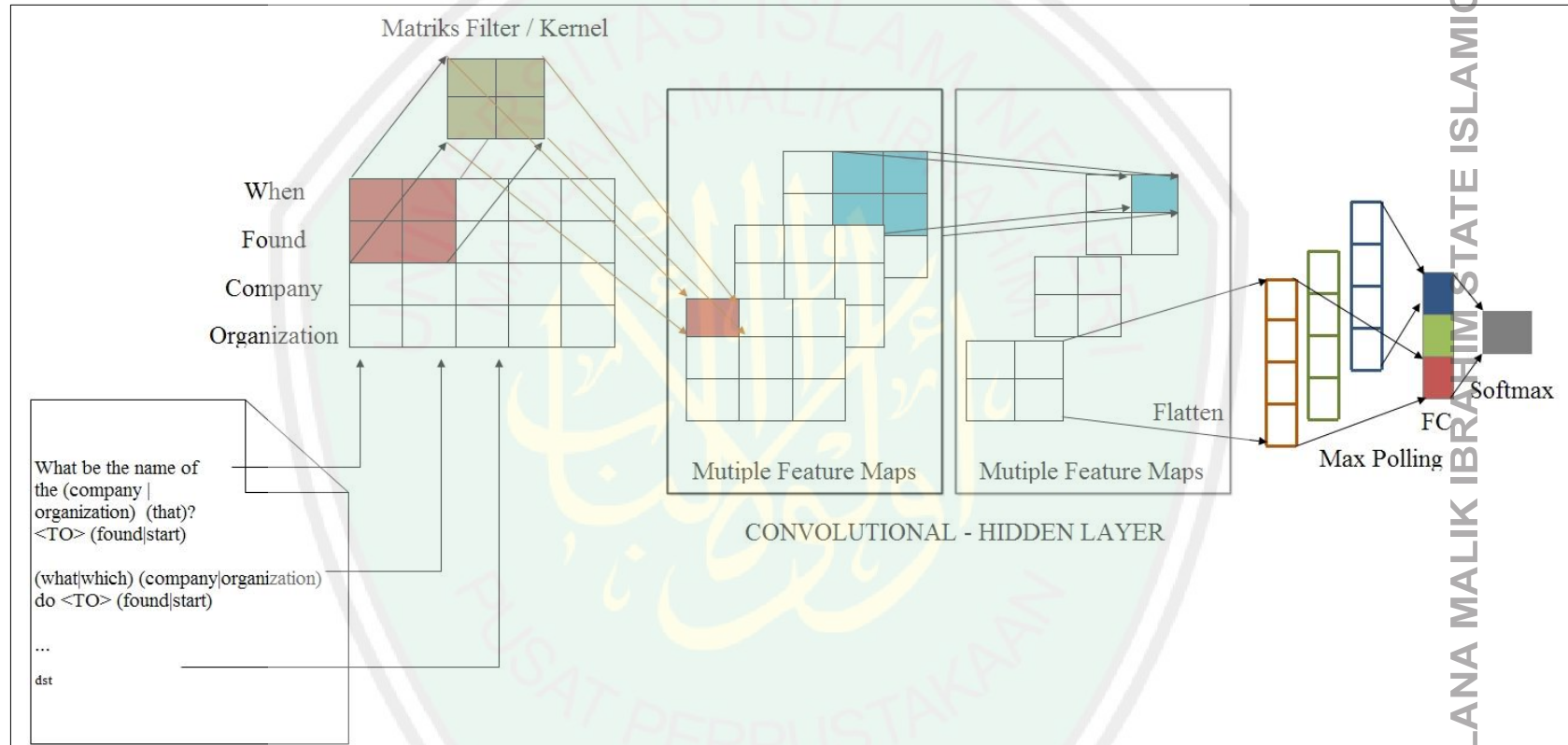
Tabel 3.6 hasil ekstraksi dari Babelnet dari kata “*famous*”, “*work*” dan “*create*”. Yang mana terdapat 9 *similarity* dari kata *famous*, 11 *similarity* dari kata *work* dan 15 *similarity* dari kata *create*.

Tabel 3.6 Hasil Ekstraksi dari Babelnet

NO	FAMOUS	WORK	CREATE
1.	Famous-Barr	plant	Invention
2.	Celebrated	works	Innovation
3.	Famed	deeds	Excogitation
4.	far-famed	workings	Conception
5.	Illustrious	Factory	Design
6.	Notable	backed	Novelism
7.	Noted	craftworks	Fangled
8.	renowned	manufactory	Fangles
9.		manufacture	Innovate
10.		workshop	Innovations
11.		whole	Innovative
12.			Innovator
13.			creativity
14.			Creative
15.			Trendsetting

3.2.2.2 Convolutional Neural Network (CNN)

Proses menemukan *class* yang sesuai. pertama pengecekan suatu term dalam setiap *class* dengan metode TF-IDF. Lalu diambil *class* yang mempunyai nilai IDF 10 teratas yaitu C5, C13, C16, C22, C48, C57, C83, C84, C87, dan C91. Lalu pemberian bobot setiap kata berdasarkan kalimat yang ada di dalam suatu dokumen agar menjadi sebuah matriks 2D. kemudian akan dikalikan dengan matriks *filter* atau *kernel*. Hasil matriks tersebut dikalikan dengan matriks *filter* atau *kernel* lagi sampai menghasilkan matriks terkecil. Lalu di flatten menjadi sebuah vector agar bisa di Max Polling. Kemudian fully-connected layer nilai Max Polling setiap *class* agar dapat di softmax untuk mendapat nilai *class* yang sesuai. Penjelasan bisa divisualisasikan pada Gambar 3.5. Dalam penelitian hasil dari softmax diambil 3 *class* yang berada di peringkat teratas.



Gambar 3.5 Proses CNN

Dibawah ini merupakan proses perhitungan CNN pada Class 37. Pertama dhitunga TF-IDF kata per-pattern kalimat yang ada di class tersebut, agar dapat membuat sebuah matrik yang dapat diconvolutional.

Tabel 3.7 TF dari Class 37

CLASS	WORD	QT	K0	K1	K2	K3
C84	WHAT	0	0	1	1	0
	FAMOUS	0	0	0	0	0
	WORK	0	0	0	0	0
	CREATE	0	0	0	0	0
	PERSON	0	0	0	0	0
	INVENTION	0	0	0	0	0
	DESIGN	0	0	1	0	1

Tabel 3.8 IDF dari Class 37

DF	LOG(N/DF)	QT	K0	K1	K2	K3
2	0.4	0	0	0.4	0.4	0
0	0.0	0	0	0	0	0
0	0.0	0	0	0	0	0
0	0.0	0	0	0	0	0
0	0.0	0	0	0	0	0
0	0.0	0	0	0.4	0	0.4

Cell yang ditebali angkanya menjadi sebuah matriks *input*-an dari CNN. Matriks tersebut dikalikan dengan matriks filter square 2x2. Matriks filter dapat diisi dengan nilai random seperti Gambar 3.6 .

0.0	0.0	0.4	0.4	0.0	x	<table border="1"> <tbody> <tr> <td>0.6</td><td>0.7</td> </tr> <tr> <td>0.3</td><td>1.0</td> </tr> </tbody> </table>	0.6	0.7	0.3	1.0	=	0.0	0.1	0.4	0.2
0.6	0.7														
0.3	1.0														
0.0	0.0	0.0	0.0	0.0			0.0	0.0	0.0	0.0					
0.0	0.0	0.0	0.0	0.0			0.0	0.0	0.0	0.0					
0.0	0.0	0.0	0.0	0.0			0.0	0.0	0.0	0.0					
0.0	0.0	0.0	0.0	0.0			0.0	0.0	0.0	0.0					
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0							
0.0	0.0	0.4	0.0	0.4	0.0	0.4	0.3	0.4							

Gambar 3.6 Perhitungan Matriks Convolutional tahap 1

Pada Gambar 3.7, merupakan proses perhitungan *convolutional* tahap ke 2. Ukuran matriks *Output* dari yang dihasilkan menjadi lebih kecil daripada sebelumnya.

0.0	0.1	0.4	0.2
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.4	0.3	0.4

 \times

1	1
1	1

 $=$

0.1	0.5	0.6
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.4	0.7	0.7

Gambar 3.7 Perhitungan Matriks *Convolutional* tahap 2

Penelitian hanya menggunakan 2 *Hidden Layers*. Hasil dari matriks diubah menjadi flatten kemudian di *Max Polling* setiap *Class*-nya. Hasil dari *Max Polling* di Full-connected seperti Tabel 3.9 untuk dihitung softmax dan dicari 3 class yang memiliki nilai tertinggi. Jadi class yang mempunyai nilai tertinggi yaitu C48 (*INVENTOR*), dan C83 (*POPULATION*) dan C91 (*SPECIALTY*). Di *Max Polling* banyak yang bernilai 0, disebabkan karena saat perhitungan TF-IDF per-*class*nya, terdapat kata yang mempunyai kemunculan disetiap pattern pertanyaan, sehingga semakin sering kata tersebut muncul maka semakin dianggap tidak penting.

Tabel 3.9 Perhitungan Softmax

NO	CLASS	MAX POLLING	EXP	SOFTMAX (EXP N / TOTAL EXP)	
1	C5	=	0.32	1.37	0.04
2	C13	=	1.26	3.54	0.09
3	C16	=	0.32	1.37	0.04
4	C22	=	1.41	4.11	0.11
5	C48	=	1.86	6.43	0.17
6	C57	=	1.57	4.78	0.13
7	C83	=	1.86	6.43	0.17
8	C84	=	0.68	1.97	0.05
9	C87	=	1.05	2.85	0.07
10	C91	=	1.68	5.35	0.14
TOTAL			38.22		

3.2.3 Ekstraksi Judul Halaman

Term yang mempunyai NER menjadi topik dari pertanyaan tersebut. Dan ekstraksi judul di wikipedia dengan cara N-Gram (Sharma dan Mittal 2016). Jadi “Muhammad ibn Musa Al Khwarizmi” jika di unigram menjadi “Muhammad”, “ibn”, “Musa”, “Al” dan “Khwarizmi”.

Tabel 3.10 Hasil Pengambilan Judul dari Wikipedia berdasarkan Unigram

UNIGRAM	NO	HASIL EKSTRAKSI
Muhammad	1.	Muhammad
	2.	Muhammad's Mosque Number Seven
	3.	Muhammad's Prophet
	4.	Muhammad's Tomb
	5.	Muhammad's Year of Sadness
	6.	Muhammad's Year of Sorrow
	7.	Muhammad's attitude toward animals
	8.	Muhammad's attitude towards animals
	9.	Muhammad's birthday
	10.	Muhammad's companions
ibn	1.	<i>Ibn</i>
	2.	<i>Ibn-Ali al-Sanusi</i>
	3.	<i>Ibn-Batuta</i>
	4.	<i>Ibn-Baveh</i>
	5.	<i>Ibn-Fadlan</i>
	6.	<i>Ibn-Hawqal</i>
	7.	<i>Ibn-Hazm</i>
	8.	<i>Ibn-Musa al-Qarizmi</i>
	9.	<i>Ibn-Rushd</i>
	10.	<i>Ibn-Rushd (crater)</i>
Musa	1.	<i>Musa</i>

	2.	<i>Musa'ab Al-Madhwani</i>
	3.	<i>Musa'ab Omar Al Madhwani</i>
	4.	<i>Musa'ab Omar Al Mudwani</i>
	5.	<i>Musa'ab al Madhwani</i>
	6.	<i>Musa'id bin Abdul-Aziz</i>
	7.	<i>Musa'id bin Abdul Aziz</i>
	8.	<i>Musa'id bin Abdul Aziz Al Saud</i>
	9.	<i>Musa'id bin Abdul Aziz al-Saud</i>
	10.	<i>Musa'id bin Abdulaziz Al Saud</i>
Al	1.	<i>Al</i>
	2.	<i>Al&d</i>
	3.	<i>Al'Ain</i>
	4.	<i>Al'Akbar</i>
	5.	<i>Al'Arish</i>
	6.	<i>Al'Asayl Cycling Team</i>
	7.	<i>Al'Ayn</i>
	8.	<i>Al'Garb</i>
	9.	<i>Al'Garb Al'Andalus</i>
	10.	<i>Al'Istaan</i>
Khwarizmi	1.	<i>Khwarizmi</i>
	2.	<i>Khwarizmi International Award</i>
	3.	<i>Khwarizmi Science Society</i>
	4.	<i>Khwarizmi festival</i>
	5.	<i>Khwarizmian</i>
	6.	<i>Khwarizmian Empire</i>
	7.	<i>Khwarizmian Language</i>
	8.	<i>Khwarizmian language</i>
	9.	<i>Khwarizmim</i>
	10.	<i>Khwarizmim Empire</i>

Pada

Tabel 3.11 merupakan hasil dari ekstraksi judul halaman wikipedia berdasarkan bigram. “Muhammad ibn Musa Al Khwarizmi” setelah di bigram yaitu “Muhammad ibn”, “ibn Musa”, “Musa Al”, dan “Al Khwarizmi”.

Tabel 3.11 Hasil Pengambilan Judul dari Wikipedia berdasarkan Bigram

BIGRAM	NO	HASIL EKTRAKSI
Muhammad ibn	1.	<i>Muhammad ibn-Abd-al-Wahab</i>
	2.	<i>Muhammad ibn-Musa al-Khwarizmi</i>
	3.	<i>Muhammad ibn 'Abdallah ibn Malik al-Khuza'i</i>
	4.	<i>Muhammad ibn 'Abdallah ibn Muhriz</i>
	5.	<i>Muhammad ibn 'Abdallah ibn Tahir</i>
	6.	<i>Muhammad ibn 'Ali</i>
	7.	<i>Muhammad ibn 'Ali 'Abd ash-Shakur</i>
	8.	<i>Muhammad ibn Abbas</i>
	9.	<i>Muhammad ibn Abd-Allah</i>
	10.	<i>Muhammad ibn Abd Al-Haqq</i>
ibn Musa	1.	<i>Ibn Mu'adh al-Jayyani</i>
	2.	<i>Ibn Mu'adh</i>
	3.	<i>Ibn Naghdala</i>
	4.	<i>Ibn Mājah</i>
	5.	<i>Ibn Mājid</i>
	6.	<i>Ibn Na'ima al-Himsi</i>
	7.	<i>Ibn Nadeem</i>
	8.	<i>Ibn Nadim</i>
	9.	<i>Ibn Nafess Hospital</i>
	10.	<i>Ibn Nafis</i>
Musa Al	1.	<i>Musa Al-Kadhim</i>
	2.	<i>Musa Al-Koni</i>
	3.	<i>Musa Al-Sadr</i>

	4.	<i>Musa Al-Taamari</i>	
	5.	<i>Musa Al-Zoubi</i>	
	6.	<i>Musa Al-Zubi</i>	
	7.	<i>Musa Al Madany</i>	
	8.	<i>Musa Al Wahab</i>	
	9.	<i>Musa Alami</i>	
	10.	<i>Musa Ali</i>	
	Al Khwarizmi	1.	<i>Al Khwarizmi</i>
		2.	<i>Al Khwarizmi International College</i>
		3.	<i>Al Kibar</i>
4.		<i>Al Kidwah</i>	
5.		<i>Al Kifaf</i>	
6.		<i>Al Kifah Refugee Center</i>	
7.		<i>Al Kifah refugee center</i>	
8.		<i>Al Kifl</i>	
9.		<i>Al Kikume</i>	
10.		<i>Al Kilgore</i>	

Tabel 3.12 merupakan hasil dari ekstraksi judul halaman wikipedia berdasarkan trigram. Hasil dari kata “Muhammad ibn Musa Al Khwarizmi” yaitu “Muhammad ibn Musa”, “ibn Musa Al”, dan “Muhammad ibn Musa Al Khwarizmi”.

Tabel 3.12 Hasil Pengambilan Judul dari Wikipedia berdasarkan Trigram

TRIGRAM	NO	JUDUL HALAMAN
Muhammad ibn Musa	1.	<i>Muhammad ibn Musa</i>
	2.	<i>Muhammad ibn Musa Al-Khwarizmi</i>
	3.	<i>Muhammad ibn Musa al-A'raj</i>
	4.	<i>Muhammad ibn Musa al-Hwarizmi</i>
	5.	<i>Muhammad ibn Musa al-Kadhim (Sabze Ghaba)</i>
	6.	<i>Muhammad ibn Musa al-Khawarazmi</i>
	7.	<i>Muhammad ibn Musa al-Khwarizmi</i>

	8.	<i>Muhammad ibn Musa al-Kwarizmi</i>
	9.	<i>Muhammad ibn Musa ibn Shakir</i>
	10.	<i>Muhammad ibn Musa ibn Tulun</i>
ibn Musa Al	1.	<i>Ibn Mu'adh al-Jayyani</i>
	2.	<i>Ibn Mu'adh</i>
	3.	<i>Ibn Mu'adh al-Jayyānī</i>
	4.	<i>Ibn Mājah</i>
	5.	<i>Ibn Mājid</i>
	6.	<i>Ibn Na'ima al-Himsi</i>
	7.	<i>Ibn Nadeem</i>
	8.	<i>Ibn Nadim</i>
	9.	<i>Ibn Nafess Hospital</i>
	10.	<i>Ibn Nafis</i>
Musa Al Khwarizmi	1.	<i>Musa Al Madany</i>
	2.	<i>Musa Al Wahab</i>
	3.	<i>Musa Alami</i>
	4.	<i>Musa Ali</i>
	5.	<i>Musa Aman</i>
	6.	<i>Musa Amer</i>
	7.	<i>Musa Amer Obaid</i>
	8.	<i>Musa Anter</i>
	9.	<i>Musa Arafat</i>
	10.	<i>Musa Araz</i>

Dari hasil ekstraksi judul dengan unigram, bigram dan trigram. Terdapat 1 judul yang sering muncul yaitu “**Muhammad ibn Musa Al Khwarizmi**”. Jika ternyata yang dihasilkan lebih dari 1, maka pengguna akan disuruh memilih topik mana yang menjadi pembahasan dari pertanyaan tersebut.

3.2.4 Ekstraksi Jawaban dengan SPARQL dari DBpedia

Ekstraksi jawaban menggunakan query SPARQL dengan parameter judul halaman dan EAT (Teixeira, Ketsmur, dan Rodrigues 2017). EAT-nya adalah C48 (*INVENTOR*), dan C83 (*POPULATION*) dan C91 (*SPECIALTY*).

Tabel 3.13 *Template Query* pada SPARQL

```

PREFIX ontology : <http://dbpedia.org/ontology>
PREFIX dbo      : <http://dbpedia.org/ontology/>
PREFIX dbc     : <http://dbpedia.org/resource/Category:>
PREFIX dct     : <http://purl.org/dc/terms/>
PREFIX dbp    : <http://dbpedia.org/property/>

SELECT *WHERE {
  dbr: JUDUL_HALAMAN ?p ?o .
  FILTER regex(str(?p), "EAT/Leksikal", "i") .
}

```

Error! Reference source not found. hasil query berdasarkan EAT. Ada beberapa label class ontology yang tidak *matching* dengan EAT sehingga sebuah *query* tidak menampilkan hasil atau jawaban. Untuk mengatasi hal tersebut, sistem dapat mengganti pencarian *ontology class* pada *query filter* dengan kata atau *term* dari data *word embedding* atau *graph* yang sudah didapatkan sebelumnya. Jika masih belum menampilkan hasil juga penyebabnya adalah dikarenakan di dalam Dbpedia, belum tersedianya ontology class atau *label property* di Dbpedia atau kata per-*label property* yang digunakan tidaklah umum. Pada tabel dibawah ini proses *query* SPARQL dengan kata EAT yang sudah didapatkan. Namun pada proses ini tidak menampilkan hasil.

Tabel 3.14 Hasil SPARQL berdasarkan EAT

QUERY	HASIL				
<pre>select*where { dbr: Muhammad_ibn_Musa_al-Khwarizmi ?p ?o FILTER regex(str(?p)," <i>inventor</i> ","i") . }</pre>	<table border="1"> <thead> <tr> <th>P</th> <th>O</th> </tr> </thead> <tbody> <tr> <td>-</td> <td>-</td> </tr> </tbody> </table>	P	O	-	-
P	O				
-	-				
<pre>select *where { dbr: Muhammad_ibn_Musa_al-Khwarizmi ?p ?o FILTER regex(str(?p)," <i>population</i> ","i") . }</pre>	<table border="1"> <thead> <tr> <th>P</th> <th>O</th> </tr> </thead> <tbody> <tr> <td>-</td> <td>-</td> </tr> </tbody> </table>	P	O	-	-
P	O				
-	-				
<pre>select *where { dbr: Muhammad_ibn_Musa_al-Khwarizmi ?p ?o FILTER regex(str(?p)," <i>specialty</i> ","i") . }</pre>	<table border="1"> <thead> <tr> <th>P</th> <th>O</th> </tr> </thead> <tbody> <tr> <td>-</td> <td>-</td> </tr> </tbody> </table>	P	O	-	-
P	O				
-	-				

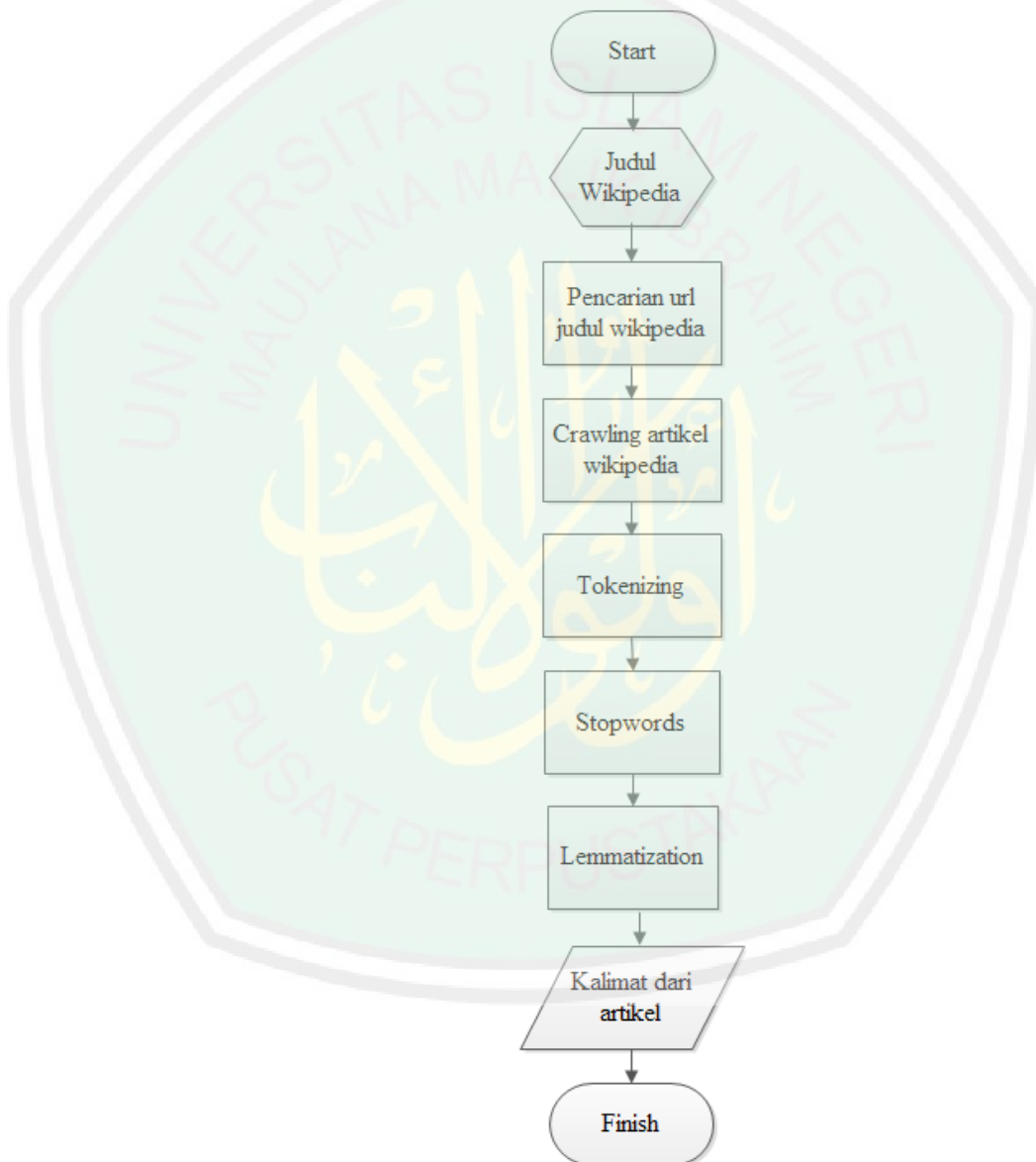
Karena hasil Query SPARQL pada sebelumnya tidak ditemukan, posisi kata EAT diganti dengan kata yang sudah didapatkan pada *word embedding*. Sehingga kata yang sesuai dengan menghasilkan jawaban yaitu kata "*Notable*".

Tabel 3.15 Proses Query SPARQL dengan kata "*Notable*"

QUERY	HASIL				
<pre>select *where { dbr: Muhammad_ibn_Musa_al- Khwarizmi ?p ?o FILTER regex(str(?p)," <i>notable</i>","i") . }</pre>	<table border="1"> <thead> <tr> <th>P</th> <th>O</th> </tr> </thead> <tbody> <tr> <td> http://dbpedia.org/ property/notableIdeas </td> <td> "Treatises on algebra and Indian numerals" </td> </tr> </tbody> </table>	P	O	http://dbpedia.org/ property/notableIdeas	"Treatises on algebra and Indian numerals"
P	O				
http://dbpedia.org/ property/notableIdeas	"Treatises on algebra and Indian numerals"				

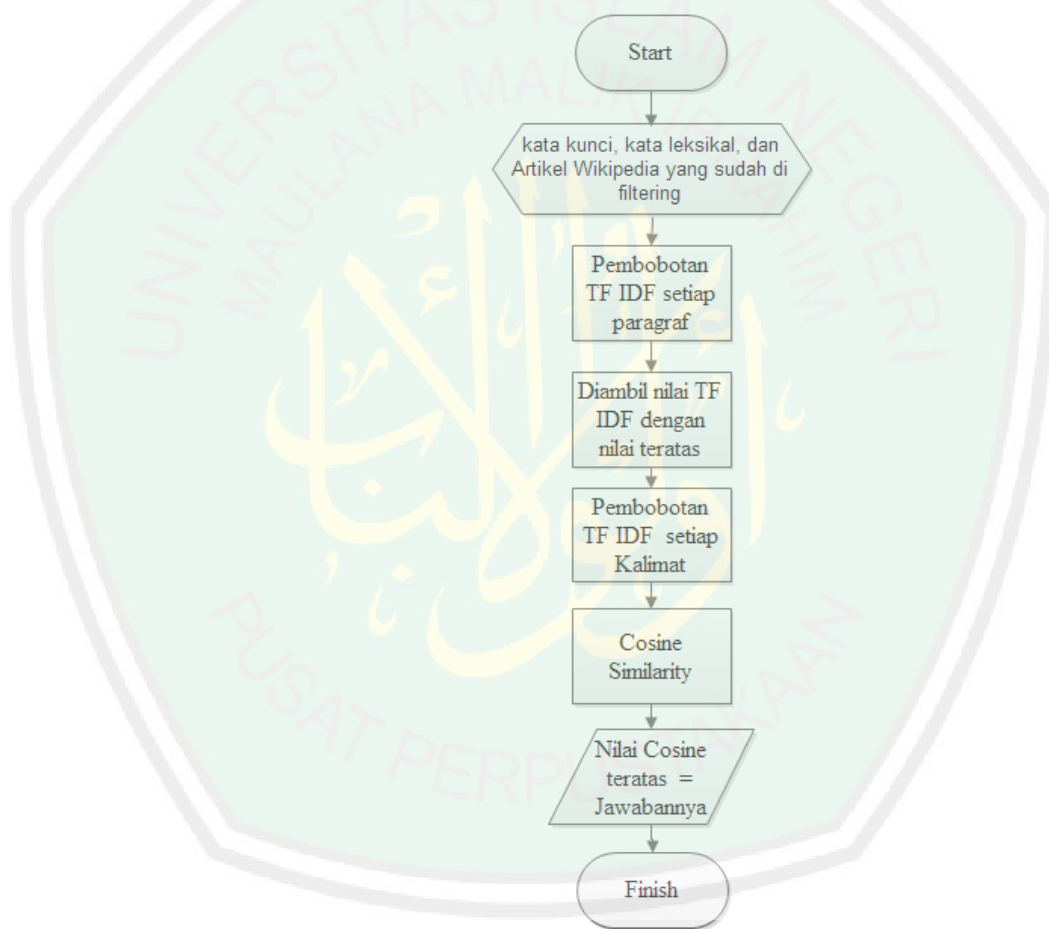
3.2.5 Ekstraksi Jawaban dengan Cosine Similarity dari Wikipedia

Cara ke dua untuk mendapatkan jawaban, yaitu mencari kemiripan kalimat tanya pada suatu artikel di wikipedia. Cara ini dimulai dari pencarian URL judul wikipedia sudah diperoleh yaitu “*Muhammad ibn Musa Al Khwarizmi*”. Sehingga judul tersebut akan di *crawling* artikelnya. Lalu difiltering mulai dari *tokenizing*, *stopwords*, *lemmatization*.



Gambar 3.8 Flowchart Ekstraksi Jawaban dengan Cosine Similarity

Kemudian untuk menemukan jawabannya data yang diperlukan yaitu artikel wikipedia yang sudah di *filtering*, kata kunci dan leksikal yang didapatkan pada tahap *preprocessing* pertanyaan. Lalu dihitung nilai TF IDF kata kunci dan leksikal pada setiap paragraf, dengan nilai TF IDF teratas yang akan diambil. Kemudian dihitung kembali nilai TF IDF setiap kalimat didalam paragraf tersebut. Dan dihitung *cosine similarity* untuk mendapatkan jawabannya. Nilai *cosine similarity* teratas yang menjadi jawaban dari kalimat tanya yang diajukan oleh pengguna.



Gambar 3.9 Flowchart pencarian jawaban

Pada Tabel 3.16 dan Tabel 3.17 merupakan perhitungan pembobotan TF IDF pada setiap paragraf di artikel “Muhammad ibn Musa al-Khwarizmi”. Jika sebuah

term seperti kata “*famous*”, “*work*” dan “*create*” tidak ada di dalam suatu dokumen, sistem bisa mengambil atau memakai kembali data *word embedding* atau *graph* yang sudah didapatkan sebelumnya.

Tabel 3.16 TF pada setiap paragraf

TERM	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	TF
Famous	0	0	0	0	0	0	0	0	0	0	0.00
Work	3	2	2	5	3	1	0	1	1	1	9.00
Create	0	0	0	0	0	0	0	0	0	0	0.00
Muhammad Ibn Mūsā Al-Khwārizmī	1	1	0	0	0	0	0	0	0	0	0.00

Tabel 3.17 IDF pada setiap paragraf

TERM	Log (n/df)	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Famous	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Work	0.05	0.14	0.09	0.09	0.23	0.14	0.05	0.00	0.05	0.05	0.05
Create	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Muhammad Ibn Mūsā Al-Khwārizmī	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TOTAL		0.84	0.79	0.09	0.23	0.14	0.05	0.00	0.05	0.05	0.05

Pada Tabel 3.18 dan

Tabel 3.19 merupakan perhitungan dari TF dan IDF dari paragraf 1. Kalimat yang memiliki nilai diatas angka 0 yaitu kalimat ke 1, 11 dan 12.

Tabel 3.18 TF setiap kalimat pada paragraf 1

TERM/	KALIMAT KE -													TF
	Q	1	2	3	4	5	6	7	8	9	10	11	12	
Famous	1	0	0	0	0	0	0	0	0	0	0	0	0	1
Work	1	1	0	0	0	0	0	0	0	0	0	1	1	4
Create	1	0	0	0	0	0	0	0	0	0	0	0	0	1
Muhammad Ibn Mūsā Al-Khwārizmī	1	1	0	0	0	0	0	0	0	0	0	0	0	2

Tabel 3.19 IDF setiap kalimat pada paragraf 1

TERM	ID F	Q	KALIMAT KE -											
			1	2	3	4	5	6	7	8	9	10	11	12
Famous	1.1	1.1	0	0	0	0	0	0	0	0	0	0	0	0
Work	0.5	0.5 1	0.51	0	0	0	0	0	0	0	0	0	0.5 1	0.5 1
Create	1.1	1.1 1	0	0	0	0	0	0	0	0	0	0	0.0	0.0
Muhammad Ibn Mūsā Al-Khwārizmī	0.8	0.8 1	0.81	0	0	0	0	0	0	0	0	0	0.0	0.0
TOTAL		3.5 5	1.3 2	0	0	0	0	0	0	0	0	0	0.5 1	0.5 1

Berikut merupakan 3 kalimat dari paragraf 1 yang telah didapatkan dari TF-IDF sebelumnya. Lalu dihitung *Cosinus Similarity*-nya untuk mendapatkan sebuah jawaban.

1. *Muhammad ibn Mūsā al-Khwārizmī* (Persian: محمد بن موسى خوارزمی; c. 780 – c. 850), formerly Latinized as *Algorithmi*, was a Persian scholar who produced works in mathematics, astronomy, and geography under the patronage of the Caliph Al-Ma'mun of the Abbasid Caliphate.
11. In addition to his best-known works, he revised Ptolemy's *Geography*, listing the longitudes and latitudes of various cities and localities.
12. He further produced a set of astronomical tables and wrote about calendaric works, as well as the astrolabe and the sundial.

Pada Tabel 3.20 merupakan awal dari proses perhitungan *Cosinus Similarity* dari beberapa kalimat yang sudah didapatkan sebelumnya.

Tabel 3.20 *Scalar* dan Panjang Vektor Kalimat Pada Paragraf 5

	PARAGRAF KE 5						
	SKALAR = WD*WDI			PANJANG VEKTOR			
	K1	K11	K12	Q	K1	K11	K12
	0	0	0	1	0	0	0
	1	1	1	1	1	1	1
	0	0	0	1	0	0	0
	1	0	0	1	1	0	0
SUM	2	1	1	4.0	2.0	1.0	1.0
SQRT				2.0	1.4	1.0	1.0

Pada Tabel 3.21 merupakan proses perhitungan skalar antara Q dengan kalimat yang lainnya. Menunjukkan bahwa nilai tertinggi terdapat pada Ke 1.

Tabel 3.21 Perhitungan *Cosine Similarity*

PARAGRAF 1
• $\text{Cos}(Q,K1) = \frac{2}{2*1.4} = 0.71$
• $\text{Cos}(Q,K11) = \frac{1}{2*1} = 0.50$
• $\text{Cos}(Q,K12) = \frac{1}{2*1} = 0.50$

Jadi jawaban dari pertanyaan “*What famous works created by Muhammad ibn Musa Al Khwarizmi?*” yaitu sebagai berikut:

- Treatises on algebra and Indian numerals
- *Muḥammad ibn Mūsā al-Khwārizmī* (Persian: محمد بن موسى خوارزمی; c. 780 – c. 850), formerly Latinized as *Algorithmi*, was a Persian scholar who produced works in mathematics, astronomy, and geography under the patronage of the Caliph Al-Ma'mun of the Abbasid Caliphate.

BAB IV

UJI COBA DAN PEMBAHASAN

Bab ini membahas uji coba system dengan metode Graph Convolutional Neural Network (GCNN) dan Cosinus Similarity. Bertujuan untuk mengetahui keakurasian system yang dibangun apakah mampu memberikan jawaban sesuai dengan pertanyaan yang diberikan oleh user.

4.1 Platform yang digunakan

Pada proses uji coba sistem spesifikasi perangkat keras dan perangkat lunak yang digunakan sebagai berikut:

1. Perangkat Keras (*Hardware*)

- Processor : Intel ® Core™ i3-4030U CPU @1.90 Ghz
- RAM : 6.00 GB
- Sytem Type : 64-bit Operating System Windows 10

2. Perangkat Lunak (*Software*)

Adapun software atau perangkat lunak yang digunakan dalam pembuatan klasifikasi dan pengujian metode yaitu:

- Visual Studio Build Tools 2017, sebagai platform yang digunakan untuk merancang sistem.
- DB Browser for SQLite, sebagai manajemen basis data relasional.
- Mozilla Firefor 66.03 (x64 id), sebagai media menjalankan system tanya jawab
- Bahasa pemrograman Python versi 3.7.2 (32-bit).
- PIP3 (Pip Install Packages - Python 3) 19.0.3. Package yang diperlukan dalam sistem ini:

Tabel 4.1 Library yang digunakan

No	Nama	Versi	Keterangan
1.	NLTK	3.4	<i>Tool</i> untuk <i>Natural Language Processing</i> pada preprocessing. Seperti tokenizing, lemmatization, dll.
2.	Standford NLP Group	3.9.2	<i>Library</i> untuk menemukan <i>Named Entity Recognition</i> (NER) pada kalimat tanya
3.	Numpy	1.16.2	<i>Library</i> yang mengelola <i>array</i> atau matriks multi dimensi seperti penggunaan transpose, mencari nilai maximum, dll.
4.	SPARQLWrapper	1.24.1	Membantu membuat URI query dan mengubah dari RDF menjadi hasil yang mudah dikelola
5.	Wikipedia-API	0.4.4	Untuk mengakses dan memparsing data dari Wikipedia

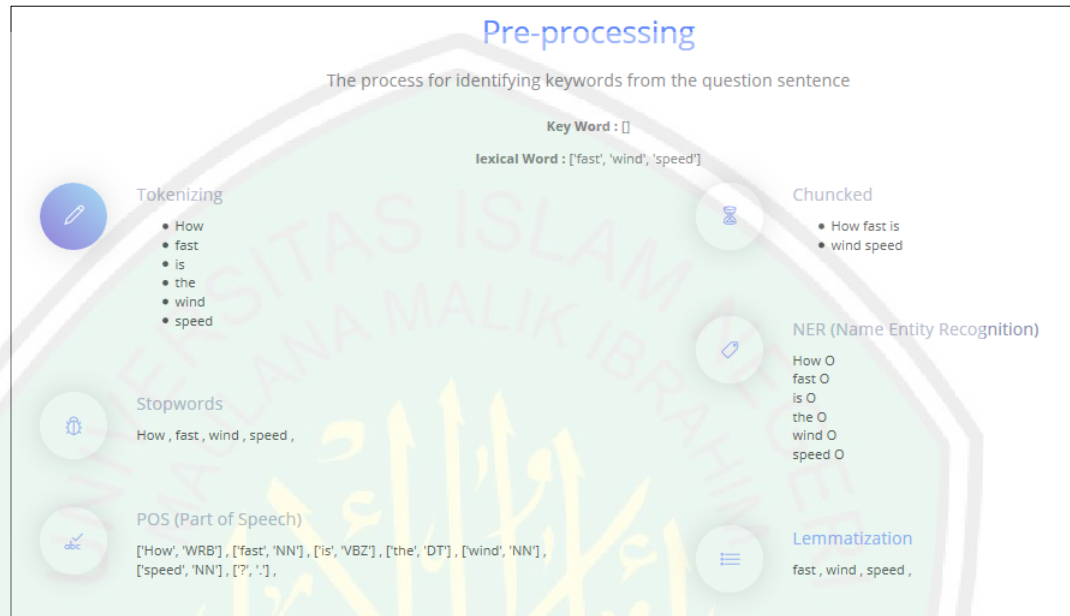
4.2 Implementasi Sistem

Pada sistem ini hanya terdiri dari 1 halaman seperti *landing page*. Yang berisi form pertanyaan dan proses untuk mendapatkan jawabannya. *Section* pertama menampilkan form, untuk memasukan pertanyaan kepada sistem.

The image shows a web interface for a question-and-answer system. At the top, it says "QAs on WIKIPEDIA" followed by a subtitle: "Question and answer system themed about history and mathematical scientists with the wikipedia knowledge base". Below this is a search bar with the text "How fast is the wind speed?" and a blue "SEARCH" button. At the bottom of the interface is a blue button with a downward arrow and the word "ANSWER".

Gambar 4.1 *Form* pertanyaan

Setelah pertanyaan diinputkan maka kalimat tanya akan diproses pada tahapan preprocessing yaitu ada tokenizing, stopword, Part of Speech (POS), Chunked, Named Entity Recognition (NER), dan Lemmatization. Hasil dari processing dapat dilihat di Gambar 4.2.



Gambar 4.2 Tampilan proses Preprocessing

Jika pada tahapan preprocessing tidak mampu menemukan topik pada kalimat tanya. Maka sistem akan menampilkan form untuk mengisi topik yang dimasukkan dalam pertanyaan tersebut seperti pada Gambar 4.3..

Sorry your question topic is not detected. Please enter the topic you want

How fast is the wind speed?

wind speed

SUBMIT

Gambar 4.3 Form untuk memasukan topik

Sistem akan mencari artikel wikipedia yang memiliki kemiripan dengan topik pertanyaan menggunakan N-grams yaitu unigram, bigram dan trigram. Karena jumlah kata pada topik yang didapatkan hanya 2 kata. Maka N-grams hanya dilakukan sampai bigram saja, seperti gambar dibawah ini..



Gambar 4.4 Hasil dari proses N-Gram

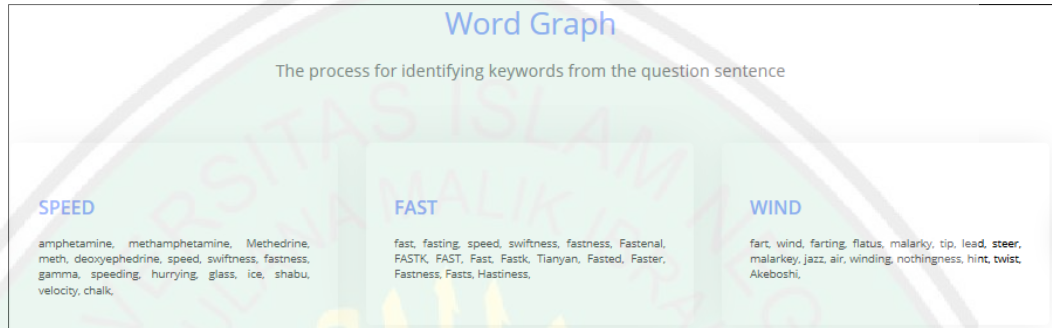
Jika judul wikipedia yang didapatkan lebih dari satu. Maka sistem akan menampilkan *form option* seperti di bawah ini. Penggunaan diminta untuk memilih salah satu topik dari beberapa topik yang tersedia seperti Gambar 4.5.

- Wind Surf (ship)
- Wind Talker sound suppressor
- Wind controller
- Wind hybrid power systems
- Wind speed
- Wind's Poem
- Wind, Sand and Stars
- Wind, Sky and Diamonds
- Wind-Del Velodrome
- Winds Fairground
- Windstar Cruises
- Windsurfing
- Windtalkers

SAVE

Gambar 4.5 *Form option* untuk memilih topik

Lalu jika topik atau judul wikipedia sudah didapatkan atau ditemukan. Maka sistem akan mencari *word graph* atau *thesaurus* atau *similarity* dari kata leksikal yang sudah didapatkan sebelumnya, tampilannya dapat dilihat di Gambar 4.6. Proses ini pengambilan data dari API babalnet dengan memnfilter beberapa hal seperti language yang dipakai berbahasa inggris, kata tidak boleh ada simbol, dan yang diambil hanya 15kata.



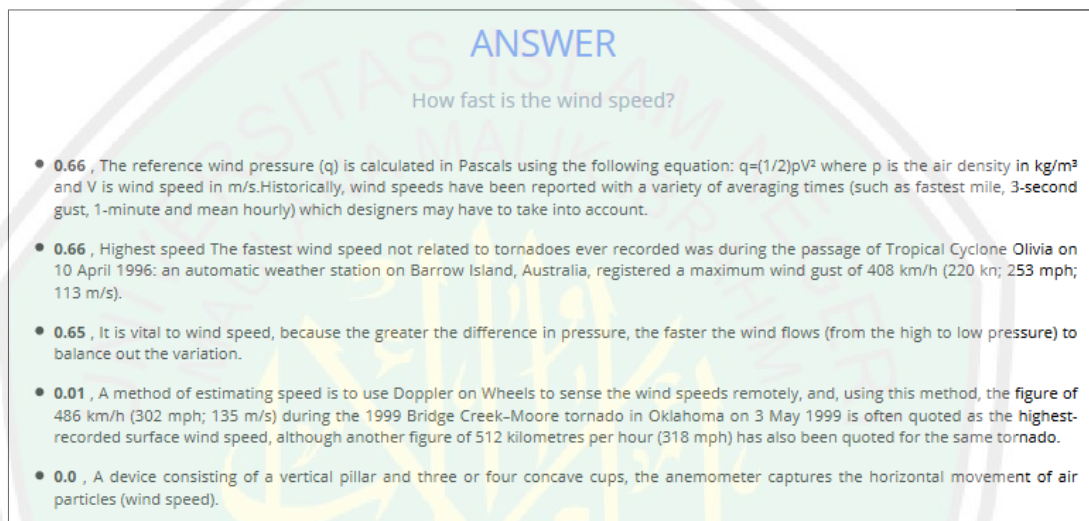
Gambar 4.6 Tampilan Hasil *Word Graph*

Hasil dari word graph diatas akan masuk kedalam proses CNN. Dibawah ini merupakan tampilan hasil proses Convolution Neural Network (CNN) dengan 3 *classification*. Proses ini dilakukan proses TF IDF kemudian mengambil 10 Class teratas untuk dimasukan ke CNN. Lalu setiap *class* tersebut akan dilakukan perkalian dengan matrik filter sebanyak 2 kali. Kemudian dihitung max polling setiap class. Dan terakhir masuk ke softmax untuk dihitung exponennya antara class satu dengan class yang lain untuk diambil 3 class teratas.



Gambar 4.7 Tampilan hasil proses *Convolutional Neural Network* (CNN)

Terakhir desain tampilan pada hasil jawaban yang didapatkan dari beberapa proses sebelumnya. Jawaban didapatkan dari 2 macam sumber yaitu dbpedia dan wikipedia. Pengambil data menggunakan query SPARQL di dbpedia dengan parameter judul wikipedia, key dari babelnet, dan key dari kalimat tanya yang sudah didapatkan sebelumnya. Sedangkan Pengambil data di artikel wikipedia, proses tersebut dibutuhkan perhitungan skalar dan panjang vektor untuk mendapatkan nilai cosine similaritynya.



Gambar 4.8 Hasil Jawaban

4.3 Uji Coba Sistem

Kriteria pengujian sistem ini yaitu mengukur keakurasian jawaban yang ditemukan dengan algoritma *Graph Convolutional Neural Network* (GCNN) dan *Cosinus Similarity*. Kriteria pertanyaan yang diijinkan yaitu:

- Menggunakan bahasa inggris baku,
- Kalimat tanya tertutup atau *factoid*,
- Kalimat tanya memiliki kata leksikal atau predikat (Contoh: What is **definition** of Geometry?)
- Huruf yang dipakai adalah latin, tidak menggunakan huruf yunani (Contoh: Where was **al-Kāshī** born?).

- Penggunaan huruf besar pada awal kata topik, dan di awal kalimat tanya (Contoh: **When does Geometry exist?**)

Pada pengujian ini dapatkan beberapa kemungkinan yang dapat dianalisa:

1. *Posisi Jawaban* : Posisi jawaban yang dihasilkan dari 2 proses yaitu Query SPARQL dan Cosine Similarit
2. *True Positive (TP)* : Jawaban yang dihasilkan sistem benar
3. *False Positive (FP)* : Jawaban yang dihasilkan salah atau sistem tidak menghasilkan jawaban.
4. *True Negative (TN)* : Pertanyaan yang diajukan tidak sesuai dengan ketentuan dan sistem tidak menghasilkan jawaban.
5. *False Negative (FN)* : Pertanyaan yang diajukan tidak sesuai dengan ketentuan tetapi sistem menghasilkan jawaban.
6. *Inpt* : Sistem membutuhkan penjelasan topik pertanyaan secara khusus
7. *OPT* : Sistem mendapatkan lebih dari 1 judul artikel
8. Σ Key : Jumlah keseluruhan kata yang didapat dari babelnet
9. Σ Kalimat : Jumlah kalimat di dalam artikel
10. Waktu : Durasi eksekusi sistem

Pada pengujian juga dapat diukur kita ukur accuracy melalui recall dan precision nya. *Recall* adalah kemampuan sistem untuk memanggil dokumen yang relevan. *Precision* adalah kemampuan untuk tidak memanggil dokumen yang tidak relevan. Berikut perhitungannya:.

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (4.1)$$

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%$$

Dibawah ini Uji coba yang dilakukan terhadap sistem dengan kebenaran jawaban yang di dapatkan.

1. How (Bagaimana)

Kata tanya “*How Many*,” “*How Fast*” digunakan untuk pertanyaan yang memberikan jawaban bertipe jumlah. Pada Tabel 4.2 mendapatkan hasil pengukuran uji coba yaitu precisionnya bernilai 80%, recall 100% dan *Accuracy* 80%. Uji coba tersebut mampu menghasilkan jawaban di cosine similarity ada di urutan pertama. Dengan dilakukan percobaan 10 kali terjadi kesalahan pada sistem sebanyak 2 kali



Tabel 4.2 Uji Coba Pertanyaan menggunakan kata tanya “How”

NO	PERTANYAAN	Jawaban	Posisi Jawaban		JAWABAN			
			Sparql	Cosine	TP	FP	TN	FN
1	How many stars according to Ulugh Beg?	He compiled the 1437 Zij-i-Sultani of 994 stars	0	1 & 2	✓			
2	How many years is determined by Ulugh Beg?	Ulugh Beg determined the length of the sidereal year as 365.2570370...d = 365d 6h 10m 8s (an error of +58 seconds)	0	1	✓			
3	How many zones of earth are divided according to Pythagoras?	the first to divide the globe into five climactic zones.	0	3	✓			
4	How much volume of tetrahedron on the cube	One such regular tetrahedron has a volume of 1/3 of that of the cube	0	1	✓			
5	How fast is the wind speed?	The fastest wind on any known planet is on HD 80606 b located 190 light years away, where it blows at more than 11,000 mph or 5 km/s.	0	1	✓			
6	How many vertice of the cube?	vertices : 8	1	2 & 4	✓			

7	How many interior angles in a triangle	A triangle with an interior angle of 180° (and collinear vertices) is degenerate	0	0		✓		
8	How many Ulugh Beg's wife?	Marriages Ulugh Beg had thirteen wives	0	1	✓			
9	How many seconds in a minute?	As a unit of time, the minute is most of times equal to $1/60$ (the first sexagesimal fraction) of an hour, or 60 seconds	0	0		✓		
10	How fast is the speed of light?	From this effect he determined that light must travel 10210 times faster than the Earth in its orbit (the modern figure is 10066 times faster).	0	3 & 4	✓			

Berikut hasil dari Tabel 4.3 analisa dari kata tanya how.

- Parameter input, sistem hanya mampu mendeteksi topik pertanyaan sebanyak 3x.
- Parameter opt, sistem memberikan rekomendasi lebih dari 1 judul artikel wikipedia pada semua query pertanyaan.
- Parameter key, leksikal yang didapat pada babelnet paling sedikit yaitu 20 kata dan terbanyak dengan jumlah 47 kata
- Parameter kalimat artikel yang didapat didalam wikipedia kalimat paling sedikit yaitu 22 kalimat dan terbanyak yaitu 312 kalimat.
- Parameter waktu, sistem mampu menghasilkan jawaban dengan waktu tercepat yang dibutuhkan sebanyak 79 detik, dan waktu terlama yaitu 130 detik.

Tabel 4.3 Analisa kata tanya “How”

No	input	opt	Σ Key	Σ Kalimat	Waktu (s)
1		✓	24	104	79
2		✓	35	104	81
3		✓	47	312	89
4	✓	✓	40	67	94
5	✓	✓	43	35	98
6	✓	✓	20	67	105
7	✓	✓	77	315	119
8	✓	✓	27	104	121
9	✓	✓	24	22	124
10	✓	✓	43	264	130

2. What (Apa)

Kata tanya “What” digunakan untuk pertanyaan yang memberikan jawaban bertipe benda. Pada tabel Tabel 4.4 mendapatkan hasil pengukuran uji coba ini precisionnya bernilai 80%, recall 100% dan Accuracy 80%. Bahwa pada uji coba kata tanya what jawaban sering muncul di SPARQL pada urutan pertama. Dengan dilakukan percobaan 10 kali terjadi kesalahan pada sistem sebanyak 2 kali.

Tabel 4.4 Uji Coba Pertanyaan menggunakan kata tanya “What”

NO	PERTANYAAN	Jawaban	Posisi Jawaban		JAWABAN			
			Sparql	Cosine	TP	FP	TN	FN
1	What type of Paraboloid?	There are two types of paraboloid, elliptic and hyperbolic	0	0		✓		
2	What is known as Isaac Newton?	knownFor : Newtonian_mechanics	1	0	✓			
3	What nationality of Isaac Newton?	Nationality : English	1	0	✓			
4	What era did Ibnu Sina succeed?	era : Islamic_Golden_Age	1	0	✓			
5	What is Al Biruni main interest	MainInterests : Astronomy			✓			
6	What was the full name of the Al Kindi?	Fullname : 'Abū Yūsuf Ya‘qūb ibn 'Ishāq al-Kindī	1	0	✓			
7	What is use of Natural Numbers ?	The natural numbers are those used for counting (as in "there are six coins on the table") and ordering (as in "this is the third largest city in the country")	0	0		✓		

8	What famous works created by Al Khwarizmi?	Notable Ideas : Treatises on algebra and Indian numerals	1	3	✓		
9	What is the name of the substitution and transposition of the password published by Al Qalqashandi?	The Compendious Book on Calculation by Completion and Balancing	0	1	✓		
10	What is the lateral surface area of a cone?	Surface area The lateral surface area of a right circular cone is $L S A = \pi r l$	0	1	✓		

Berikut hasil dari Tabel 4.5 analisa dari kata tanya what.

- Parameter input, sistem tidak mampu mendeteksi topik pertanyaan sebanyak 3 kali
- Parameter opt, sistem memberikan rekomendasi lebih dari 1 judul artikel wikipedia pada hampir semua query pertanyaan.
- Parameter key, leksikal yang didapat pada babelnet paling sedikit yaitu 17 kata dan terbanyak dengan jumlah 56 kata
- Parameter kalimat artikel yang didapat didalam wikipedia kalimat paling sedikit yaitu 13 kalimat dan terbanyak yaitu 304 kalimat.
- Parameter waktu, sistem mampu menghasilkan jawaban dengan waktu tercepat yang dibutuhkan sebanyak 59 detik, dan waktu terlama yaitu 158 detik.

Tabel 4.5 Analisa kata tanya “What”

No	input	opt	Σ Key	Σ Kalimat	Waktu (s)
1	✓	✓	22	62	59
2		✓	19	286	65
3		✓	17	286	72
4		✓	21	304	74
5			34	162	84
6		✓	41	245	86
7	✓	✓	38	116	94
8		✓	41	117	95
9		✓	56	13	139
10	✓	✓	53	59	158

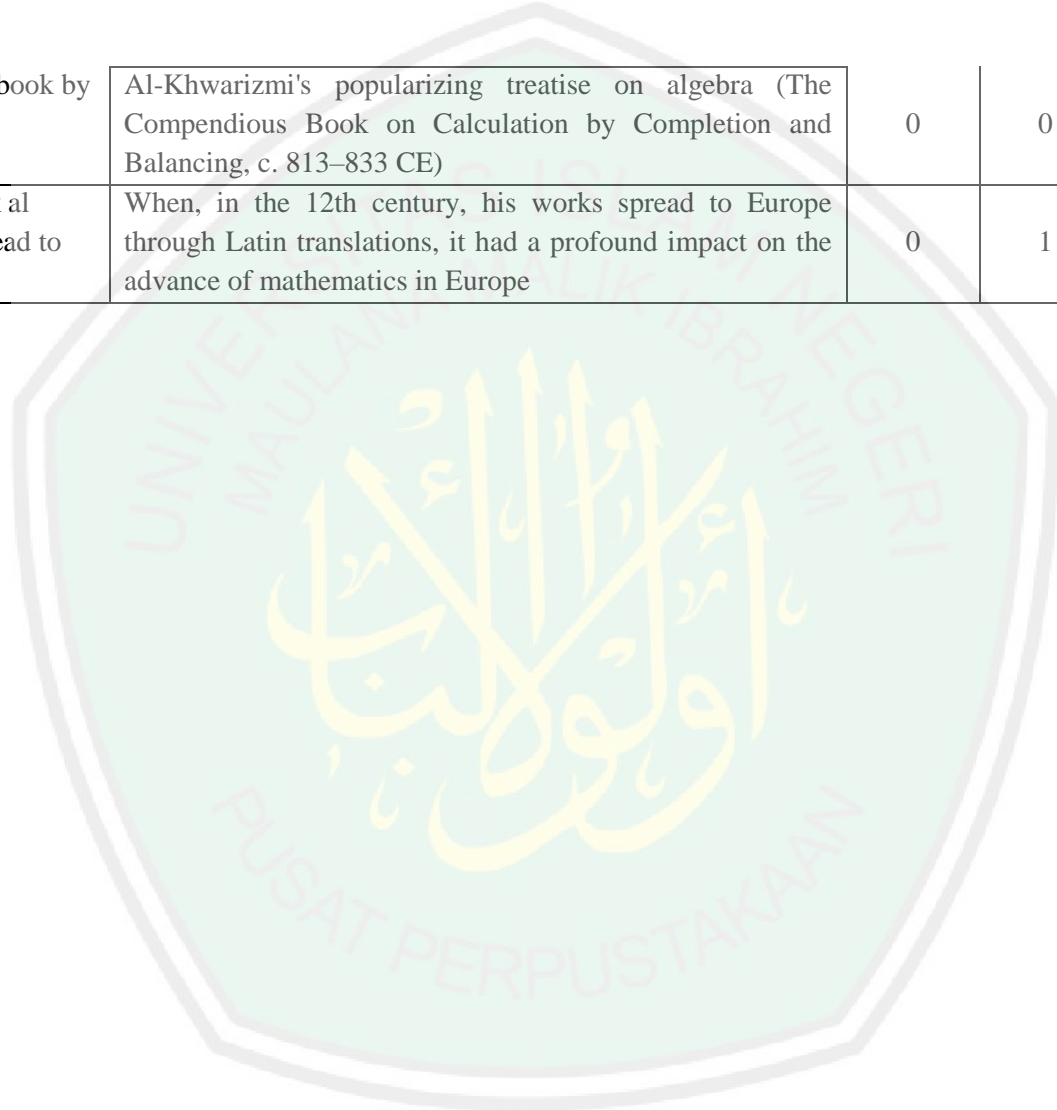
3. *When* (Kapan)

Kata tanya “*When*” digunakan untuk pertanyaan yang memberikan jawaban bertipe waktu. Pada Tabel 4.6 mendapatkan hasil pengukuran uji coba ini precisionnya bernilai 88.89%, recall 100% dan *Accuracy* 90%. Bahwa pada uji coba kata tanya when jawaban sering muncul di cosine pada urutan pertama. Dengan dilakukan percobaan 10 kali terjadi kesalahan sebanyak 2 kali.

Tabel 4.6 Uji Coba Pertanyaan menggunakan kata tanya “When”.

No	Pertanyaan	Jawaban	Posisi Jawaban		Jawaban			
			Sparql	Cosine	TP	FP	TN	FN
1	When Geometry appear?	Geometry arose independently in India, with texts providing rules for geometric constructions appearing as early as the 3rd century BC.	0	1	✓			
2	When numbers are found?	Abstract : During the 19th century, mathematicians began to develop many different abstractions which share certain properties of numbers and may be seen as extending the concept	1	0	✓			
3	When is the triangle found?	Jawaban tidak ada karena didalam artikel TRIANGLE tidak ada pembahasan tentang sejarah	0	0			✓	
4	When was ibn Sinan died?	He died in June 1037, in his fifty-eighth year, in the month of Ramadan and was buried in Hamadan, Iran	0	1	✓			
5	When was the Algebra invented?	In 1637, René Descartes published La Géométrie, inventing analytic geometry and introducing modern algebraic notation	0	1	✓			
6	When did Al-Birjandi translate his book into Sanskrit?	The 11th chapter of the book was translated to Sanskrit in 1729 at Jaipur by Nayanasukhopadhyaya	0	1	✓			
7	When was the prime number theorem proven?	Abstract : The first result in that direction is the prime number theorem, proven at the end of the 19th century	1	0	✓			
8	When was the book of optics written by Alhazen?	Book of Optics Alhazen's most famous work is his seven-volume treatise on optics Kitab al-Manazir (Book of Optics), written from 1011 to 1021	0	1	✓			

9	When was the book by Al Khwarizmi published?	Al-Khwarizmi's popularizing treatise on algebra (The Compendious Book on Calculation by Completion and Balancing, c. 813–833 CE)	0	0		✓		
10	When his work al khwarizmi spread to europe?	When, in the 12th century, his works spread to Europe through Latin translations, it had a profound impact on the advance of mathematics in Europe	0	1	✓			



Berikut hasil dari Tabel 4.7 analisa dari kata tanya when:

- Parameter input, sistem tidak mampu mendeteksi toopik pertanyaan sebanyak 6 kali
- Parameter opt, sistem memberikan rekomendasi lebih dari 1 judul artikel wikipedia pada hampir semua query pertanyaan.
- Parameter key, leksikal yang didapat pada babelnet paling sedikit yaitu 17 kata dan terbanyak dengan jumlah 72 kata
- Parameter kalimat artikel yang didapat didalam wikipedia kalimat paling sedikit yaitu 15 kalimat dan terbanyak yaitu 264 kalimat.
- Parameter waktu, sistem mampu menghasilkan jawaban dengan waktu tercepat yang dibutuhkan sebanyak 47 detik, dan waktu terlama yaitu 167 detik

Tabel 4.7 Analisa kata tanya “When”

No	input	opt	Σ Key	Σ Kalimat	Waktu (s)
1	✓	✓	17	186	47
2	✓	✓	27	305	63
3	✓	✓	27	315	67
4			30	304	70
5	✓	✓	33	190	81
6		✓	51	15	90
7	✓	✓	55	264	109
8		✓	55	256	109
9			43	117	122
10	✓	✓	72	117	167

4. Where (Dimana)

Kata tanya “Where” digunakan untuk pertanyaan yang memberikan jawaban bertipe tempat. Pada Tabel 4.8 mendapatkan hasil pengukuran uji coba ini precisionnya bernilai 88.89%, recall 100% dan Accuracy 90%. Bahwa pada uji coba kata tanya when jawaban sering muncul di cosine pada urutan pertama. Dengan dilakukan percobaan 10x terjadi kesalahan sebanyak 2 kali.

Tabel 4.8 Uji Coba Pertanyaan menggunakan kata tanya “Where”.

NO	PERTANYAAN	Jawaban	Posisi Jawaban		JAWABAN			
			Sparql	Cosine	TP	FP	TN	FN
1	Where was Pythagoras born?	<ul style="list-style-type: none"> • birthPlace : Samos • Life Early life Herodotus, Isocrates, and other early writers agree that Pythagoras was the son of Mnesarchus and that he was born on the Greek island of Samos in the eastern Aegean. 	1	1	✓			
2	Where did Pythagoras advise the elite?	He served as an advisor to the elites in Croton and gave them frequent advice	0	1	✓			
3	Where is the school of Ibnu Sina?	The Avicenna Directories (2008–15; now the World Directory of Medical Schools) list universities and schools where doctors, public health practitioners, pharmacists and others, are educated.	0	4			✓	
4	Where was born Ibn Khaldun?	<ul style="list-style-type: none"> • birthPlace : Tunis • "Ibn Khaldūn" after a remote ancestor, was born in Tunis in AD 1332 (732 AH) into an upper-class Andalusian family of Arab descent, 	1	1	✓			
5	Where was Al Farabi died?	<ul style="list-style-type: none"> • deathPlace : Damascus • He later spent time in Damascus and in Egypt before returning to Damascus where he died in 950-1. His name was Abū 	1	4	✓			

		Naṣr Muḥammad b. Muḥammad Farabi, sometimes with the family surname al-Ṭarḳānī, i.e., the element Ṭarḳān appears in a nisba					
6	Where was Al-Battani born?	<ul style="list-style-type: none"> • birthPlace : Harran • Life Little is known about al-Battānī's life beside that he was born in Harran near Urfa, in Upper Mesopotamia, which is now in Turkey, and his father was a famous maker of scientific instruments. 	1	1	✓		
7	Where Al Khazini was freed from slaves?	Al-Khazini was an emancipated slave in Marv	0	0		✓	
8	Where is Ulugh's burial place Beg?	deathPlace : Timurid_Empire	1	0	✓		
9	Where is the residence of Ibn Al Haytham?	residence : Cairo	2	0	✓		
10	where is the book titled "The Nine Chapters on the Mathematical Art" becomes an influence in the development of mathematics?	The influence of The Nine Chapters greatly assisted the development of ancient mathematics in the regions of Korea and Japan.	0	1	✓		

Berikut hasil dari Tabel 4.9 analisa dari kata tanya where.

- Parameter input, hampir semua sistem mampu mendeteksi topik pertanyaan
- Parameter opt, sistem memberikan rekomendasi lebih dari 1 judul artikel wikipedia pada 7 kali uji coba, sisanya sistem mampu mengambil kesimpulan sendiri judul artikel yang sesuai dengan pertanyaan.
- Parameter key, leksikal yang didapat pada babelnet paling sedikit yaitu 19 kata dan terbanyak dengan jumlah 93 kata
- Parameter kalimat artikel yang didapat didalam wikipedia kalimat paling sedikit yaitu 27 kalimat dan terbanyak yaitu 312 kalimat.
- Parameter waktu, sistem mampu menghasilkan jawaban dengan waktu tercepat yang dibutuhkan sebanyak 58 detik, dan waktu terlama yaitu 201 detik.

Tabel 4.9 Analisa kata tanya “Where”

No	input	opt	Σ Key	Σ Kalimat	Waktu (s)
1		✓	30	312	58
2		✓	28	312	64
3		✓	19	304	66
4			30	252	67
5			30	270	72
6		✓	30	62	76
7			41	27	88
8		✓	50	104	94
9		✓	21	264	95
10	✓	✓	93	61	201

5. Who (Siapa)

Kata tanya “Who” digunakan untuk pertanyaan yang memberikan jawaban bertipe orang. Pada Tabel 4.10 hasil pengukuran uji coba ini precisionnya bernilai 100%, recall 100% dan Accuracy 100%. Bahwa pada uji coba kata tanya when jawaban sering muncul di cosine pada urutan pertama. Dengan dilakukan percobaan 10x terjadi kesalahan pada pertanyaan yang tidak tepat sehingga tidak dapat menghasilkan jawaban yang sesuai sebanyak 1 kali.

Tabel 4.10 Uji Coba Pertanyaan menggunakan kata tanya “Who”.

NO	PERTANYAAN	Jawaban	Posisi Jawaban		JAWABAN			
			Sparql	Cosine	TP	FP	FN	FN
1	Who is the teacher of Pythagoras ?	On the grounds of all these references connecting Pythagoras with Pherecydes, Riedweg concludes that there may well be some historical foundation to the tradition that Pherecydes was Pythagoras's teacher.	0	1	✓			
2	Who are the students of Avicenna?	(The only critical edition of Ibn Sina's autobiography, supplemented with material from a biography by his student Abu 'Ubayd al-Juzjani. Al-Biruni correspondence Correspondence between Ibn Sina (with his student Ahmad ibn 'Ali al-Ma'sumi) and Al-Biruni has survived in which they debated Aristotelian natural philosophy and the Peripatetic school	0	1 & 2	✓			
3	Who is the brother of Ibn Khaldun?	His brother, Yahya Khaldun	0	1	✓			
4	Who developed the Fibonacci intervals into melody?	Jawaban tidak ada karena memang didalam artikel FIBONACCI tidak ada pembahasan tentang sejarah pengembangan fibonacci ke dalam melody.	0	0				
5	Who developed Calculus in the 17th century?	during the 17th century, when Isaac Newton and Gottfried Wilhelm Leibniz built on the work of earlier mathematicians to introduce its basic principles. History Modern calculus was developed in 17th-century Europe by Isaac Newton and Gottfried Wilhelm Leibniz	0	1 & 3	✓			

6	Who use geometry to calculate the pyramid?	South of Egypt the ancient Nubians established a system of geometry including early versions of sun clocks. In the 7th century BC, the Greek mathematician Thales of Miletus used geometry to solve problems such as calculating the height of pyramids and the distance of ships from the shore.	0	1	✓		
7	Who said Mathematics as "Queen of Science" ?	Mathematics as science The German mathematician Carl Friedrich Gauss referred to mathematics as "the Queen of the Sciences".	0	1 & 2	✓		
8	Who uses the form of Exponentiation notation in the 15th century?	the notation is introduced in Book I. Nicolas Chuquet used a form of exponential notation in the 15th century	0	1	✓		
9	Who included 1 as the first prime number in the mid 18th century?	In the mid-18th century Christian Goldbach listed 1 as prime in his correspondence with Leonhard Euler; however, Euler himself did not consider 1 to be prime	0	1	✓		
10	Who wrote the book on Statistics titled "Manuscript on Deciphering Cryptographic Messages" ?	History The earliest writing on statistics was found in a 9th-century book entitled Manuscript on Deciphering Cryptographic Messages, written by Arab scholar Al-Kindi (801–873).	0	1	✓		

Berikut hasil dari Tabel 4.11 analisa dari kata tanya who:

- Parameter input, hampir sebagian sistem tidak mampu mendeteksi topik pertanyaan sebanyak 6 kali.
- Parameter opt, sistem memberikan rekomendasi lebih dari 1 judul artikel pada 7 kali uji coba, sisanya dapat mengambil kesimpulan judul yang sesuai dengan pertanyaan.
- Parameter key, leksikal yang didapat pada babelnet paling sedikit yaitu 16 kata dan terbanyak dengan jumlah 82 kata
- Parameter kalimat artikel yang didapat didalam wikipedia kalimat paling sedikit yaitu 186 kalimat dan terbanyak yaitu 312 kalimat.
- Parameter waktu, sistem mampu menghasilkan jawaban dengan waktu tercepat yang dibutuhkan sebanyak 52 detik, dan waktu terlama yaitu 199 detik.

Tabel 4.11 Analisa kata tanya “Who”

No	input	opt	Σ Key	Σ Kalimat	Waktu (s)
1		✓	19	312	52
2		✓	16	304	57
3			19	255	73
4	✓	✓	36	295	81
5	✓	✓	37	237	97
6			49	186	98
7	✓	✓	52	211	117
8	✓	✓	67	365	167
9	✓		78	322	193
10	✓	✓	82	287	199

6. Question Yes or No

Kalimat tanya digunakan yang digunakan untuk memvalidasi kebenarannya. Jawaban dianggap True Positif jika mampu menampilkan kalimat yang sesuai dengan pernyataan kalimat tanya tersebut. Pada Tabel 4.12 hasil pengukuran pada uji coba tersebut precisionnya bernilai 88.89%, recall 100% dan Accuracy 90%, dengan jawaban sering muncul pada cosine di urutan pertama. Dengan dilakukan percobaan 10x terjadi kesalahan pada sistem sebanyak 1 kali dan pertanyaan yang tidak tepat sebanyak 1 kali.

Tabel 4.12 Uji Coba Pertanyaan “Yes or No”.

No	Pertanyaan	Jawaban	Posisi jawaban		Jawaban			
			Sparql	Cosine	TP	FP	TN	FN
1	was Pythagoras born in Cairo?	However, Ulugh Beg later measured another more precise value of the tropical year as 365d 5h 49m 15s, which has an error of +25 seconds, making it more accurate than Copernicus's estimate which had an error of +30 seconds.	1	1	✓			
2	Whether exponential notation is used in the 15th century?	Nicolas Chuquet used a form of exponential notation in the 15th century, which was later used by Henricus Grammateus and Michael Stifel in the 16th century.	0	1	✓			
3	has cube 8 vertices?	Point in space For a cube whose circumscribing sphere has radius R, and for a given point in its 3-dimensional space with distances d_i from the cube's eight vertices	0	1	✓			
4	Is cube two dimensional object ?	In geometry, a cube is a three-dimensional solid object bounded by six square faces, facets or sides, with three meeting at each vertex.	0	1	✓			
5	Did Pythagoras divide the world into three zones	, It was said that he was the first man to call himself a philosopher ("lover of wisdom") and that he was the first to divide the globe into five climatic zones.	0	1	✓			

6	Whether al-Khwarizmi Physics scientist?	Tidak ada jawabannya dikarena di dalam artikel Al Khwarizmi tidak ada kalimat yang memverifikasi pertanyaan tersebut	0	0			✓	
7	Does Carl Gauss refer to mathematics as Queen of Science?	He referred to mathematics as "the queen of sciences"[66] and supposedly once espoused a belief in the necessity of immediately understanding Euler's identity as a benchmark pursuant to becoming a first-class mathematician	0	0		✓		
8	Does Ulugh beg determine the length of the year as 365?	However, Ulugh Beg later measured another more precise value of the tropical year as 365d 5h 49m 15s, which has an error of +25 seconds, making it more accurate than Copernicus's estimate which had an error of +30 seconds.	0		✓			
9	Whether in the 18th century mathematicians use one as a prime number	In the mid-18th century Christian Goldbach listed 1 as prime in his correspondence with Leonhard Euler; however, Euler himself did not consider 1 to be prime.	0	1	✓			
10	Is the speed of light faster than Earth in its orbit?	From this effect he determined that light must travel 10210 times faster than the Earth in its orbit (the modern figure is 10066 times faster) Bradley used this method to derive that light travelled 10210 times faster than the Earth in its orbit (the modern figure is 10066 times faster)	0	1.5	✓			

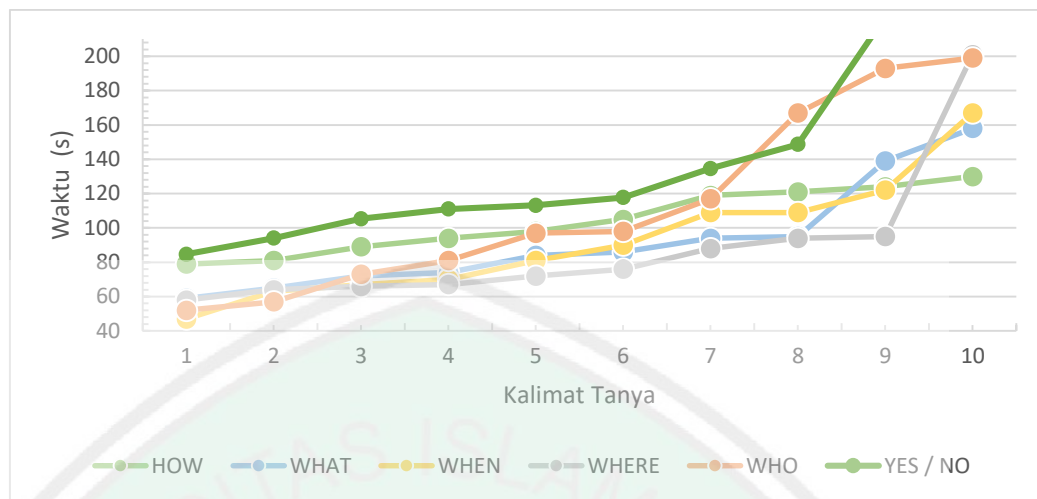
Berikut hasil dari Tabel 4.13 analisa dari kata tanya *Yes or No*:

- Parameter input, hampir sebagian sistem tidak mampu mendeteksi topik pertanyaan sebanyak 6 kali.
- Parameter opt, uji coba yang dilakukan berimbang sebagian sistem memberikan rekomendasi, sebagian mampu mengambil kesimpulan sendiri pada judul artikel.
- Parameter key, leksikal yang didapat pada babelnet paling sedikit yaitu 29 kata dan terbanyak dengan jumlah 98 kata
- Parameter kalimat artikel yang didapat didalam wikipedia kalimat paling sedikit yaitu 65 kalimat dan terbanyak yaitu 265 kalimat.
- Parameter waktu, sistem mampu menghasilkan jawaban dengan waktu tercepat yang dibutuhkan sebanyak 85 detik, dan waktu terlama yaitu 241 detik.

Tabel 4.13 Analisa kata tanya "*Yes or No*".

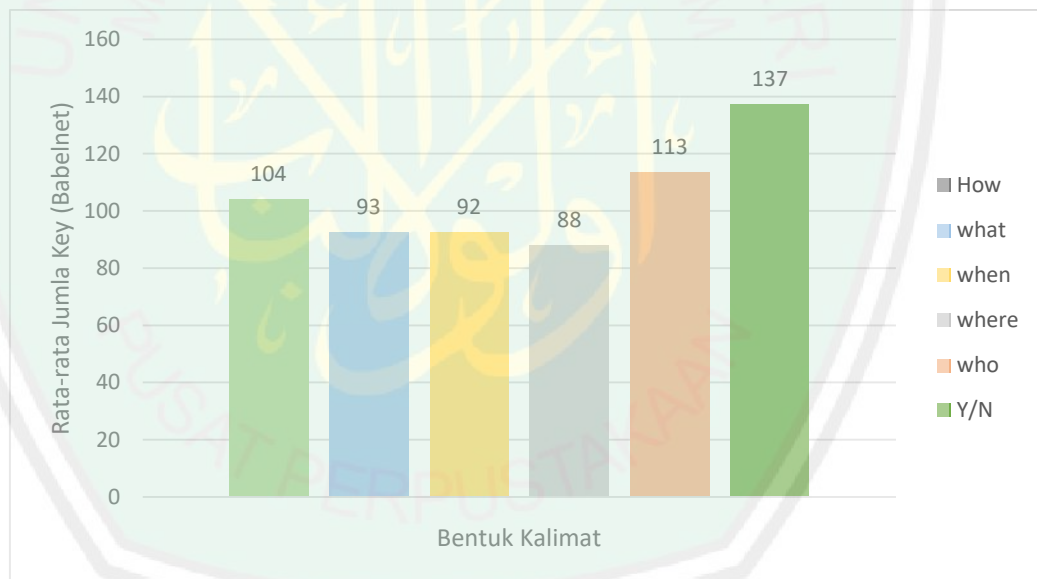
No	input	opt	Σ Key	Σ Kalimat	Waktu (s)
1			29	309	85
2	✓	✓	41	387	94
3			54	65	105
4	✓	✓	54	65	111
5		✓	66	312	113
6	✓		31	317	118
7			51	322	135
8	✓	✓	57	104	149
9	✓		91	322	223
10	✓	✓	98	265	241

Dari uji coba yang dilakukan di berbagai pertanyaan, dapat divisualisasika waktu eksekusinya seperti pada Gambar 4.9. Waktu tercepat yaitu 47 detik di kalimat tanya when, sedangkan waktu terlama yaitu 201 detik di kalimat tanya Y/N



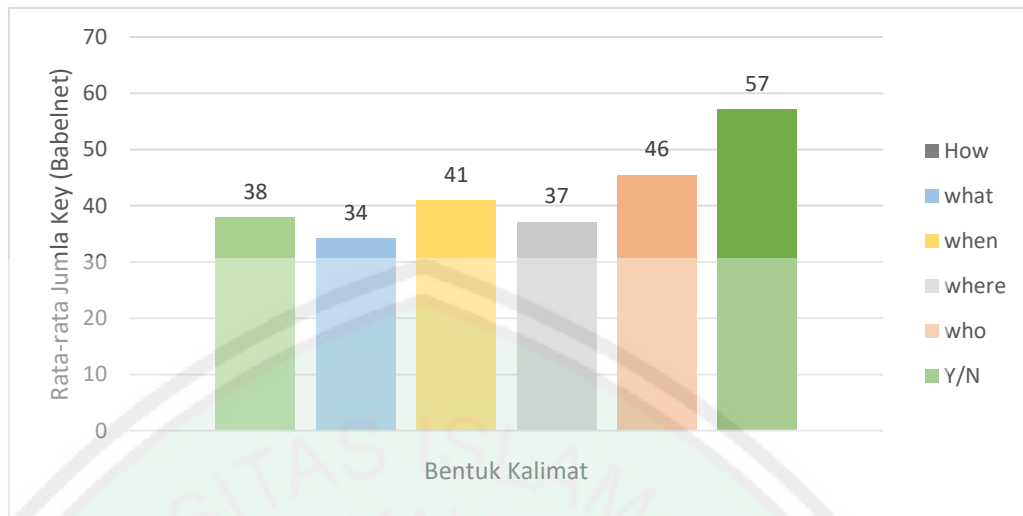
Gambar 4.9 Grafik waktu eksekusi

Gambar 4.10 merupakan waktu eksekusi pada setiap pertanyaan dirata-rata. Maka waktu tercepat yang dihasilkan terdapat pada kata tanya where. Dan untuk waktu terlama ada di kata tanya Y/N.



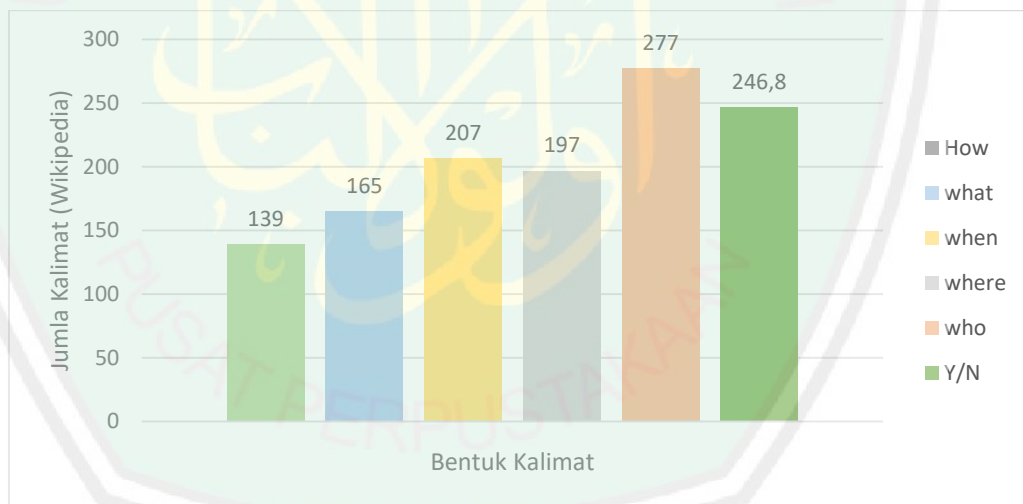
Gambar 4.10 Grafik rata-rata waktu eksekusi

Gambar 4.11 merupakan grafik rata-rata jumlah key dari Babelnet pada setiap uji coba bentuk kalimat pertanyaan. Kata terbanyak didapatkan di kata tanya who. Sedangkan yang paling sedikit ada pada kata tanya what.



Gambar 4.11 Grafik rata-rata jumlah key dari Babelnet

Gambar 4.12 merupakan grafik rata-rata hasil jumlah kalimat dari *crawling* Wikipedia. Kalimat terbanyak ada di kata tanya *who*. Sedangkan yang paling sedikit ada pada kata tanya *how*.



Gambar 4.12 Grafik rata-rata jumlah kalimat dari Wikipedia

Dari hasil pengujian semua data pada tabel diatas, maka jumlah keseluruhan nilai precision recall dan accuracynya adalah sebagai berikut

Tabel 4.14 Hasil *Precision*, *Recall*, dan *Accuracy*

<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>
87.23 %	100 %	87.76 %

4.4 Analisa Hasil dan Pembahasan

Berdasarkan uji coba yang telah dilakukan, bahwa sistem tanya jawab ini memiliki tingkat keakurasian mencapai 87,76%. Adapaun beberapa masalah yang terjadi didalam sistem ini.

A. Masalah didalam pengklasifikasian, seperti pada penelitiannya Makkar, dkk (2017) yaitu sebagai berikut:

1. Biaya komputasi yang tinggi. Semakin datanya banyak, maka semakin lama waktu yang dibutuhkan.
2. CNN adalah salah satu teknik yang kuat dan mencerminkan akurasi klasifikasi tinggi tetapi membutuhkan banyak data pelatihan

B. Penyebab jawaban dari dbpedia tidak muncul atau salah, hampir sama dengan penelitiannya Ketsmur, dkk (2017) yaitu sebagai berikut:

1. Ontology tidak tersedia atau belum ada diDBpedia. (Contoh : pada Dbpedia <http://dbpedia.org/page/Al-Khazini>, tidak ada class **deathPlace** yang menunjukkan tempat meninggalnya seseorang, sedangkan di <http://dbpedia.org/page/Al-Kindi> terdapat class ontology **deathPlace**)
2. Nama ontology pada Dbpedia bersifat khusus. (Contoh: **dbo:abstract**, nama class ontoloy, dengan *valuenya* berisi tentang definisi, penjelasan ringkas pada suatu artikel di wikipedia).
3. SPARQL menggunakan regex, untuk mencari atau mencocokkan nama ontology di Dbpedia dengan entitynya yaitu hasil class CNN, *word graph*, dan leksikal.

C. Penyebab ketidakakuratan hasil jawaban dari wikipedia, terjadi karena beberapa faktor yang sama juga dengan penelitiannya Purwarianti dan Yusliani (2012), yaitu sebagai berikut:

1. Semakin banyak *term* dalam 1 kalimat, maka kalimat tsb berpeluang memiliki nilai teratas dan menjadi sebuah jawaban. Contoh kalimat tanya “*What type of Paraboloid?*”. Pada percobaan ini tidak menghasilkan jawaban yang baik dikarenakan, dalam artikel tersebut banyak kata yang mengandung Paraboloid dalam 1 kalimat. Sehingga kalimat tersebut memiliki peluang untuk sebuah jawaban.
2. Kalimat yang salah dengan pola jawaban yang benar memiliki nilai yang tinggi. Contoh pada percobaan kata tanya “*How many seconds in a minute?*”. Seharusnya jawaban yang benar adalah “*As a unit of time, the minute is most of times equal to 1/60*”. Tapi kandidat jawaban yang mendekati benar adalah “*In the UTC time standard, a minute on rare occasions has 61 seconds, a consequence of leap seconds*”, yang artinya jawabannya kurang tepat.
3. Nilai *Cosinus Similarity* sama antara kandidat jawaban yang 1 dengan yang lainnya. Contoh kalimat “*How fast is the speed of light?*”, jawabannya berada di posisi 3 dan 4.
4. Jawaban yang dihasilkan tidak berada pada urutan pertama. Contoh pada uji coba pertanyaan “*Where was Al Farabi died?*”, jawaban yang benar adalah “*He later spent time in Damascus and in Egypt before returning to Damascus where he died in 950-1.His name was Abū Naṣr Muḥammad b. Muḥammad Farabi, sometimes with the family surname al-Ṭarḳānī, i.e., the element Ṭarḳān appears in a nisba.*”. jawaban tersebut berada di posisi ke 4.

D. Penyebab berdurasi lama yaitu disebabkan oleh :

1. Kecepatan internet, semakin lambat kecepatannya maka semakin lama pula untuk meng-*crawl*-ing data dari wikipedia dan babelnet.

2. Jumlah kata kunci yang didapat dari babelnet terlalu banyak. Sehingga pencarian topik pada CNN, jawaban melalui SPARQL dan Wikipedia semakin memakan waktu
3. Jumlah Kalimat dari wikipedia juga dapat mempengaruhinya. Semakin banyak kalimatnya, semakin pula untuk mencari jawabannya semakin lama.

Sehingga dari beberapa masalah tersebut, pada sistem ini ditampilkan kemungkinan jawaban pada pertanyaan pengguna lebih dari 1. Dan kata kunci, kata leksikal memiliki pengaruh yang tinggi untuk menemukan sebuah jawaban. Semakin tinggi nilai lexical pada suatu kalimat didalam artikel wikipedia, maka semakin menjadi suatu jawaban. Seperti menurut penelitiannya Jumadi (2016), bahwa *cosine similarity* memiliki konsep normalisasi panjang vektor data dengan membandingkan n-gram yang sejajar satu sama lain dari 2 pembanding.

Sistem ini jika dilihat dari sudut pandang islam yaitu seperti di dalam Al Quran, yang mana Allah menurunkan wahyunya tentang kisah dari rasul terdahulu dapat menjadi pembelajaran umat manusia saat ini di masa depan. Seperti pada ayat Al Quran dibawah ini.

وَحَقُّ آلِ بَدِئِهِ فِي جَاءِكَ وَفُؤَادِكَ بِمِ نُتَبِّتَ مَا الرُّسُلِ أَنْبَاءٍ مِنْ عَلَيْكَ تَقْصُ كُلاًَّ وَ
لِلْمُؤْمِنِينَ ذِكْرًا وَ مَوْعِظَةً

Artinya: “Dan semua kisah dari rasul-rasul Kami ceritakan kepadamu, ialah kisah-kisah yang dengannya Kami teguhkan hatimu, dan dalam surat ini telah datang kepadamu kebenaran serta pengajaran dan peringatan bagi orang-orang yang beriman.”. (QS. Hud 11:120)

Menurut Jalaluddin al-Mahalli & Jalaluddin as-Suyuthi, (Dan setiap) lafal kullan ini dinashabkan dengan alamat naqsh sedangkan tanwinnya merupakan pergantian dari mudhaf ilaih, artinya semua kisah rasul-rasul yang diperlukan (Kami ceritakan kepadamu, yaitu kisah-kisah para rasul) lafal maa di sini menjadi badal daripada lafal kullan (yang dengannya Kami teguhkan) Kami tenangkan (hatimu) kalbumu

(dan dalam surah ini telah datang kepadamu kebenaran) yang dimaksud adalah kisah-kisah para rasul ini atau ayat-ayat ini (serta pengajaran dan peringatan bagi orang-orang yang beriman) orang-orang yang beriman disebutkan di sini secara khusus, mengingat hanya merekalah yang dapat memanfaatkan adanya kisah-kisah atau ayat-ayat ini untuk mempertebal keimanan mereka, berbeda dengan orang-orang kafir.

Allah juga meminta hambanya untuk mempelajari suatu ilmu tidak tergesa-gesa, dan memintalah kepada-Nya akan tambahan ilmu. Seperti pada ayat Al Quran di bawah ini.

وَيُوحِ إِلَيْكَ يُقْضَىٰ أَنْ قَبْلَ مِنْ بِالْقُرْآنِ تَعْجَلْ لَا وَءَ الْحَقُّ الْمَلِكُ اللَّهُ فَتَعْلَىٰ
عِلْمًا زِدْنِي رَبِّ قُلْ

Artinya: *Maka Maha Tinggi Allah Raja Yang sebenar-benarnya, dan janganlah kamu tergesa-gesa membaca Al Qur'an sebelum disempurnakan mewahyukannya kepadamu, dan katakanlah: "Ya Tuhanku, tambahkanlah kepadaku ilmu pengetahuan". (QS. Taha 20:114)*

Firman Allah subhanahu wa ta'ala "Maka Mahatinggi Allah, Raja yang sebenar-benarnya". Artinya, Mahasuci Allah, Raja yang sebenar-benarnya, janji-Nya benar, ancaman-Nya benar, rasul-rasul-Nya benar, surga benar, neraka benar (adanya), dan segala sesuatu yang datang dari-Nya adalah benar belaka. Sifat Mahaadil Allah ialah Dia tidak mengazab seseorang sebelum memberikan peringatan dan mengutus rasul-rasul-Nya dan sebagai alasan-Nya kepada makhluk-Nya, agar tidak ada lagi hujah dan keraguan bagi seorang pun terhadap apa yang telah diputuskan oleh-Nya kelak.

Firman Allah subhanahu wa ta'ala, "Dan janganlah kamu tergesa-gesa membaca Al-Qur'an sebelum disempurnakan mewahyukannya kepadamu". Ayat ini semakna dengan apa yang disebutkan oleh Allah subhanahu wa ta'ala dalam surat lainnya yang mengatakan: Janganlah kamu gerakkan lidahmu untuk (membaca) Al-Qur'an

karena hendak cepat-cepat (menguasai)nya Sesungguhnya atas tanggungan Kamilah mengumpulkannya (di dadamu) dan (membuatmu pandai) membacanya. Apabila Kami telah selesai membacaknya, maka ikutilah bacaannya itu. Kemudian atas tanggungan Kamilah penjelasannya. (Q.S. Al-Qiyaamah [75]: 16-19).

Dalam surat berikut ini disebutkan oleh firman-Nya, *“dan janganlah kamu tergesa-gesa membaca Al-Qur’an sebelum disempurnakan mewahyukannya kepadamu”*. Melainkan dengarlah dengan penuh perhatian. Apabila malaikat telah selesai membacaknya kepadamu, mulailah kamu membacanya. *“... dan katakanlah, “Ya Tuhanku, tambahkanlah kepadaku ilmu pengetahuan.”*. Maksudnya, berilah aku tambahan ilmu dari-Mu. Ibnu Uyaynah mengatakan bahwa Rasulullah ﷺ terus-menerus mendapat tambahan ilmu hingga Allah subhanahu wa ta’ala mewafatkannya. Karena itulah di dalam sebuah hadis telah disebutkan: Sesungguhnya Allah menurunkan wahyu kepada Rasul-Nya secara berturut-turut, sehingga wahyu banyak diturunkan di hari-hari beliau menjelang wafatnya.

Jadi sistem tanya jawab ini diharapkan dapat memberikan pengetahuan sejarah matematika dan biografi ilmuwan matematika. Agar dapat memberikan wawasan tambahan kepada pengguna.

BAB IV

KESIMPULAN

5.1 Kesimpulan

Hasil penelitian dari sistem tanya jawab yang sudah dilakukan dengan menggunakan metode *Deep Graph Convolutional Neural Network* (DGCNN) yang berbasis web. Bentuk pertanyaan yang diajukan untuk sistem ini mulai dari how, what, when, where, who. Adapun Selain itu metode tersebut mampu menjawab pertanyaan dengan kata tanya yes dan no.

Namun ada beberapa faktor yang mempengaruhi nilai akurasi dan kecepatan eksekusi. Faktor yang mempengaruhi nilai akurasi seperti belum lengkapnya ontology pada dbpedia, sistem yang belum mampu mendeteksi kata ganti orang ketiga, jumlah kata dalam suatu kalimat yang dicari juga mempengaruhi akurasi kebenaran. Untuk faktor dari kecepatan eksekusi yaitu kecepatan internet (bandwith) yang kurang.

Sehingga akurasi yang didapatkan pada penelitian ini sebesar 87, 76%. Akurasi tersebut didapatkan dari 60 percobaan yang dilakukan.

5.2 Saran

Untuk pengembangan sistem tanya jawab ini di kemudian hari diperlukan beberapa tambahan atau perbaikan untuk mendapatkan akurasi yang lebih maksimal, yaitu sebagai berikut:

1. Menambahkan metode pendeteksian untuk kata ganti orang, seperti dia, mereka dan lain-lain. Agar jika didalam dokumen ada kalimat yang memiliki kata ganti, maka bisa langsung merujuk ke kalimat sebelumnya
2. Membangun aplikasi dengan bentuk kalimat tanya non-factoid. Kalimat tanya yang memberikan jawaban yang lebih luas, seperti why, how to, dan lain-lain.

REFERENSI

1. Abdiansah, Abdiansah, dan Anny K. Sari. 2015. "Survei: Question Classification untuk Question Answering System." *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)* 1 (1). <http://journal.uui.ac.id/Snati/article/view/3531>.
2. Agarap, Abien Fred. 2018. "Deep Learning using Rectified Linear Units (ReLU)." *arXiv:1803.08375 [cs, stat]*, Maret. <http://arxiv.org/abs/1803.08375>.
3. Archana, S. M., Naima Vahab, Rekha Thankappan, dan C. Raseek. 2016. "A Rule Based Question Answering System in Malayalam Corpus Using Vibhakthi and POS Tag Analysis." *Procedia Technology*, International Conference on Emerging Trends in Engineering, Science and Technology (ICETEST - 2015), 24 (Januari): 1534–41. <https://doi.org/10.1016/j.protcy.2016.05.124>.
4. Biswas, P., A. Sharan, dan R. Kumar. 2014. "Question Classification using syntactic and rule based approach." Dalam *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 1033–38. <https://doi.org/10.1109/ICACCI.2014.6968434>.
5. Bouziane, Abdelghani, Djelloul Bouchiha, Nouredine Doumi, dan Mimoun Malki. 2015. "Question Answering Systems: Survey and Trends." *Procedia Computer Science*, International Conference on Advanced Wireless Information and Communication Technologies (AWICT 2015), 73 (Januari): 366–75. <https://doi.org/10.1016/j.procs.2015.12.005>.
6. Breck, Eric, John D. Burger, Lisa Ferro, Lynette Hirschman, David House, Marc Light, dan Inderjeet Mani. 2000. "How to Evaluate Your Question Answering System Every Day and Still Get Real Work Done," April. <https://arxiv.org/abs/cs/0004008>.
7. Casteel, Clifton A. 1988. "Effects of Chunked Reading among Learning Disabled Students: An Experimental Comparison of Computer and Traditional

- Chunked Passages.” *Journal of Educational Technology Systems* 17 (2): 115–21. <https://doi.org/10.2190/EBNP-6Q4W-1BJW-G5HM>.
8. Danukusumo, Kefin Pudi. 2017. “Implementasi Deep Learning Menggunakan Convolutional Neural Network Untuk Klasifikasi Citra Candi Berbasis Gpu.” S1, UAJY. <http://e-journal.uajy.ac.id/12425/>.
 9. Hirschman, L., dan R. Gaizauskas. 2001. “Natural Language Question Answering: The View from Here.” *Nat. Lang. Eng.* 7 (4): 275–300. <https://doi.org/10.1017/S1351324901002807>.
 10. Ho, H., V. C. Mawardi, dan A. B. Dharmawan. 2017. “Question answering system with Hidden Markov Model speech recognition.” Dalam *2017 3rd International Conference on Science in Information Technology (ICSITech)*, 257–62. <https://doi.org/10.1109/ICSITech.2017.8257121>.
 11. Ingason, Anton Karl, Sigrún Helgadóttir, Hrafn Loftsson, dan Eiríkur Rögnvaldsson. 2008. “A Mixed Method Lemmatization Algorithm Using a Hierarchy of Linguistic Identities (HOLI).” Dalam *Advances in Natural Language Processing*, disunting oleh Bengt Nordström dan Aarne Ranta, 5221:205–16. Catatan Kuliah Dalam, Ilmu Komputer. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-85287-2_20.
 12. Iyyer, Mohit, Jordan Boyd-Graber, Leonardo Claudino, Richard Socher, dan Hal Daumé III. 2014. “A Neural Network for Factoid Question Answering over Paragraphs.” Dalam , 633–44. <https://doi.org/10.3115/v1/D14-1070>.
 13. Kipf, Thomas N., dan Max Welling. 2016. “Semi-Supervised Classification with Graph Convolutional Networks.” *arXiv:1609.02907 [cs, stat]*, September. <http://arxiv.org/abs/1609.02907>.
 14. Manning, Christopher, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, dan David McClosky. 2014. “The Stanford CoreNLP Natural Language Processing Toolkit.” Dalam *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 55–60. Baltimore, Maryland: Association for Computational Linguistics. <https://doi.org/10.3115/v1/P14-5010>.

15. Navigli, Roberto, dan Simone Paolo Ponzetto. 2012. "BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network." *Artificial Intelligence* 193 (Desember): 217–50. <https://doi.org/10.1016/j.artint.2012.07.001>.
16. Noh, Hyeonwoo, Paul Hongsuck Seo, dan Bohyung Han. 2015. "Image Question Answering using Convolutional Neural Network with Dynamic Parameter Prediction." *arXiv:1511.05756 [cs]*, November. <http://arxiv.org/abs/1511.05756>.
17. Novreni, Anggraini. 2012. "RANCANG BANGUN APLIKASI QUESTION ANSWERING (QA) SYSTEM PADA TERJEMAHAN AL QURAN MENGGUNAKAN EPHYRA FRAMEWORK." Skripsi, UNIVERSITAS ISLAM NEGERI SULTAN SYARIEF KASIM RIAU. <http://repository.uin-suska.ac.id/1101/>.
18. Ns, Abdiansah, dan Edi Winarko. 2015. "Question Classification Menggunakan Support Vector Machines dan Stemming." Dalam . ResearchGate.
19. Peng, Hao, Jianxin Li, Yaopeng Liu, Mengjiao Bao, Lihong Wang, Yangqiu Song, dan Qiang Yang. 2018. "Large-Scale Hierarchical Text Classification with Recursively Regularized Deep Graph-CNN - Semantic Scholar." *Semantic Scholar*. 2018. <https://www.semanticscholar.org/paper/Large-Scale-Hierarchical-Text-Classification-with-Peng-Li/a1a197449aeca81a39cb2213b41cef4831d6983e>.
20. Purwarianti, Ayu, dan Novi Yusliani. 2012. "SISTEM QUESTION ANSWERING BAHASA INDONESIA UNTUK PERTANYAAN NON-FACTOID." *Jurnal Ilmu Komputer dan Informasi* 4 (1): 10–14. <https://doi.org/10.21609/jiki.v4i1.151>.
21. Rawat, Waseem, dan Zenghui Wang. 2017. "Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review." *Neural Computation* 29 (9): 2352–2449. https://doi.org/10.1162/neco_a_00990.

22. Rodrigo, Álvaro, Joaquín Pérez-Iglesias, Anselmo Peñas, Guillermo Garrido, dan Lourdes Araujo. 2013. "Answering questions about European legislation." *Expert Systems with Applications* 40 (15): 5811–16. <https://doi.org/10.1016/j.eswa.2013.05.008>.
23. Segaran, Toby, Colin Evans, Jamie Taylor, Segaran Toby, Evans Colin, dan Taylor Jamie. 2009. *Programming the Semantic Web*. 1st ed. O'Reilly Media, Inc.
24. Setiawan, Ryan Agus, dan Kristoforus Jawa Bendi. 2014. "CHITCATCINEMA: APLIKASI QUESTION ANSWERING SYSTEM UNTUK DOMAIN FILM BIOSKOP." ResearchGate. Mei 2014. https://www.researchgate.net/publication/290222527_CHITCATCINEMA_APLIKASI_QUESTION_ANSWERING_SYSTEM_UNTUK_DOMAIN_FILM_BIOSKOP.
25. Sharma, Vijay Kumar, dan Namita Mittal. 2016. "Exploiting Wikipedia API for Hindi-english Cross-language Information Retrieval." *Procedia Computer Science*, Twelfth International Conference on Communication Networks, ICCN 2016, August 19– 21, 2016, Bangalore, India Twelfth International Conference on Data Mining and Warehousing, ICDMW 2016, August 19-21, 2016, Bangalore, India Twelfth International Conference on Image and Signal Processing, ICISP 2016, August 19-21, 2016, Bangalore, India, 89 (Januari): 434–40. <https://doi.org/10.1016/j.procs.2016.06.094>.
26. Skeith, Leslie, Andrew Aw, Julia Hews-Girard, dan Natalia Rydz. 2017. "A case that illustrates the challenges of managing pregnant patients with antithrombin deficiency: More questions than answers." *Thrombosis Research* 157 (September): 1–6. <https://doi.org/10.1016/j.thromres.2017.06.029>.
27. Teixeira, António, Maksym Ketsmur, dan Mário Rodrigues. 2017. "A QUERY/ANSWERING SYSTEM FOR FACTUAL QUERIES IN PORTUGUESE ON DBPEDIA." Dalam .
28. Triana, Ana, Ristu Saptono, dan Meiyanto Eko Sulisty. 2016. "Pemanfaatan Metode Vector Space Model dan Metode Cosine Similarity pada Fitur Deteksi

- Hama dan Penyakit Tanaman Padi.” *ITSMART: Jurnal Teknologi dan Informasi* 3 (2): 90–95. <https://doi.org/10.20961/itsmart.v3i2.704>.
29. Wahyu, Kamirsyah, dan Sofyan Mahfudy. 2016. “SEJARAH MATEMATIKA: ALTERNATIF STRATEGI PEMBELAJARAN MATEMATIKA” 9: 89–110.
30. Xu, Ying. 2017. “Relation Extraction and its Application to Question Answering.” *ERA*. Juni 2017. <https://doi.org/10.7939/R3QB9VJ17>.
31. Yang, Min-Chul, Do-Gil Lee, So-Young Park, dan Hae-Chang Rim. 2015. “Knowledge-based question answering using the semantic embedding space.” *Expert Systems with Applications* 42 (23): 9086–9104. <https://doi.org/10.1016/j.eswa.2015.07.009>.
32. Zheng, Zhiping. 2003. “Question Answering Using Web News as Knowledge Base.” *ResearchGate*. 2003. <http://dx.doi.org/10.3115/1067737.1067797>.
33. Zhou, Guangyou, Yin Zhou, Tingting He, dan Wensheng Wu. 2016. “Learning semantic representation with neural networks for community question answering retrieval.” *Knowledge-Based Systems* 93 (Februari): 75–83. <https://doi.org/10.1016/j.knosys.2015.11.002>.