PROPOSAL PENCARIAN INFORMASI MEDIA ONLINE

Kualitas Jawaban Pada Philoit



KELOMPOK:

1772041 - Natasya Dey 1872021 - Hosea Adipratama Santoso 1872043 - Dinda Ayu Febriani 1872045 - Yolanda Trixie Hernita

PROGRAM STUDI S1 TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI UNIVERSITAS KRISTEN MARANATHA BANDUNG 2021

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Philoit merupakan sebuah aplikasi yang berdiri pada tahun 2019 sebagai platform tanya-jawab, yang ditujukan untuk para milenial dapat berbagi dan berdiskusi tentang informasi apapun yang diketahuinya. Ide aplikasi ini lahir dari 'unconscious problem' atau fenomena yang tidak disadari oleh masyarakat, para generasi muda indonesia banyak menghabiskan energinya kepada konten-konten yang tidak membangun literasi dan pengetahuan dalam dunia digital. Dari permasalahan ini, aplikasi dibangun agar generasi muda mendapatkan informasi dan aktif kepada literasi dan pengetahuan tetapi dengan konsep yang menyenangkan.

Dilakukan eksplorasi terhadap aplikasi *Philoit*, dilihat terdapat ketertarikan para pengguna terutama generasi muda untuk dapat berdiskusi dan berbagi informasi tentang berbagai macam topik menarik melalui aplikasi ini, sehingga dilakukan analisis terhadap pertanyaan dan jawaban dari topik-topik yang terdapat dalam aplikasi. Setelah melakukan analisa, maka didapat beberapa jawaban yang tidak sesuai dengan topik atau pertanyaan yang diajukan, sehingga membuat kualitas jawaban menjadi tidak relevan.

Dari hasil analisis tersebut, maka dilakukan penelitian untuk menangani permasalahan, dilakukan analisa dengan *text mining* menggunakan "Python" untuk mengidentifikasi setiap kata dari jawaban, sehingga didapat beberapa kata yang kemudian menjadi kata kunci. Dari proses tersebut, kata kunci dilakukan pencocokan dengan pertanyaan yang kemudian dikategorikan sebagai jawaban yang berkualitas atau layak.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah, maka dapat diambil kesimpulan rumusan masalah yaitu :

1. Bagaimana cara mengurangi kualitas jawaban yang tidak relevan dari topik pertanyaan?

1.3 Tujuan

Berdasarkan rumusan masalah, maka dapat diambil tujuan dari penelitian yaitu:

1. Untuk mendapatkan jawaban yang berkualitas sesuai dengan pertanyaan.

1.4 Manfaat

Berikut adalah manfaat dari penelitian, yaitu:

- Agar mempermudah pengguna dalam mencari solusi jawaban sesuai dengan topik
- 2. Agar pengguna dari *Philoit* memberikan jawaban yang berkualitas.
- 3. Agar Pengunjung *Philoit* meningkat dikarenakan *platform* dan jawaban yang berkualitas.

1.5 Ruang Lingkup

Ruang lingkup yang digunakan pada pembuatan laporan ini adalah sebagai berikut :

1. Dapat menganalisis jawaban dari topik pertanyaan untuk membandingkan suatu jawaban tersebut relevan atau tidak relevan.

1.6 Sumber Data

Sumber data yang digunakan dalam penelitian ini adalah sebagai berikut :

1. Sumber data primer

Data didapatkan melalui penghimpunan hasil data selama 1 tahun terakhir

2. Sumber data sekunder

Sumber data diperoleh dari situs-situs di internet serta referensi ilmiah yang mendukung penelitian ini.

1.7 Sistematika Penyajian

Sistematika penyajian laporan penelitian yang dibuat untuk mempermudah dalam penulisan laporan supaya terarah dan terfokus. Sistematika penyajian laporan penelitian adalah sebagai berikut :

BAB 1: PENDAHULUAN

Bab ini membahas tentang latar belakang, rumusan masalah, tujuan, ruang lingkup kajian, sumber data, serta sistematika pembahasan dari penelitian ini.

BAB 2: KAJIAN TEORI

Bab ini menjelaskan mengenai pengertian dan penjelasan yang terkait dalam penyelesaian penelitian ini.

BAB 3: ANALISA DAN RANCANGAN SISTEM

Bab ini menjelaskan analisis dan rancangan desain sistem yang akan dibangun pada penelitian ini.

BAB 4: IMPLEMENTASI

Bab ini menjelaskan implementasi yang ada dari Bab 3 sehingga menjadi sebuah model yang dapat menganalisis sentimen pada setiap teks percakapan pada forum diskusi.

BAB 5 : PENGUJIAN

Bab ini berisi hasil pengujian terhadap sistem yang telah dirancang dan dibangun sesuai dengan tujuan dari penelitian.

BAB 6: KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari seluruh proses perancangan dan pengujian aplikasi yang dibuat untuk pengembangan aplikasi lebih lanjut.

BAB 2 KAJIAN LITERATUR

2.1 Text Mining

Text mining merupakan analisis teks dimana data pada umumnya diperoleh dari dokumen, dan bertujuan untuk mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan, keterkaitan dan kelas antar dokumen. Text mining melingkupi sebuah proses ekstraksi informasi yang berpola yang berasal dari sejumlah besar sumber data teks, seperti dokumen Word, PDF, kutipan teks, atau bahkan sms. Text mining dibagi menjadi 2 tahap, pertama diawali dengan merubah data teks tidak terstruktur ke data semi atau terstruktur dilanjutkan dengan melakukan ekstraksi informasi yang diteliti dari data teks terstruktur. (Hartanto, 2017)

Dengan dapat melakukan klasifikasi kata yang muncul adalah langkah yang baik untuk *text mining* ini. Sejauh ini software analisis teks hanya dapat digunakan secara optimal dalam bahasa Inggris. Sehingga, terdapat beberapa kendala untuk dapat menganalisis teks menggunakan bahasa indonesia dikarenakan penelitian — penelitian mengenai text mining berbahasa Indonesia lebih sedikit dibandingkan dengan penelitian yang lain.

2.2 Natural Language Toolkit (NLTK)

Natural Language Toolkit (NLTK) diciptakan pada tahun 2001 sebagai bagian dari program komputasi linguistik di Departemen Ilmu Komputer dan Informasi di University of Pennsylvania. Natural Language Toolkit mendefinisikan infrastruktur yang dapat digunakan untuk membangun program pemrosesan bahasa manusia menggunakan "Python". Natural Language Toolkit menyediakan class untuk mewakili data yang relevan dengan pemrosesan bahasa manusia, interface standar untuk melakukan tugas-tugas (penandaan part-of-speech, parsing sintaksis, dan klasifikasi teks) dan implementasi standar untuk setiap tugas yang dapat dikombinasikan untuk memecahkan masalah yang kompleks. (Bird et al., 2019)

2.3 Python

Python didesain oleh Guido van Rossum dan pertama kali dirilis pada tahun 1991. Python adalah bahasa pemrograman yang digunakan pada berbagai platform yang berfokus pada tingkat keterbacaan kode dan merupakan salah satu bahasa populer yang berkaitan dengan *Data Science*, *Machine Learning*, dan *Internet of Things* (IOT). Beberapa keunggulan python diantaranya python memiliki sintaks sederhana yang mirip dengan bahasa inggris, python dapat bekerja pada platform yang berbeda (Windows, Linux, Mac OS, dan lainnya), serta python dapat dijalankan dengan cara prosedur, cara berorientasi objek atau cara fungsional.

Python mendukung multi paradigma pemrograman, utamanya; namun tidak dibatasi; pada pemrograman berorientasi objek, pemrograman imperatif, dan pemrograman fungsional. Python umumnya digunakan sebagai bahasa skrip meski pada praktiknya penggunaan bahasa ini lebih luas mencakup konteks pemanfaatan yang umumnya tidak dilakukan dengan menggunakan bahasa skrip. Python dapat digunakan untuk berbagai keperluan pengembangan perangkat lunak dan dapat berjalan di berbagai platform sistem operasi.

2.3.1 Google Colab

Google menyediakan fasilitas pemrograman online yang dapat diakses di melalui website sebagai Integrated Environment Development (IDE) merujuk Jupyter Notebook dengan ekstensi *.ipynb. Struktur IDE Google Collab yaitu jendela kode berupa kumpulan sel (cell), jendela file yang secara default tertutup untuk keperluan unggah dan unduh file pendukung program, dan menu.

Untuk memulai Google Collab, masuk ke menu dan pilih "new notebook". Karakter berupa sel pada Google Collab berfungsi untuk memisahkan antara satu grup kode dengan grup lainnya. Keunggulan IDE ini adalah fasilitas yang disediakan oleh Google untuk menggunakan GPU dan TPU milik mesin pencari ini yang terkenal tangguh. Sementara jika menggunakan Jupyter Notebook harus menyediakan kartu grafis GPU sendiri sebagai accelerator. Untuk memanfaatkan GPU/TPU pengguna dapat

masuk ke menu edit lalu memilih GPU atau TPU setelah masuk ke notebook settings. (Handayanto & Herlawati, 2020)

2.4 UML (Unified Modelling Language)

Suatu alat untuk memvisualisasikan dan mendokumentasikan hasil analisa dan desain yang berisi sintak dalam memodelkan sistem secara visual. Juga merupakan satu kumpulan konvensi pemodelan yang digunakan atau menggambarkan sebuah sistem software yang terkait dengan objek. Analisis dan desain sistem yang ekspresif dari suatu bahasa pemodelan visual ini dilakukan, agar dapat mengembangkan dan melakukan pertukaran data yang bermakna dan mendukung konsep pembangunan tingkat tinggi seperti kolaborasi kerangka pola dan komponen terhadap suatu sistem. (Haviluddin, 2011)

2.4.1 *Use Case* Diagram

Diagram yang bersifat statis, yang memperlihatkan himpunan Use Case dan aktor-aktor dan menggambarkan apa saja aktivitas yang dilakukan oleh suatu sistem dari sudut pandang pengamatan luar. Diagram ini sangat penting untuk mengorganisasi dan memodelkan perilaku suatu sistem yang dibutuhkan serta diharapkan pengguna yang menjadi persoalan itu "apa yang dilakukan" bukan "bagaimana melakukannya". Diagram ini menggambarkan fungsionalitas yang diharapkan dari sistem. (Y & E, 2015)

Syarat penamaan pada use case adalah nama didefinisikan sesimpel mungkin dan dapat dipahami. Ada dua hal utama pada *use case* yaitu pendefinisian apa yang disebut aktor dan *use case*.

- Aktor merupakan orang, proses, atau sistem lain yang berinteraksi dengan sistem informasi yang akan dibuat di luar sistem informasi yang akan dibuat itu sendiri, jadi walaupun simbol dari aktor adalah gambar orang, tetapi aktor belum tentu merupakan orang.
- 2. Use case merupakan fungsionalitas yang disediakan sistem.

2.4.2 Activity Diagram

Diagram aktivitas yang menggambarkan workflow (aliran kerja) atau aktivitas dari sebuah sistem atau proses bisnis. Diagram aktivitas menggambarkan aktivitas sistem yang dapat dilakukan oleh sistem bukan oleh aktor. (Setiady & Yulistia, 2013)

Dalam pembuatannya yang menjadi penting adalah penamaan aktivitas dilakukan dengan menggunakan kata kerja, kemudian terdapat beberapa simbol yang digunakan, sebagai berikut.

NO	GAMBAR	NAMA	KETERANGAN
1		Activity	Memperlihatkan bagaimana masing- masing kelas antarmuka saling berinteraksi satu sama lain
2		Action	State dari sistem yang mencerminkan eksekusi dari suatu aksi
3	•	Initial Node	Bagaimana objek dibentuk atau diawali.
4	•	Activity Final Node	Bagaimana objek dibentuk dan diakhiri
5	\Diamond	Decision	Diguanakan untuk menggambarkan suatu keputusan / tindakan yang harus diambil pada kondisi tertentu
6	↓↑	Line Connector	Digunakan untuk menghubungkan satu simbol dengan simbol lainnya

Gambar 2.1 Komponen Activity Diagram

2.5 ERD (Entity Relationship Diagram)

Komponen-komponen entitas dan himpunan relasi yang masing-masing dilengkapi dengan atribut-atribut yang merepresentasikan seluruh fakta yang ditinjau sehingga dapat diketahui hubungan antara entity-entity yang ada dengan atribut-atributnya. Selain itu juga bisa menggambarkan hubungan yang ada dalam pengolahan data, seperti hubungan many to many, one to many, one to one . Berikut simbol-simbol dari ERD:

Notasi	Keterangan
	Entitas, adalah suatu objek unik yang dapat diidentifikasi dalam lingkungan pemakai.
\Diamond	Relasi, menunjukan adanya hubungan diantara sejumlah entitas yang berbeda.
	Atribut, adalah property dari entitas atau tipe relasi.
	Garis, sebagai penghubung antara relasi dengan entitas, relasi dan entitas dengan atribut.

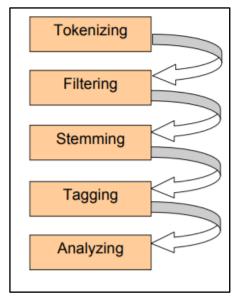
Gambar 2.2 Komponen ERD

BAB 3

ANALISIS DAN DESAIN

3.1 Analisis

Secara garis besar penelitian yang direncanakan terbagi ke dalam lima tahapan besar. Pada Gambar 3.1 memberikan ilustrasi tahapan penelitian tersebut



Gambar 3.1 Tahapan Penelitian

3.1.2 Tahap Pertama

Tahap pertama adalah *Tokenizing* untuk memisahkan per-kata dari sebuah kalimat. *Tokenizing* merupakan proses penguraian deskripsi yang semula berupa kalimat menjadi kata. Pada memberikan ilustrasi tahapan *Tokenizing*.

Text Input	Hasil Token
Go until jurong point, crazy Available only in bugis n great world la e buffet Cine there got amore wat	go until jurong point crazy available wat

Gambar 3.2 Ilustrasi Tokenizing

3.1.3 Tahap Kedua

Tahap kedua adalah *Filtering* untuk mengambil kata penting dr *tokenizing*. *Filtering* adalah tahap mengambil kata penting dari hasil proses token. Bisa menggunakan algoritma *stoplist* atau *word list*, *Filtering* dapat juga diartikan sebagai proses mengambil kata – kata penting dari hasil proses token atau penghapusan *stopwords*. *Stopwords* merupakan kosakata yang bukan merupakan ciri (kata unik) dari suatu dokumen. Pada Gambar 3.3 memberikan ilustrasi tahapan *filtering*.

Hasil Token	Hasil Filtering
go	until
until	jurong
jurong	point
point	crazy
crazy	available
available	bugis
in	
bugis	
wat	wat

Gambar 3.3 Ilustrasi Filteing

3.1.4 Tahap Ketiga

Tahap ketiga adalah *Stemming* untuk *filter* kata pokok dari *filtering*, kata tambahan me-an contohnya. *Stemming* merupakan tahap untuk mencari root kata dari hasil *filtering*. *Stemming* adalah proses pemetaan dan penguraian berbagai bentuk (*variants*) dari suatu kata menjadi bentuk kata dasarnya (*stem*). Pada Gambar 3.4 memberikan ilustrasi tahapan *Stemming*.

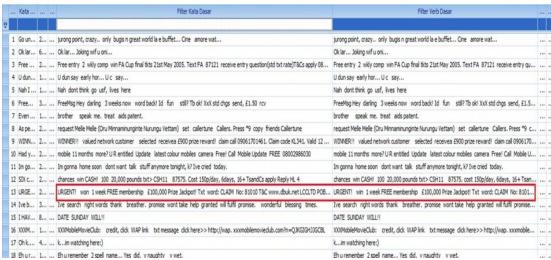


Gambar 3.4 Ilustrasi Stemming

Pada kalimat dengan ID 5 terdapat kalimat hasil proses filtering yaitu "Nah don't think goes usf, lives here", setelah adanya proses stemming kata "goes" berubah menjadi "go".

3.1.5 Tahap Keempat

Tahap keempat adalah *Tagging* untuk mencari bentuk awal dari *stemming*. *Tagging* merupakan tahap untuk mencari bentuk awal atau *root* dari tiap kata lampau atau hasil dari proses *stemming*.



Gambar 3.5 Ilustrasi Tagging

3.1.6 Tahap Kelima

Tahap kelima adalah *Analyzing* untuk tahap keterhubungan dengan dokumen atau kata kunci yang ada. *Analyzing* merupakan tahap penentuan seberapa jauh keterhubungan antara suatu kata atau term terhadap suatu dokumen atau kalimat dengan menghitung nilai atau bobot keterhubungan.

(Ratniasih et al., 2017)

3.2 Data Penelitian

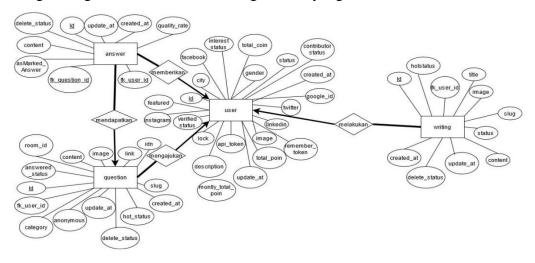
Untuk tahap awal penelitian dilakukan adanya penghimpunan data dari *philoit.* Data telah berhasil terhimpun, data tersebut diantaranya data *user*, data *question*, data *answer*, dan data *writing*

3.3 Rancangan Basis Data

Dengan melakukan rancangan sistem ini dapat membantu untuk memberikan gambaran fungsionalitas sistem yang akan dibuat. Perancangan sistem akan dijelaskan menggunakan *Entity Relationship Diagram*.

3.3.1 Entity Relationship Diagram

Entity Relationship Diagram menggambarkan pemodelan basis data yang menghubungkan antara satu tabel dengan tabel yang lain.



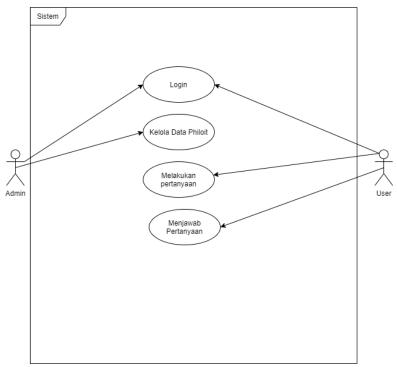
Gambar 3.6 ERD

3.4 Unified Modelling Language

Unified modelling language (UML) merupakan proses pemodelan yang menggambarkan keseluruhan proses atau data yang mendukung aplikasi, pemodelan yang dilakukan menggunakan Use Case dan Activity Diagram.

3.4.1 *Use Case*

Use case diagram menggambarkan hubungan antara aktor dengan sistem. Use case diagram dapat mendeskripsikan sebuah interaksi atau keterkaitan antara satu atau lebih aktor dengan sistem yang akan dibuat.



Gambar 3.7 Use Case

Berdasarkan Gambar Use Case Diagram yang telah dibuat, berikut skenario yang menjelaskan use case diagram tersebut

Tabel 3.1 Skenario Login

Tuber out pachatro Login		
Nama	Login	
Aktor	Admin, User	
Deskripsi	Aktor melakukan login supaya dapat masuk kedalam sistem	
Pre-Condition	Aktor masuk ke menu login dan belum melakukan login	
Post-Condition	Aktor telah melakukan login untuk dapat menggunakan sistem	

Tabel 3.2 Skenario Kelola Data Philoit

Nama	Kelola Data Philoit
Aktor	Admin
Deskripsi	Dapat memperbaharui data <i>philoit</i> yang berhasil di saring melalui text mining
Pre-Condition	Data lama belum diperbaharui
Post-Condition	Data telah diperbaharui dan tersimpan dalam sistem

Tabel 3.3 Skenario Melakukan pertanyaan

Nama	Melakukan pertanyaan
Aktor	User
Deskripsi	Dapat mengajukan pertanyaan pada philoit
Pre-Condition	Mengisi kolom pertanyaan dan mengajukan pertanyaan tersebut
Post-Condition	Pertanyaan telah diajukan dalam sistem

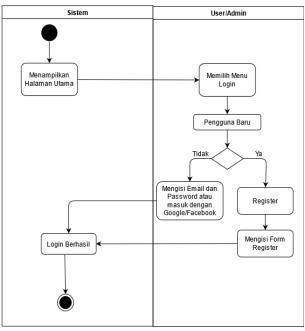
Tabel 3.4 Skenario Menjawab pertanyaan

Nama	Menjawab pertanyaan
Aktor	User
Deskripsi	Dapat menjawab pertanyaan pada philoit
Pre-Condition	Mengisi kolom jawaban dan mengajukan jawaban tersebut
Post-Condition	Jawaban telah diajukan dalam sistem

3.4.3 Activity Diagram

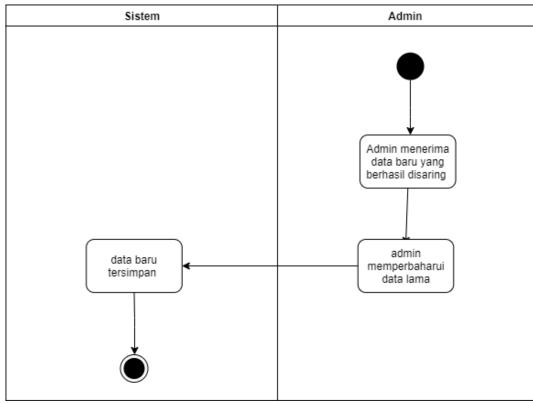
Activity Diagram menggambarkan sebuah aliran kerja atau aliran aktivitas dari beberapa use case yang mendukung dalam perancangan sistem.

1. Login



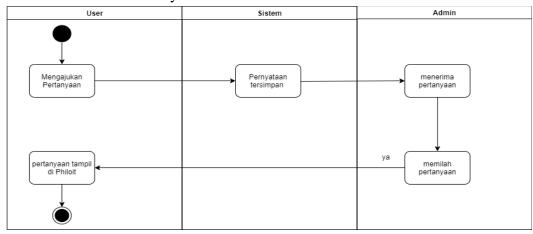
Gambar 3.8 Activity Diagram Login

2. Kelola Data Philoit



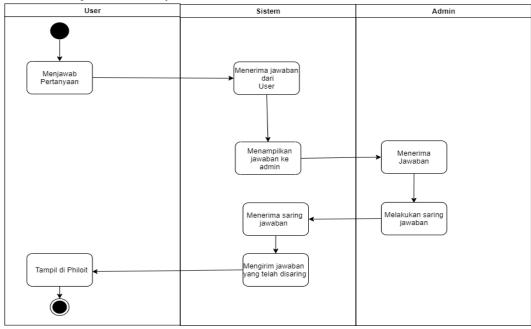
Gambar 3.9 Activity Diagram Kelola Data Philoit

3. Melakukan Pertanyaan



Gambar 3.10 Activity Diagram Melakukan Pertanyaan

4. Menjawab Pertanyaan



Gambar 3.11 Activity Diagram Menjawab Pertanyaan

Daftar Pustaka

Bird, S., Klein, E., & Loper, E. (2019). *Natural Language Processing with Python* (3.0 ed.). O'Reilly.

Handayanto, R. T., & Herlawati. (2020, Mei). Machine Learning Berbasis

Desktop dan Web dengan Metode Jaringan Syaraf Tiruan Untuk Sistem

Pendukung Keputusan. *Jurnal Komtika*, 4, 17-18.

Hartanto. (2017). TEXT MINING DAN SENTIMEN ANALISIS TWITTER
PADA GERAKAN LGBT. TEXT MINING DAN SENTIMEN ANALISIS
TWITTER PADA GERAKAN LGBT, 8.

Haviluddin. (2011, Februari 1). Memahami Penggunaan UML (Unified Modelling Language). *Jurnal Informatika Mulawarman*, 6, 1-2.

Ratniasih, N. L., Sudarma, M., & Gunantara, N. (2017, September - Desember). *Penerapan Text Mining dalam Spam Filtering untuk Aplikasi Chat*, 16.

Setiady, H., & Yulistia. (2013). Sistem Informasi Pemesan dan Penjualan Berbasis Web Pada Dewi Florist. 4.

Y, P. A., & E, J. (2015). Sistem Pakar Diagnosa Penyakit Ginjal dan Saluran Kemih Dengan Metode Breadth First Search. *Jurnal Informatika*, 2, 217.