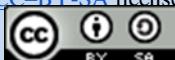


Konversi Bahasa Indonesia ke Perintah *Data Manipulation Language* pada *Structured Query Language* menggunakan *Natural Language Processing*

Sirajuddin^{a,1}, Yulita Salim^{a,2}, Mardiyyah Hasnawi^{a,3,*}

^a Universitas Muslim Indonesia, Jalan Urip Sumoharjo KM.5, Makassar dan 90245, Indonesia
¹ sirajuddin.cloudstring@gmail.com; ² yulita.salim@umi.ac.id; ³ mardiyyah.hasnawi@umi.ac.id
^{*}corresponding author

INFORMASI ARTIKEL	ABSTRAK
Diterima : 16 – 02 – 2022 Direvisi : 01 – 07 – 2022 Diterbitkan : 31 – 08 – 2022	Bahasa merupakan alat yang digunakan manusia untuk berkomunikasi. Agar bahasa alami dapat dimengerti oleh komputer, maka diperlukan penerjemah bahasa yang mampu memahami arti bahasa alami misal Bahasa Indonesia. <i>Structured Query Language (SQL)</i> merupakan salah satu bahasa pemrograman yang digunakan dalam database. <i>Data Manipulation Language (DML)</i> termasuk jenis perintah dasar dalam Bahasa SQL. Perintah tersebut memungkinkan pengguna memanipulasi database antara lain menampilkan, menambah, mengubah dan menghapus data. Namun, tidak semua pemrogram pemula mempelajari tentang DML dapat memahami penggunaan perintah-perintah tersebut dengan cepat karena adanya kendala bahasa. Oleh karena itu, penelitian ini bertujuan untuk membuat aplikasi yang dapat mengonversi bahasa Indonesia menjadi perintah DML menggunakan <i>Natural Language Processing (NLP)</i> . Aplikasi ini dibangun dengan melakukan analisis leksikal (<i>Scanner</i>), analisis sintaksis (<i>Parser</i>), analisis semantik (<i>Translator</i>) dan analisis pragmatik (<i>Evaluator</i>) hingga implementasi dan pengujian aplikasi. Hasil penelitian ini berupa aplikasi yang dapat melakukan konversi Bahasa Indonesia menjadi perintah dasar DML pada SQL menggunakan NLP. Berdasarkan hasil pengujian sebanyak 40 kali uji coba, 29 perintah yang berhasil dikonversi dengan benar dengan tingkat akurasi sebesar 72.5%.
<i>Kata Kunci:</i> Konversi Bahasa Bahasa Indonesia SQL DML NLP	This is an open access article under the CC-BY-SA license 

I. Pendahuluan

Bahasa merupakan alat yang digunakan manusia untuk berkomunikasi. Fungsi bahasa sebagai alat komunikasi yaitu sebagai sarana penyampaian informasi kepada orang lain baik secara lisan maupun tulisan. Sedangkan bahasa alami atau *natural language* adalah bahasa yang dapat dimengerti dan dipahami oleh seseorang pada lokasi tertentu, sebagai contoh bahasa alami dari orang Indonesia adalah bahasa Indonesia. Tidak hanya untuk melakukan komunikasi antara manusia dengan manusia yang lainnya, bahasa juga dapat menjadi perantara komunikasi manusia dengan komputer. Bahasa alami tidak akan langsung dimengerti oleh komputer. Agar bahasa alami dapat dimengerti oleh komputer, maka harus diolah pada setiap tahapan *Natural Language Processing (NLP)*. NLP atau pengolahan bahasa alami merupakan salah satu bidang ilmu *Artificial Intelligence* (Kecerdasan Buatan) yang mempelajari komunikasi antara manusia dengan komputer melalui bahasa alami [1].

Bahasa yang digunakan manusia untuk berkomunikasi dengan komputer dikenal dengan bahasa pemrograman. Bahasa *Structured Query Language (SQL)* merupakan salah satu bahasa pemrograman yang digunakan dalam *database*. *Database* dapat diungkapkan sebagai suatu pengorganisasian data dengan bantuan komputer yang memungkinkan data dapat diakses dengan mudah [2]. Bahasa SQL terbagi dalam dua bagian yaitu *Data Definition Language (DDL)* dan *Data Manipulation Language (DML)*. DDL adalah perintah yang digunakan untuk mendefinisikan skema ke dalam *database* sedangkan DML adalah perintah yang digunakan untuk memanipulasi *database* misalnya menghapus, mengubah dan menambah [3]. Pada DML terdapat beberapa perintah seperti *Select*, *Insert*, *Update* dan *Delete*. Untuk memanipulasi database pengguna dapat menggunakan perintah-perintah tersebut. Namun, tidak semua pemula dalam dunia pemrograman yang baru mempelajari tentang DML dapat memahami penggunaan perintah-perintah tersebut dengan cepat karena adanya kendala bahasa. Oleh karena itu, dengan menggunakan bahasa alami khususnya bahasa Indonesia untuk mengeksekusi query, pengguna akan lebih mudah memahami perintah-perintah tersebut dengan cepat. Untuk

itu, maka dibangun sebuah aplikasi yang dapat mengonversi bahasa Indonesia menjadi perintah DML menggunakan NLP.

II. Metode

A. Structured Query Language(SQL)

SQL merupakan standar bahasa basis data relasional karena hampir semua DBMS telah mendukung penggunaan dari SQL. SQL terbagi menjadi 2 bagian, yaitu *Data Definition Language* (DDL) dan *Data Manipulation Language* (DML) [4]. DDL bertujuan untuk mendefinisikan struktur atau skema basis data, termasuk record, elemen data, kunci elemen dan relasinya. DDL mengacu pada kumpulan perintah yang dapat digunakan untuk mendefinisikan objek-objek basis data, seperti membuat sebuah tabel basis data atau indeks primer atau sekunder sedangkan DML adalah perintah yang memperbolehkan pemakai untuk mengakses atau memanipulasi data. DML mengacu pada kumpulan perintah yang dapat digunakan untuk melakukan manipulasi data, seperti penyimpanan data ke suatu tabel, kemudian mengubahnya dan menghapusnya atau hanya sekedar menampilkannya kembali [5]. Beberapa perintah dasar dalam DML antara lain *select* berfungsi untuk menampilkan data yang ada dalam tabel pada *database*, *insert* berfungsi untuk menambahkan data ke dalam tabel pada *database*, *update* berfungsi untuk mengubah data yang ada dalam tabel pada *database* dan *delete* berfungsi untuk menghapus data yang ada dalam tabel pada *database* [5].

B. Natural Language Processing (NLP)

Natural Language Processing (NLP) dapat didefinisikan sebagai kemampuan suatu komputer untuk memproses bahasa, baik lisan maupun tulisan yang digunakan oleh manusia dalam percakapan sehari-hari. Untuk proses komputasi, bahasa harus direpresentasikan sebagai rangkaian simbol yang memenuhi aturan tertentu. Secara sederhana, NLP mencoba untuk membuat komputer dapat mengerti perintah-perintah yang ditulis dalam standar bahasa manusia [4]. Dalam NLP terdapat beberapa tahapan yaitu Analisis Leksikal (Scanner), Analisis Sintaksis (Parser), Analisis Semantik (Translator) dan Analisis Pragmatik (Evaluator) [6].

1) Analisis Leksikal (Scanner)

Pada tahap analisis leksikal atau pre-processing terdapat 4 (empat) aktifitas yang akan dilakukan yaitu

- a) *Case Folding*. Proses mengubah semua huruf dalam dokumen menjadi huruf kecil, agar sama untuk setiap kata yang ada [7].
- b) *Tokenizing*. Proses untuk mengubah kalimat, paragraf, dokumen menjadi teks atau token-token tertentu dimana spasi dan tanda baca digunakan sebagai acuan pemisah antar token [8]
- c) *Stopword*. Tahap pemilihan kata-kata penting dari hasil token yang akan mewakili dokumen. Stop-words removal ini membuang kata-kata yang tidak penting [8].
- d) *Stemming*. Proses mencari kata dasar dari sebuah kata imbuhan, dengan menghilangkan imbuhan-imbuhan pada kata [8].

2) Analisis Sintaksis (Parser)

Proses menganalisis daftar token yang terbentuk dari proses leksikal yang akan dianalisis oleh *parser* dan *lexicon*. Analisis sintaksis melakukan pelacakan terhadap token-token yang dihasilkan oleh scanner kemudian dibandingkan dengan daftar token yang tersedia [9].

3) Analisis Semantik (Translator)

Untuk melakukan analisis semantik, terdapat istilah *understander* dan *knowledge base*. *Understander* bekerja sama dengan *knowledge base*. *Understander* akan menentukan makna sebuah kalimat dengan melihat kata-kata atau frasa-frasa yang terdapat dalam *knowledge base* bermakna [4]. *Knowledge base* merupakan basis pengetahuan yang dijadikan acuan oleh *understander* dalam menentukan makna suatu kalimat. *Knowledge base* berisi sekumpulan informasi yang terorganisasi dengan baik, yang secara unik diformat ke dalam bentuk struktur data [4]

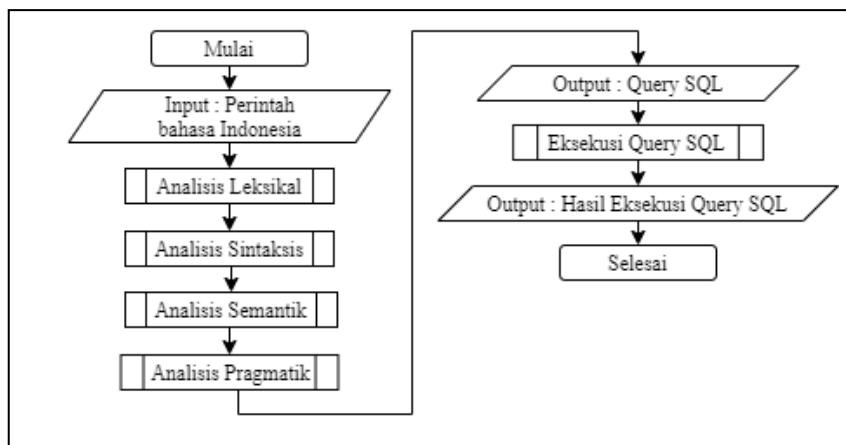
4) Analisis Pragmatik (Evaluator)

Analisis pragmatik atau *evaluator* berfungsi untuk menentukan jawaban akhir *query* berdasarkan hasil keluaran *translator* (analisis semantik) yang berupa perintah *query* menurut tipe dan atribut sesuai dengan notasi masing-masing tipe *query* [10].

III. Hasil dan Pembahasan

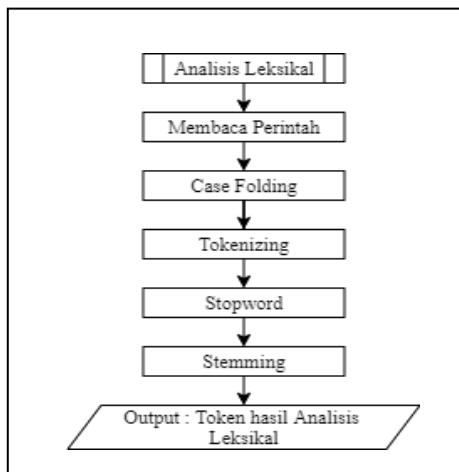
A. Algoritma Konversi Bahasa Indonesia

Konversi Bahasa Indonesia ke perintah DML menggunakan konsep *natural language processing*. Tahapan konversi dimulai dari memasukkan kata atau kalimat Bahasa Indonesia kemudian dianalisis kalimatnya dengan pendekatan leksikal, sintaksis, semantik dan pragmatik. Luaran analisis tersebut berupa perintah DML pada SQL. Setelah kalimat diterjemahkan ke dalam perintah DML maka selanjutnya dieksekusi hingga menghasilkan luaran query. Algoritma Konversi Bahasa Indonesia dengan pendekatan NLP sebagai berikut:

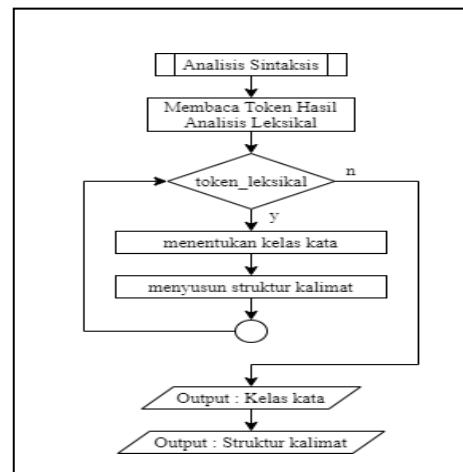


Gambar 1. Algoritma Konversi Bahasa Indonesia menggunakan NLP

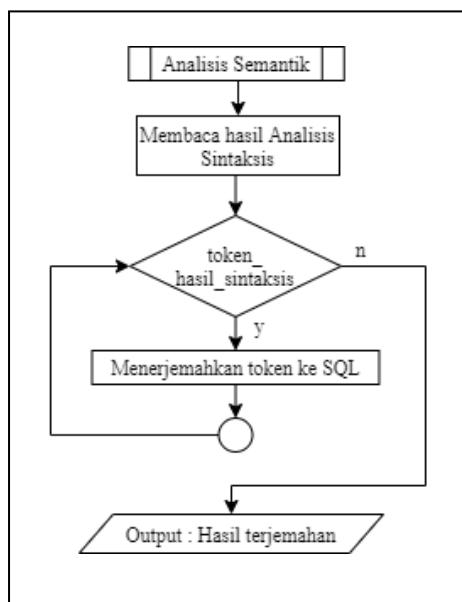
Flowchart analisis leksikal, analisis sintaksis, analisis semantik dan analisis pragmatik secara berurutan dapat dilihat pada gambar berikut ini:



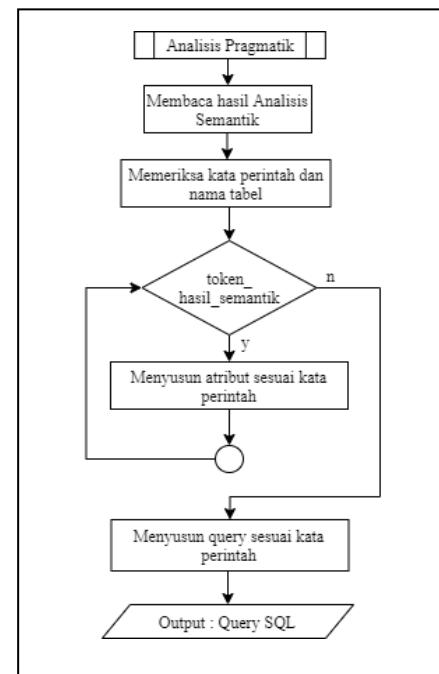
Gambar 2. Flowchart proses analisis leksikal



Gambar 3. Flowchart proses analisis sintaksis



Gambar 4. Flowchart proses analisis semantik



Gambar 5. Flowchart proses analisis pragmatik

B. Perancangan Knowledge Base

Knowledge base berupa kata atau kalimat yang disimpan dalam format file XML. Berikut ini daftar kata *stopword* yang digunakan pada proses analisis leksikal.

Tabel 1. Daftar *stopword*

Kepala Tabel
adalah
adapun
agak
agaknya
akan
...
<i>dan seterusnya</i>

Proses *parsing* atau penentuan kelas kata dan struktur kalimat pada proses analisis sintaksis memerlukan daftar kelas kata yang tidak termasuk kata *stopword* seperti pada tabel berikut:

Tabel 2. Daftar kelas kata

Kata Perintah	Kata Kuantitas	Kata Sambung	Operator
tampil	semua	jika	atau
tunjuk	jumlah	kalau	dan
cari	banyak	dengan	tidak
berapa	total	yang	bukan
hapus		mana	kecuali
ubah			selain
edit			sama
masuk			kurang
tambah			lebih
			!=
			<>
			=
			<
			<=
			>
			>=

Untuk melakukan analisis semantik, *translator* memerlukan *understander* yang berisi daftar token dan sintak SQL yang disimpan dalam *knowledge base*. Isi dari *understander* dapat dilihat pada tabel berikut:

Tabel 3. Daftar sintak SQL

Token	Sintak SQL
tampil	<i>select</i>
tunjuk	<i>select</i>
cari	<i>select</i>
berapa	<i>select</i>
tambah	<i>insert</i>
masuk	<i>insert</i>
ubah	<i>update</i>
edit	<i>update</i>
hapus	<i>delete</i>

Token	Sintak SQL
jika	<i>where</i>
mana	<i>where</i>
dan	<i>and</i>
atau	<i>or</i>
tidak	<i>!=</i>
bukan	<i>!=</i>
kecuali	<i>!=</i>
selain	<i>!=</i>
sama	<i>=</i>
tidak sama dengan	<i>!=</i>

Token	Sintak SQL	Token	Sintak SQL
kurang dari	<	\geq	\leq
kurang atau_sama dengan	\leq	>	>
lebih dari	>	\neq	\neq
lebih atau sama dengan	\geq	jumlah	$count(*)$
kurang lebih	\approx	banyak	$count(*)$
\neq	\neq	total	$count(*)$
\leq	\leq	semua	$count(*)$
<	<		

C. Pemrosesan Konversi Bahasa ke DML

Sebelum sistem menampilkan hasil konversi dari perintah bahasa Indonesia yang dimasukkan *user*, terdapat beberapa proses yang terjadi. Pemrosesan Bahasa melalui beberapa tahap yang dimulai dari proses analisis leksikal sampai proses analisis pragmatik. Perintah diproses dengan menerapkan metode *natural language processing* sampai menghasilkan *query DML*. Perintah DML antara lain *insert*, *update*, *delete*.

1) Perintah Masukan Bahasa

Misalkan pengguna memasukkan perintah berikut ini:

Perintah *Insert*: Tambahkan data “1304BB802”, “Tugas Akhir”, “4”, “8” ke dalam tabel mata_kuliah

2) Analisis Leksikal

- *Perintah Insert*

Perintah: Tambahkan data “1304BB802”, “Tugas Akhir”, “4”, “8” ke dalam tabel mata_kuliah

Case Folding: tambahkan data “1304BB802”, “Tugas Akhir”, “4”, “8” ke dalam tabel mata_kuliah

Tokenizing: tambahkan | data | “1304BB802” | “Tugas Akhir” | “4” / “8” | ke | dalam | tabel | mata_kuliah

Stopword: tambahkan | “1304BB802” | “Tugas Akhir” | “4” | “8” | mata_kuliah

Stemming: tambah | “1304BB802” | “Tugas Akhir” | “4” | “8” | mata_kuliah

3) Analisis Sintaksis

- Perintah Insert

Perintah: Tambahkan data “1304BB802”, “Tugas Akhir”, “4”, “8” ke dalam tabel mata_kuliah

Kelas Kata: kata perintah -> tambah

data -> “1304BB802” “Tugas Akhir” “4” “8”

nama tabel -> mata_kuliah

Struktur Kalimat: kata_perintah, data, nama_tabel

4) Analisis Semantik

- *Perintah Insert*

Perintah: Tambahkan data “1304BB802”, “Tugas Akhir”, “4”, “8” ke dalam tabel mata_kuliah

Terjemahan: tambah -> insert

data -> “1304BB802” “Tugas Akhir” “4” “8”

nama tabel -> mata_kuliah

5) Analisis Pragmatik

- *Perintah Insert*

Perintah: Tambahkan data “1304BB802”, “Tugas Akhir”, “4”, “8” ke dalam tabel mata_kuliah

Query SQL: insert into mata_kuliah values ('1304BB802', 'Tugas Akhir', '4', '8')

D. Implementasi Konversi Bahasa Indonesia

Desain konversi Bahasa Indonesia ke perintah DML diimplementasi ke dalam program sehingga dapat mengonversi kalimat perintah DML sesuai sintak Query SQL. Beberapa perintah yang dapat diproses adalah perintah menampilkan, menambah, mengubah dan menambah data. Tampilan aplikasi konversi Bahasa ke perintah DML SQL pada Gambar 6.

Convert Indonesian to DML

NLP

Analisis Leksikal

Analisis Sintaksis

Analisis Semantik

Analisis Pragmatik

Database

Info Database

tampilkan data pada tabel mata_kuliah

Hasil Konversi

```
select * from mata_kuliah
```

Jalankan SQL Bersihkan

kode_mata_kuliah	nama_mata_kuliah	sls	tingkat_semester
1301KB202	Praktikum Basis Data	1	2
1301KB206	Praaktikum Microprocessor & M... 1		3
1301KB308	Praaktikum Pemrograman Java F... 1		3
1301KB310	Praktikum Basis Data 2	1	3
1301KB413	Praktikum Jaringan Komputer 1		4
1301KB415	Praktikum Pemrograman Java L... 1		4
1301KB417	Praktikum Pemrograman Web 1		4
1301KB521	Praaktikum Sistem Terdistribusi 1		5
1301KB528	Praaktikum Pemrograman Enter... 1		5

Gambar 6. Form Hasil Konversi Bahasa ke DML

Uji coba konversi perintah *insert* atau memasukkan data dapat dilihat pada Gambar 7.

Tambahkan data "1304BB802", "Tugas Akhir", "4", "8" ke dalam tabel mata_kuliah

Hasil Konversi

```
insert into mata_kuliah values ('1304BB802', 'Tugas Akhir', '4', '8')
```

Jalankan SQL Bersihkan

Gambar 7. Hasil konversi perintah *select*

Knowledge Base berupa daftar kata untuk verifikasi Bahasa. Implementasi *form knowledge base* pada Gambar 8.

Knowledge Base

Kategori

- select
- insert
- update
- delete

Kosakata

- tampil = select
- cari = select
- berapa = select
- tambah = insert
- masuk = insert
- ubah = update
- edit = update

Kosakata

+

Simpan Batal

Gambar 8. Form *Knowledge Base*

E. Pengujian Aplikasi

Pengujian konversi Bahasa dilakukan dengan teknik pengujian *White Box* dan *Black Box*. Berdasarkan pengujian *White Box* pada proses *stopword*, *stemming*, analisis sintaksis, dan analisis semantik masing-masing memiliki nilai *cyclomatic complexity* dan *independent path* yang sama dan menunjukkan bahwa kode program dari proses tersebut sudah sesuai dan benar dalam alur logika. Pada proses *case folding*, *tokenizing* dan analisis pragmatik ditemukan bahwa masing-masing nilai *cyclomatic complexity* dan *independent path* terdapat perbedaan. Sedangkan hasil pengujian *Black Box* sebanyak 40 percobaan, maka ditemukan 29 perintah yang berhasil dikonversi.

Tabel 4. Perfomansi Sistem

No	Perintah	Total Uji	Keberhasilan Perintah	Persentase
1	Menampilkan (<i>select</i>)	10	7	70%
2	Menambahkan (<i>insert</i>)	10	6	60%
3	Mengubah (<i>update</i>)	10	8	80%
4	Menghapus (<i>delete</i>)	10	8	80%
Akurasi Keseluruhan		40	29	72%

IV.Kesimpulan

Berdasarkan penelitian ini melalui proses perancangan, implementasi dan pengujian, maka diperoleh kesimpulan bahwa Aplikasi yang dibuat mampu memproses perintah menampilkan (*select*), menambah (*insert*), mengubah (*update*) dan menghapus data (*delete*). Aplikasi yang dibuat juga mampu mengonversi perintah bahasa Indonesia menjadi perintah DML. Berdasarkan hasil uji coba konversi perintah, didapatkan akurasi sebesar 72.5%.

Daftar Pustaka

- [1] J. Karlos, H. Sujaini and H. Anra, "Konversi Bahasa Indonesia ke SQL (Structured Query Language) dengan Pendekatan Mesin Penerjemah Statistik," *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, vol. 3, no. 1, pp. 1-6, 2016.
- [2] Y. Priyadi, Kolaborasi SQL & ERD dalam Implementasi Database, Yogyakarta: Andi Offset Yogyakarta, 2014.
- [3] Rachmawati, "Analisis Kesalahan Menerapkan Bahasa SQL (Structure Query Language) Mata Kuliah Basis Data," *Jurnal Pendidikan dan Riset Matematika Prismatika*, vol. 1, no. 2, pp. 27-34, 2019.
- [4] E. A. Lisangan, "Natural Language Processing Dalam Memperoleh Informasi Akademik Mahasiswa Universitas Atma Jaya Makassar," *TEMATIKA*, vol. 1, no. 1, pp. 34-46, 2013.
- [5] N. K. D. A. Jayanti and N. K. Sumiari, Teori Basis Data, Yogyakarta: Penerbit Andi, 2018.
- [6] D. Soyusiawaty and A. H. S. Jones, "Penerapan Antar Muka Bahasa Alami Dalam Pencarian Informasi Skripsi Pada Suatu Program Studi," *SEMNASTEKNOMEDIA ONLINE*, vol. 6, no. 1, pp. 13-18, 2018.
- [7] R. Rosyadi, S. Al-Faraby and Adiwijaya, "Penerapan Question Answering System Pada Pembahasan Agama Islam Dengan Pendekatan Metode Pattern Based," *MEDIA INFORMATIKA BUDIDARMA*, vol. 2, no. 4, pp. 140-149, 2018.
- [8] E. Supriyati and M. Iqbal, "Pengukuran Similarity Tema Pada Juz 30 Al Qur'an Menggunakan Teks Klasifikasi," *SIMETRIS*, vol. 9, no. 1, pp. 261-370, 2018.
- [9] S. Raharjo and S. Hartati, "Antarmuka Bahasa Alami Untuk Melakukan Query Terhadap Terjemahan Al-Quran," *Jurnal Teknologi*, vol. 7, no. 1, pp. 12-19, 2014.
- [10] S. Wibisono, "Aplikasi Pengolah Bahasa Alami untuk Query Basisdata Akademik dengan Format Data Xml," *Jurnal Teknologi Informasi DINAMIK*, vol. 18, no. 1, pp. 65-79, 2013.