

Alat Bantu Anotasi Pengenalan Entitas Manusia AI Qur'an Fasilitas Saran Secara Otomatis

Abu Abdirrohman Rosikhun Nur Muttaqin¹, Moch Arif Bijaksana², Donni Richasdy³

^{1;2;3}Fakultas Informatika, Universitas Telkom, Bandung

¹abuabdirrohman@students.telkomuniversity.ac.id, ²arifbijaksana@telkomuniversity.ac.id,

³donnir@telkomuniversity.ac.id

Abstrak

Setiap agama tentu memiliki kitabnya masing-masing sebagai panduan para penganutnya. Al-Quran merupakan kitab suci umat Islam dan merupakan mukjizat yang diturunkan kepada Nabi Muhammad SAW. Al-Qur'an juga adalah kitab suci umat Islam yang wajib dipelajari oleh seluruh umat Islam yang ada di dunia. Dikarenakan Al-Qur'an memakai bahasa Arab sehingga masih banyak orang Indonesia yang ke-seulitan memahaminya. Dibutuhkan suatu alat yang bisa membantu memudahkan pemahaman pada isi Al-Qur'an dan untuk membuat itu diperlukan terlebih dahulu alat yang bisa membantu mengumpulkan dataset yang nantinya bisa digunakan untuk membangun berbagai alat atau teknologi canggih yang membantu pemahaman isi Al-Qur'an. Dataset dapat dibuat dengan melakukan pemberian informasi pada sebuah teks ataupun sebuah kalimat atau bisa disebut anotasi. Pembuatan dataset dapat dilakukan secara manual ataupun otomatis. Namun, pembuatan dataset secara otomatis lebih mudah dan membutuhkan waktu yang lebih singkat daripada pembuatan secara manual. Penelitian ini bertujuan untuk membuat alat bantu untuk melakukan proses anotasi pada teks Al-Qur'an. Alat bantu ini dibangun dengan menerapkan metode rule based yang merupakan bagian dari NER (Named Entity Recognition) untuk mendeteksi entitas dari suatu teks dan digunakan sebagai fitur saran dalam alat bantu ini. Hasil penelitian menunjukkan bahwa alat bantu ini berhasil membantu proses anotasi data teks pada Al-Qur'an. Alat bantu ini dibuat menggunakan teknologi ReactJS, Flask dan Mongo. Penerapan rule based pada alat bantu ini berhasil menghasilkan dataset yang bisa digunakan sebagai fitur saran yang ada di dalam alat bantu.

Kata Kunci: al-quran, entitas, entitas manusia, alat bantu, reactjs, flask, mongodb, name entity recognition, rule based

Abstract

Each religion certainly has its own book as a guide for its adherents. Al-Quran is the holy book of Muslims and is a miracle revealed to the Prophet Muhammad SAW. Al-Qur'an is also a holy book of Muslims that must be studied by all Muslims in the world. Because the Qur'an uses Arabic, there are still many Indonesians who find it difficult to understand. A tool is needed that can help facilitate understanding of the contents of the Qur'an and to make it necessary beforehand a tool that can help collect data sets that can later be used to build various tools or sophisticated technology that helps understand the contents of the Qur'an. The dataset can be created by giving information to a text or a sentence or it can be called an annotation. Creating a dataset can be done manually or automatically. However, creating a dataset automatically is easier and requires a shorter time than creating it manually. This study aims to make a tool for annotating the Qur'anic text. This tool was built by implementing a method based rule that is part of the NER (Named Entity Recognition) to detect entities from a text and is used as a suggestion feature in this tool. The results showed that this auxiliary tool was successful in helping the text data annotation process in the Qur'an. This tool is made using react, flask and mongo technology. The application of rule based on this tool succeeds in producing a dataset that can be used as a suggestion feature contained in the tool.

Keywords: al-quran, entitys, human entity, annotation tool, reactjs, flask, mongodb, name entity recognition, rule based



1. Pendahuluan

Latar Belakang

Al-Qur'an merupakan dokumen terpenting bagi umat Islam. Dokumen yang keberadaannya akan terus terjaga. Sayangnya pada saat ini masih banyak sekali orang-orang yang beragama Islam yang punya kewajiban untuk membaca, memahami dan mempraktikkan segala perintah dan larangan yang ada di dalam Al Qur'an masih belum bisa melakukannya. Oleh karena itu, penulis berharap bisa membantu kebanyakan orang-orang yang beragama Islam untuk bisa lebih mudah dan bisa memahami dan mempraktikkan isi Al-Qur'an. Dikarenakan Al-Qur'an memiliki banyak teks yang berupa surat dan ayat di dalamnya, alangkah lebih baiknya jika proses pembelajaran Al-Qur'an dibantu dengan alat bantu yang dibangun dengan teknologi terkini.

Saat ini sudah ada banyak alat bantu yang memudahkan untuk orang-orang Islam yang tidak bisa berbahasa Arab untuk memahami Al Qur'an, seperti Al Qur'an terjemahan. Namun untuk pemrosesan teks Al-Qur'an sendiri diperlukan notasi yang selengkap mungkin. Sebagai dokumen yang ditujukan untuk petunjuk hidup manusia, salah satu notasi penting adalah notasi entitas manusia. Untuk kemudahan komputasi, basis data notasi entitas manusia perlu dimiliki dan disimpan untuk memudahkan pemrosesan oleh komputer. Bahasa Arab adalah bahasa yang digunakan dalam Al-Qur'an dan bahasa Arab menimbulkan banyak tantangan dalam pemrosesannya, karena secara morfologis bahasa Arab merupakan bahasa yang rumit [2]. Bahasa Arab juga masih kurang memiliki sumber daya dan alat untuk pekerjaan anotasi [11].

Saat ini sudah banyak alat bantu untuk memudahkan orang dalam melakukan proses anotasi pada teks, namun sejauh yang peneliti ketahui belum ada alat bantu anotasi yang khusus dibuat untuk keperluan anotasi pada Al-Qur'an. Oleh karena itu di tugas akhir ini peneliti melakukan pembuatan alat bantu yang digunakan untuk pembangunan basis data pada Al-Qur'an. Selain itu alat bantu ini juga dibangun dengan menerapkan metode rule based yang merupakan bagian dari NER (Named Entity Recognition) untuk mendeteksi entitas dari suatu teks dan digunakan sebagai fitur saran dalam alat bantu ini.

Topik dan Batasannya

Berdasarkan dari latar belakang yang telah dijelaskan, topik-topik yang di angkat dalam tugas akhir ini sebagai berikut:

1. Alat Bantu

Hasil dari tugas akhir ini berupa alat bantu atau aplikasi yang nantinya akan membantu dalam melakukan anotasi entitas pada Al Qur'an.

2. Fasilitas Saran

Fasilitas tambahan dari alat bantu yang akan dibuat pada tugas akhir ini adalah fasilitas saran dalam proses pengerjaan anotasi entitas manusia yang ada di dalam Al-Qur'an dengan menggunakan metode rule based untuk mendapatkan data sarannya.

Adapun batasan dari permasalahan yang ada pada tugas akhir ini sebagai berikut:

1. Penelitian ini hanya berfokus pada penelitian menemukan entitas manusia saja.
2. Pembuatan alat bantu berfokus pada fungsionalitas tampilannya.

Tujuan

Berikut adalah tujuan yang ingin dicapai pada penelitian ini:

1. Membuat alat bantu untuk melakukan anotasi pada Al Qur'an.
2. Mengimplementasikan metode rule based dalam mencari entitas manusia yang ada pada Al-Qur'an,
3. Menambahkan fitur saran anotasi entitas manusia secara otomatis

Organisasi Tulisan

Struktur penulisan tugas akhir ini tersusun atas pendahuluan yang menjelaskan permasalahan pada aktifitas anotasi pada al qur'an, dilanjutkan dengan bagian studi terkait yang memaparkan studi literatur terkait yang men-cakup istilah entitas, entitas manusia atau person dan alat bantu yang bisa digunakan untuk membantu dalam kegiatan anotasi pada Al-Qur'an. Bagian selanjutnya merupakan bagian penjelasan terkait sistem dan produk atau alat bantu yang dibangun, membahas tahapan yang dilakukan untuk menyelesaikan tugas akhir dengan penjelasan rinci dan singkat yang diawali dari pengumpulan data, pembuatan alat bantu, fitur saran hingga kesimpulan. Lalu hasil pengujian dan apa saja yang perlu ditambahkan untuk penelitian selanjutnya. Bagian terakhir adalah daftar pustaka yang merupakan daftar dari setiap referensi yang digunakan dalam penyelesaian tugas akhir.

2. Studi Terkait

2.1 Ringkasan Studi Terkait

Pada tugas akhir ini studi-studi yang terkait dengan isi dari tugas akhir ini adalah alat bantu anotasi entitas, pengenalan entitas manusia pada Al-Qur'an dengan named entity recognition, pendekatan dengan metode rule based, pengukuran sistem dengan confusion matrix untuk mencari nilai precision, recall, f-measure serta pengujian penerimaan pengguna atau user acceptance test.

Alat Bantu Anotasi Entitas adalah sebuah alat bantu atau aplikasi yang nantinya akan membantu pengguna untuk melakukan anotasi, yang sebelumnya harus dilakukan secara manual dengan alat bantu yang dibuat ini para anotator akan dibantu untuk melakukan anotasi secara lebih cepat dan mudah karena sudah dibantu dengan alat bantu ini [4]. Alat bantu ini dibangun dengan teknologi ReactJS untuk frontend, Flask untuk backend dan MongoDB untuk database.

Pengenalan entitas manusia pada Al Qur'an disini adalah pengenalan untuk mengetahui mana saja yang ter-masuk entitas manusia yang ada di dalam Al Qur'an [8]. NER digunakan untuk melakukan pengenalan ini dan metode yang digunakan adalah metode rule based.

Confusion matrix alat pengukuran yang digunakan untuk menghitung kinerja sistem dan digunakan untuk men-cari nilai precision, recall dan f-measure sebagai evaluasi sistemnya. Sedangkan user acceptance test digunakan untuk mengevaluasi performa alat bantu atau produk yang dibuat.

2.2 Named Entity Recognition

NER (Named Entity Recognition) adalah proses mengekstraksi otomatis entitas bernama yang dianggap penting di dalam sebuah teks dan menentukan kategorinya ke dalam kategori terdefinisi. Sebagai contoh untuk teks berita, NER dapat mengekstraksi nama orang, nama organisasi dan nama lokasi[14]. Tugas utama NER adalah untuk mencari named entity dan menentukan tipe named entity. Cara dasar untuk mengenali named entity adalah dengan mencari jenis dari setiap kata pada teks menggunakan kamus[1].

Entitas dapat berupa banyak hal, seperti: orang, benda, tempat, kejadian ataupun konsep. Untuk entitas manusia itu seperti: mahasiswa, dosen, penjual dan pembeli. Sedangkan untuk contoh entitas-entitas orang yang ada di dalam Al Qur'an itu seperti: orang-orang iman, manusia, orang-orang kafir, orang yang bersyukur ataupun orang yang bertobat. Nama-nama tokoh atau orang pada Al-Qur'an juga termasuk kedalam kategori ini, seperti: Yusuf, Ibrahim, Fir'aun, Zulkarnain dan Musa.

2.3 Confusion Matrix

Confusion Matrix adalah Alat pengukuran yang dapat digunakan untuk menghitung kinerja atau tingkat kebenaran dari proses klasifikasi [7]. Perbandingan dilakukan anatara hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi sebenarnya. Terdapat 4 istilah sebagai representasi hasil proses klasifikasi pada confusion matrix yaitu TP (True Positive) yang merupakan data positif yang diprediksi benar, TN (True Negative) data negatif yang diprediksi benar, FP (False Positive) data positif namun diprediksi sebagai data negatif dan FN (False Negative) data positif namun diprediksi sebagai data negatif.

2.4 Rule Based

Rule based adalah sebuah metode dimana aturan yang ada di dalam sistem dibuat sendiri berdasarkan pengetahuan linguistik. Analisis dilakukan pada tingkatan sintaksis dan semantik secara lebih mendalam merupakan kelebihan dari metode ini[1].

Grishman pada tahun 1995 mengembangkan rule based dengan memanfaatkan kamus data yang terdiri dari nama negara, kota, perusahaan dan beberapa nama-nama sejenis[5]. Dengan menggunakan pendekatan rule based pengenalan entitas dilakukan dengan mendefinisikan aturan mengenai pola-pola posisi kata anggota entitas pada sebuah frase atau kalimat. Kendala implementasi dari metode ini berada pada kemampuan definisi pola yang biasanya dilakukan oleh ahli bahasa. Rule based juga memiliki ketergantungan yang besar dengan bahasa yang digunakan.

2.4.1 Precision

Precision adalah persentasi entitas manusia yang ditemukan terhadap keseluruhan hasil entitas yang dianggap sebagai entitas manusia oleh mesin [9]. Rumus precision adalah sebagai berikut:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (1)$$

TP (True Positive) yaitu kondisi dimana anotasi yang didapat oleh sistem merupakan entitas manusia.
FN (False Negatif) kondisi dimana anotasi yang didapat oleh sistem bukan termasuk entitas manusia.

2.4.2 Recall

Recall adalah persentase entitas manusia yang ditemukan terhadap jumlah entitas manusia sebenarnya [9].
Rumus recall adalah sebagai berikut:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2)$$

FP (False Positif) yaitu kondisi dimana anotasi yang didapat oleh sistem hanya sebagiannya saja yang merupakan entitas manusia sedangkan sebagian yang lain bukan merupakan entitas manusia.

2.4.3 F-measure

F-measure merupakan rata-rata dari precision dan recall [9]. Rumus f-measure terdiri dari perhitungan hasil precision dan recall adalah sebagai berikut:

$$\text{F-measure} = \frac{2PR}{P + R} \quad (3)$$

2.4.4 User Acceptance Test

User Acceptance Test (UAT) atau uji penerimaan pengguna adalah suatu proses pengujian oleh pengguna yang dimaksudkan untuk menghasilkan dokumen yang dijadikan bukti bahwa software yang telah dikembangkan dapat diterima oleh pengguna, bahwa hasil pengujian (testing) sudah bisa dianggap memenuhi kebutuhan dari pengguna [13].

3. Sistem yang Dibangun

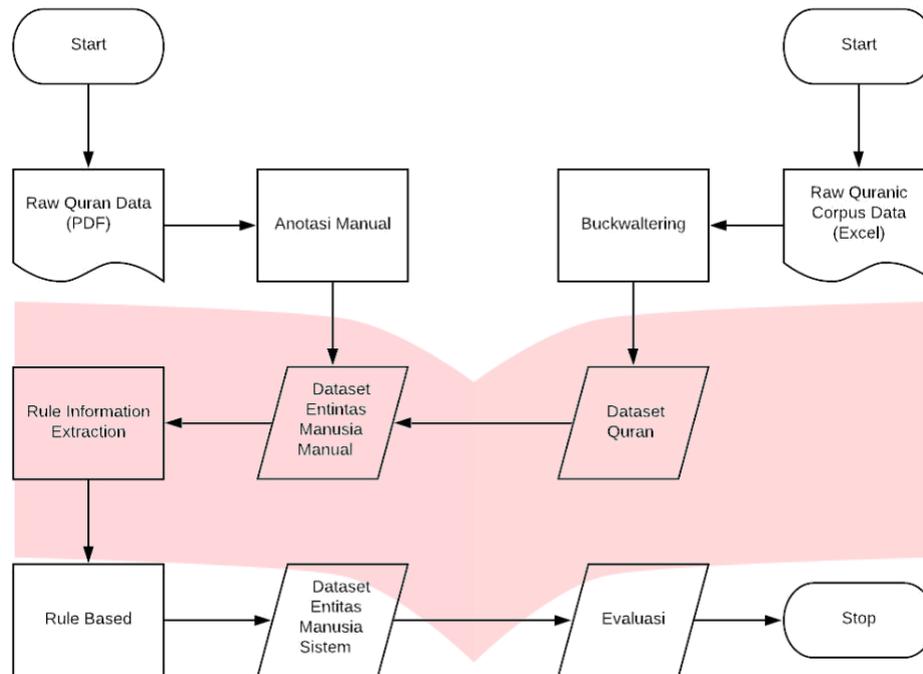
Sistem yang dibangun bertujuan untuk menghasilkan saran atau rekomendasi yang akan diberikan pada saat penganotasian entitas pada produk alat bantu yang dibuat. Pada gambar 1 digambarkan gambaran umum dari sistem dan gambar 3 digambarkan gambaran umum dari alat bantu yang akan dibangun dalam tugas akhir ini.

Alat bantu anotasi teks Al-Qur'an ini dibangun pada aplikasi berbasis web dengan teknologi ReactJS, Flask dan MongoDB. Sistemnya menyediakan seluruh data surat yang ada di dalam Al-Qur'an yang mampu melakukan anotasi serta memberikan saran teks yang termasuk entitas manusia yang ada di dalam Al-Qur'an. Saran yang diberikan ini merupakan hasil proses dari pengolahan dataset yang ada dengan menggunakan metode rule based.

Berikutnya akan dijelaskan arsitektur sistem, dataset, rule untuk entitas manusia, implementasi rule based, arsitektur alat bantu dan fitur-fitur yang dibuatnya.

3.1 Arsitektur Sistem

Hasil rekomendasi entitas manusia yang dimunculkan pada alat bantu yang dibuat, seperti yang terlihat pada gambar 5 berasal dari penelitian mencari entitas bernama melalui proses named entity recognition menggunakan metode rule based. Pada gambar 1 ditunjukkan gambaran umum dari sistem yang dibuat. Proses ini dimulai dengan pengolahan 2 dataset, lalu didapatkanlah rule yang digunakan untuk mencari entitas manusia secara otomatis dan digunakan sebagai saran pada alat bantu yang dibuat.



Gambar 1. Gambaran Umum Sistem

Penelitian ini diawali dengan 2 buah dataset mentah berbeda yakni raw quranin data yang berbentuk pdf dan raw quranic corpus data yang berbentuk excel, lalu tahapan berikutnya adalah sebagai berikut:

1. Buckwaltering

Buckwaltering adalah proses translasi data morfem yang masih dalam bentuk bahasa latin menjadi bahasa arab. Tahapan buckwaltering ini dilakukan pada dataset raw quranic corpus data yang berbentuk excel, tahapan ini dilakukan agar dataset yang awalnya hanya berisi data morfem berbentuk latin selanjutnya akan berisi juga data morfem dalam bentuk arab. Data morfem dalam bentuk arab ini digunakan sebagai tampilan data al quran yang akan ditampilkan pada alat bantu.

Hasil proses buckwaltering ini akan menghasilkan dataset qur'an yang akan menjadi dataset utama dalam sistem dan produk yang akan dibangun. Pada sistem akan digunakan sebagai dataset untuk menerapkan metode rule based dan pada produk akan digunakan sebagai database untuk menampilkan data Al-Qur'an pada alat bantu.

2. Anotasi Manual

Anotasi manual dilakukan pada dataset raw quranic data dalam bentuk pdf, ini adalah proses mendapatkan entitas manusia yang ada pada Al-Qur'an dengan tanpa bantuan alat bantu. Pekerjaan anotasi ini dilakukan dengan cara memberikan tanda kurung berwarna merah diantara kata atau kalimat yang merupakan entitas manusia.

Hasil anotasi manual yang berisi daftar kata atau kalimat yang merupakan entitas manusia ini lalu dimasukkan secara manual ke dalam dataset qur'an berbentuk excel yang merupakan hasil buckwaltering sebelumnya dan menjadi dataset entitas manusia secara manual.

3. Rule Information Extraction

Dataset entitas manusia secara manual sebelumnya lalu dimasukkan pada program yang dibuat oleh peneliti dengan pemrograman python untuk mengekstraksi pattern data entitas manusia yang ada pada dataset. Ekstraksi ini menghasilkan pattern kelas sintaksi dari data entitas manusia yang ada pada dataset dan dijadikan sebagai rule dalam pemrosesan rule based untuk mendapatkan entitas manusia secara otomatis.

4. Rule Based

Dengan menggunakan rule yang didapatkan pada proses sebelumnya, rule ini lalu digunakan sebagai dataset untuk mendapatkan dataset entitas manusia secara otomatis oleh sistem dan hasilnya akhirnya dievaluasi dengan confusion matrix.

3.2 Dataset

Untuk membuat sistem ini dilakukan pengumpulan data terlebih dahulu berupa dataset. Ada 2 dataset utama yang diolah, berikut adalah dataset tersebut:

1. Al-Qur'an PDF

Dataset Al-Qur'an dalam bentuk pdf ini merupakan pekerjaan paling awal dari penelitian ini, yaitu melakukan anotasi untuk entitas manusia secara manual dengan menandai kata-kata atau kumpulan kata yang termasuk ke dalam entitas manusia dan hasilnya akan digunakan sebagai acuan untuk anotasi manual pada dataset Al-Qur'an dalam bentuk excel. Contoh pekerjaan anotasi manual pada Al-Qur'an pdf ini dapat dilihat pada gambar 2 atau data yang lebih lengkapnya terdapat pada lampiran 5.

Anotasi dilakukan pada kata dan kalimat yang berisi nama manusia atau yang menyatakan bahwa kata atau kalimat itu berarti manusia. Anotasi diawali oleh peneliti yang lalu hasilnya pun akan dikonfirmasi pada pakar atau ahli yang paham dengan grammar bahasa arab atau nahwu shorof yang ada pada Al-Qur'an.



Gambar 2. Contoh Anotasi Manual Pada PDF

2. Al-Qur'an Excel

Dataset Al-Qur'an berbentuk excel ini berisi data seluruh morfem, posisi tiap morfem dan kategori sintaksis yang ada di dalam Al-Qur'an, didapat dari dari corpus.quran.com [3]. Data morfem yang berbentuk tulisan latin pada dataset ini lalu di diubah menjadi tulisan Arab dengan menggunakan buckwalter yaitu sebuah transliterasi bahasa Arab yang cukup populer dan dikembangkan oleh Tim Buckwalter [12], hal ini dilakukan untuk keperluan tampilan teks Al-Qur'an yang ada pada alat bantu. Lalu ditambahkan data buka kurung dan tutup kurung pada data morfem yang termasuk dari bagian entitas manusia berdasarkan data entitas manusia yang ada pada dataset Al-Qur'an dalam bentuk pdf sebelumnya agar bisa dilakukan proses pencarian rule entitas manusia. Contoh dataset ini dapat dilihat pada tabel 1 atau yang lebih lengkapnya terdapat pada lampiran 5.

OPEN TAG	LOCATION	FORM	TAG	ARAB	CLOSE TAG
	(2:160:1:1)	jil aA	EXP	B@	
(((2:160:2:1)	(l a*iyna	REL	ÆKY,@)
	(2:160:3:1)	laAbu	v	HAK	
	(2:160:3:2)	wA@	PRON	@s)

Tabel 1. Contoh Anotasi Manual Pada Excel

3.3 Rule Entitas Manusia

Dari dataset yang telah dikumpulkan sebelumnya sekarang kita mempunyai dataset yang setiap data morfem-nya memiliki kategori sintaksisnya masing-masing yang ditampilkan pada kolom TAG. Data morfem yang telah ditandai sebagai entitas manusia akan membentuk kumpulan tag yang menjadi pola atau pattern entitas manusia. Artinya, setiap terdapat kumpulan entitas manusia maka akan terdapat pula kumpulan pattern tag yang akan menjadi rule untuk menemukan entitas manusia lainnya. Untuk mendapatkan pattern tersebut, kita perlu melakukan crawl pada dataset Al-Qur'an yang berbentuk excel. Peneliti melakukan crawl pada dataset Al-Qur'an di juz 2 dan hasil dari pekerjaan ini dapat dilihat pada tabel 2. Terdapat 48 pattern yang berhasil didapat dan pattern ini akan digunakan sebagai rule untuk mencari entitas manusia lainnya dengan algoritma Rule Based.

No	Pattern
1	[COND, V, P, PRON]
2	[REL, V, P, PRON]
3	[REL, V, PRON, CONJ, REL, V, PRON]
4	[COND, V, DET, N, CONJ, V]
...
...
48	[PN, CONJ, N, PN]

Tabel 2. Tabel Hasil Pattern Yang Didapat

3.4 Rule Based

Setelah mendapatkan pattern maka langkah selanjutnya adalah melakukan pencarian entitas manusia dengan menggunakan rule based, dimana kita akan menggunakan pattern ini sebagai rule untuk memeriksa kecocokan data yang ada pada dataset dengan tiap aturan (rule 1 sampai 48). Hal ini dilakukan dengan menggunakan teknik perulangan atau looping pada sekumpulan morfem yang sesuai dengan jumlah pattern yang sedang difokuskan, perulangan tersebut dimulai dari data yang paling awal sampai data paling akhir. Pencocokannya cukup sederhana, ketika sekumpulan morfem tersebut memiliki pattern TAG yang sama dengan pattern yang sedang difokuskan maka pada kolom OPEN TAG pada morfem awal dari sekumpulan morfem tersebut ditambahkan tanda kurung buka dan pada kolom data CLOSE TAG pada morfem akhir diberi tanda kurung tutup. Data kurung buka dan kurung tutup tersebut menjadi tanda awal dan akhir dari entitas manusia yang ditemukan dengan menggunakan rule based.

Pada penelitian ini data yang akan yang dijadikan data uji adalah juz 2, hal ini dilakukan agar hasil dari pencarian entitas manusia ini bisa dijadikan sebagai baseline untuk menghitung performa sistem. Contoh dataset hasil rule based ini dapat dilihat pada tabel 3 atau yang lebih lengkapnya terdapat pada lampiran 5.

OPEN TAG	LOCATION	FORM	TAG	ARAB	CLOSE TAG
	(2:160:1:1)	il aA	EXP	B@)
(((2:160:2:1)	(l a*iyna	REL	ÆKY,@)
	(2:160:3:1)	taAbu	V	HAK)
	(2:160:3:2)	wA@	PRON	@â)

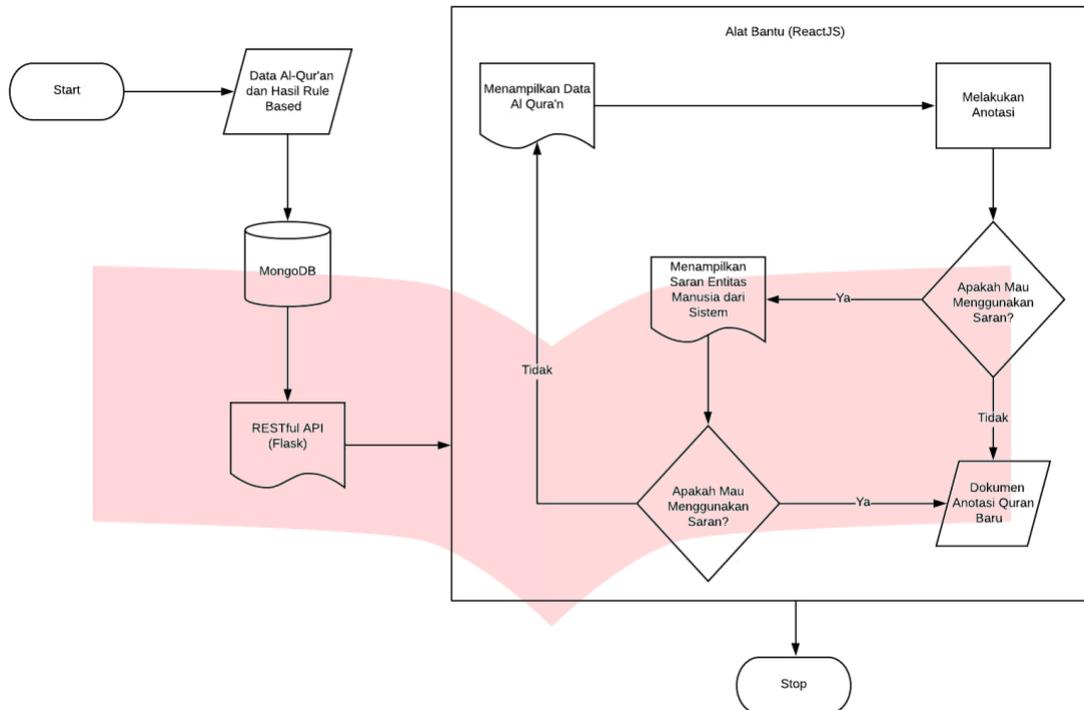
Tabel 3. Contoh Hasil Anotasi Dengan Rule Based

3.5 Arsitektur Alat Bantu

Pertama kita perlu menjelaskan arsitektur umum untuk alat bantu yang dibuat lalu dilanjutkan dengan melihat detail dari sistem dan alat bantu. Pada gambar 3 ditunjukkan gambaran umum dari alur kerja alat bantu saat melakukan proses anotasi.

Alat bantu ini akan menerima dataset seluruh isi teks yang ada di dalam Al-Qur'an dan dataset entitas manusia hasil penganotasian otomatis yang dilakukan dengan rule based, lalu dikirimkan pada tampilan alat bantu sehingga dataset mentah berupa isi teks Al-Qur'an yang masih kosong bisa dikelola dan diproses menjadi sebuah dataset yang baru yang berupa hasil anotasi yang dilakukan oleh pengguna.

Tampilan alat bantu ini menggunakan teknologi ReactJS yang merupakan framework javascript yang sangat populer saat penelitian ini dibuat, memiliki tampilan yang menarik dan performa dari alat bantu yang dibuat menjadi lebih ringan [6], lalu untuk penyimpanan datanya menggunakan MongoDB yang sangat membantu meningkatkan performa pemrosesan data [10] dan pengelolaan database dilakukan dengan teknologi Flask yang merupakan sebuah microframework dari teknologi Python yang sangat populer juga saat penelitian ini dibuat dalam hal pengelolaan dan pemrosesan data, Flask ini akan membuat sebuah RESTful API yang akan menjadi penghubung antara data yang ada pada database yaitu dataset seluruh isi teks yang ada di dalam Al-Qur'an dan dataset entitas manusia hasil penganotasian otomatis yang dilakukan dengan rule based dengan tampilan alat bantu yang dibuat oleh ReactJS.



Gambar 3. Gambaran Umum Alur Kerja Alat Bantu

3.6 Fitur Alat Bantu

Alat bantu yang dibuat ini memiliki 2 fitur utama, yaitu fitur untuk mendapatkan entitas manusia secara oto-matis dan fitur untuk melakukan anotasi pada teks Al-Qur'an yang akan dijelaskan setelah ini, adapun untuk pen-jelasan lebih lengkap terkait fitur-fitur lain pada alat bantu ini seperti fitur membuat, penyimpanan dan perubahan proyek dapat dilihat pada lampiran 5.

3.6.1 Fitur Anotasi

Fitur yang pertama adalah fitur anotasi, dimana pengguna akan dapat melakukan anotasi pada surat yang telah dipilih pada awal pembuatan proyek. Anotasi dilakukan dengan cara memilih dan menarik teks sepanjang teks yang akan dianotasi.

Setelah proses anotasi selesai dilakukan pengguna bisa menyimpan hasil anotasi mereka dan nantinya juga bisa melihat dan membuka kembali pekerjaan anotasi mereka. Karena itu alat bantu ini menyiapkan fungsi save yang tersedia di bagian paling bawah dari halaman pengerjaan anotasi seperti yang terlihat pada gambar 4 dan untuk membukanya kembali bisa melalui halaman recent projects yang ada pada menu kiri halaman.



Gambar 4. Halaman Pengerjaan Anotasi

3.6.2 Fitur Saran

Dalam melakukan proses anotasi penggunaan akan dipermudah dengan fitur saran yang disediakan untuk membantu pengguna dalam melakukan anotasi dengan tujuan mempercepat proses anotasi tanpa harus memakan waktu yang lama. Saran yang disediakan oleh alat bantu ini didapatkan melalui proses pengenalan entitas bernama atau NER (Name Entity Recognition) yang merupakan proses ekstraksi otomatis entitas bernama dengan menggunakan metode rule based. Contoh hasil saran otomatisnya dapat dilihat pada gambar 5.



Gambar 5. Tampilan Fitur Saran

4. Evaluasi

4.1 Hasil Pengujian

Pengujian yang dilakukan adalah dengan menghitung precision, recall dan f-measure dari pengenalan entitas dengan metode rule based untuk sistem dan pengujian penerimaan pengguna atau user acceptance test untuk produk. Pengujian sistem dan produk dilakukan untuk memastikan bahwa sistem dan produk yang dibuat sesuai dengan rancangan yang dibuat dan mengukur performa sistem dan keefektifan produk. Hasil pengujian sistem dapat dilihat pada persamaan 4, 5, 6 dan hasil pengujian produk dapat dilihat pada tabel 6.

4.1.1 Hasil Pengujian Sistem

Pengujian sistem dilakukan dengan membandingkan hasil dari anotasi manual dan anotasi otomatis dengan confusion matrix. Tabel 4 menunjukkan beberapa contoh hasil dari perbandingannya, yang mana kolom MANUAL menandakan hasil dari anotasi manual peneliti dan kolom AUTOMATIC merupakan hasil anotasi dengan rule based.

LOCATION	MANUAL	AUTOMATIC	TP	FP	FN	TN
(2:160:2) - (2:160:4)	@æj* @ð (@æKAK (ÆKY,@))	@æj* (@ð (@æ(KAK ÆKY,@)))	1	2	0	0
(2:160:9:1) - (2:160:11:2)	AK@ð		0	0	0	1
(2:161:1:1) - (2:161:3:2)	(@ðQfi» (ÆKY,@)) à@	(@ð@Qfi» (ÆKY,@)) à@	2	1	0	0

Tabel 4. Contoh Hasil Perbandingan Anotasi Manual dan Prediksi Sistem

Cara penentuan nilai TP, FP, FN dan TN nya adalah jika hasil dari rule based sesuai dengan hasil manual maka dihitung sebagai TP (True Positive), apabila hanya sebagian yang sesuai maka dihitung sebagai TN (True Negative), apabila tidak sesuai sama sekali maka dihitung sebagai FN (True Negative) dan apabila ada hasil manual yang tidak terdeteksi maka dihitung sebagai FN (False Negative). Sesuai atau tidaknya ditentukan oleh pattern dari masing-masing hasil anotasi, apabila pattern satu buah entitas yaitu yang diberi tanda kurung yang dimiliki hasil rule based memiliki pattern yang sama dengan yang dimiliki hasil manual maka akan dikatakan sesuai begitupun sebaliknya. Hasil dari perbandingan ini akan menghasilkan sejumlah nilai TP, TF, FN dan TN yang mana nilai ini akan digunakan untuk perhitungan precision, recall dan f-measure.

Hasil keseluruhan dari perbandingan confusion matrix ini adalah TP sebanyak 63 buah, FP sebanyak 624 buah dan FN sebanyak 0 buah dan TN sebanyak 1727 buah, yang mana ini berasal dari perbandingan 291 entitas

manusia secara manual dengan 2410 entitas manusia secara otomatis dengan sistem rule based yang dilakukan pada dataset yang ada pada juz 2 yang memiliki data morfem sebanyak 4357 buah atau dapat dilihat pada tabel 5 yang menunjukkan hasil lengkap dari confusion matrix.

Nilai Prediksi	Nilai Sebenarnya	
	TP (True Positive) Entitas Terdeteksi Dan Benar 286	FP (False Positive) Entitas Terdeteksi Tapi Tidak Sesuai 278
FN (True Positive) Entitas Benar Yang Tidak Terdeteksi 0	TN (False Positive) Entitas Terdeteksi Tapi Salah 1846	

Tabel 5. Tabel Hasil Confusion Matrix

4.1.2 Hasil Pengujian Produk

Untuk pengujian produk atau alat bantu dilakukan dengan metode user acceptance test yang kesimpulan hasilnya dapat dilihat pada tabel 6 dibawah ini. Kesimpulan hasil pengujian ini didapatkan melalui pengujian yang dilakukan oleh peneliti kepada 5 orang peserta uji yang data peserta ujinya dapat dilihat pada lampiran 7 dengan peserta uji yang dipilih berdasarkan kualifikasi yang ada pada lampiran 5 dan sudah menyatakan ketersediaan dengan surat pernyataan yang dapat dilihat pada lampiran 8. Prosedur yang dilakukan pada pengujian ini dapat dilihat pada lampiran 4 dan dengan skenario pengujian yang dijelaskan pada lampiran 6. Adapun detail hasil pengujian tiap peserta uji dijelaskan pada lampiran 9.

Fungsi Yang Diuji	Hasil Yang Diharapkan	Hasil Aktual	Kesimpulan
Buat Proyek	Akan tampil modal pilih metode dan input data anotator dan masuk ke halaman anotasi teks dengan menampilkan surat yang dipilih.	Dapat membuat proyek berdasarkan metode yang diinginkan dan masuk ke halaman anotasi teks berdasarkan surat yang dipilih.	Diterima
Anotasi Teks	Menambahkan anotasi pada teks yang diinginkan.	Teks berhasil di anotasi seperti yang pengguna inginkan.	Diterima
Tampilkan Saran	Menampilkan saran berdasarkan metode yang dipilih.	Saran berhasil ditampilkan	Diterima
Simpan Proyek	Menyimpan hasil anotasi	Data berhasil tersimpan	Diterima.
Mengubah Proyek	Dapat membuka proyek yang tersimpan dan melanjutkan atau mengubah isi proyek.	Dapat membuka proyek tersimpan yang dipilih dan melakukan pekerjaan anotasi kembali.	Diterima

Tabel 6. Tabel Hasil Pengujian Sistem

4.2 Analisis Hasil Pengujian

4.2.1 Analisis Hasil Pengujian Sistem

Pengujian dilakukan dengan cara membandingkan hasil atau output anotasi otomatis dari sistem dengan metode rule based dengan hasil anotasi yang telah dibuat secara manual. Dengan confusion matrix nilai yang didapat untuk masing-masing variabel pengukuran sebelumnya adalah TP sebanyak 63 buah, FP sebanyak 624 buah dan FN sebanyak 0 buah dan TN sebanyak 1727 buah yang mana ini berasal dari perbandingan 291 entitas manusia secara manual dengan 2410 entitas manusia secara otomatis dengan sistem rule based yang dilakukan pada dataset

yang ada pada juz 2 yang memiliki data morfem sebanyak 4357 buah atau dapat dilihat pada tabel 5 yang menunjukkan hasil lengkap dari confusion matrix. Dengan menggunakan data-data tersebut dan rumus untuk mencari nilai precision, recall dan f-measure yang ada pada persamaan 1, 2, 3 sebelumnya maka hasilnya adalah sebagai berikut:

$$\text{Precision} = \frac{286}{286 + 278} = 50:71\% \quad (4) \qquad \text{Recall} = \frac{286}{286+0} = 100\% \quad (5)$$

$$\text{F-measure} = \frac{2 (50:71\% \cdot 100\%)}{50:71\% + 100\%} = 67:29\% \quad (6)$$

Sistem pengenalan entitas manusia secara otomatis ini berhasil dibangun, akan tetapi memiliki hasil performa precision yang rendah. Faktor utama yang menyebabkan rendahnya nilai precision ini ada pada aspek kompleksitas aturan-aturan entitasnya, aturan yang dipakai pada sistem ini hanya menggunakan pattern dari kategori sintaksis tiap morfem yang ada pada dataset sehingga masih banyak pengenalan yang salah dari sistem. Hal ini dikarenakan pattern yang didapat dari kategori sintaksis masih kurang kompleks dan spesifik.

4.2.2 Analisis Hasil Pengujian Produk

Pengujian dengan user acceptance test ini dilakukan untuk mendapatkan dokumen hasil uji yang akan dijadikan bukti bahwa produk atau alat bantu ini sudah diterima dan sudah memenuhi kebutuhan diminta. Ada 5 fungsi yang diujikan dan masing-masing fungsi memiliki penjelasan mengenai hasil yang diharapkan, hasil aktual, kesimpulan dan komentar dari masing-masing peserta uji mengenai fungsi atau fitur yang telah dibuat. Hasil lengkapnya dapat dilihat pada lampiran 9. Tabel 6 sebelumnya menunjukkan bahwa alat bantu bisa digunakan sebagai bantuan dalam melakukan anotasi pada teks Al-Qur'an dengan hasil penilaian dari peserta uji yang menerima semua fungsi yang diujikan.

Pada pengujian ini para peserta uji juga memberikan komentar pada masing-masing fungsi yang diujikan, hasilnya telah peneliti rangkum dan dapat dilihat pada tabel 7. Hasil rangkuman ini menunjukkan bahwa alat bantu ini perlu diperbaiki dan bisa ditingkatkan lagi dengan menambahkan beberapa masukan yang telah diberikan oleh peserta uji.

Fungsi Yang Diuji	Rangkuman Komentar
Buat Proyek	Sudah sesuai dengan yang diharapkan dan tampilan gambar yang besar untuk membuat proyek mempermudah pengguna baru, namun tulisan penjelasan metode yang dipilih dinilai terlalu kecil
Anotasi Teks	Anotasi dapat dilakukan namun dalam pengerjaan masih sedikit kaku, saat pemilihan teks terkadang ada yang belum tepat dan sebaiknya ditambahkan keterangan pada teks yang dipilih, warna dan angka yang tampil
Tampilkan Saran	Hasilnya sudah sesuai dan membantu pemilihan entitas, tinggal peningkatan pada akurasi dan tambahan penjelasan pada hasil saran anotasi
Simpan Proyek	Tidak ada pemberitahuan bahwa proyek telah disimpan, namun programnya dapat menyimpan hasilnya dengan baik dan ditambahkan keterangan waktu penyimpanan
Mengubah Proyek	Fiturnya berjalan dengan baik, dapat membuka proyek yang disimpan dan dapat mengubah isi proyek

Tabel 7. Tabel Hasil Pengujian Sistem

5. Kesimpulan

Alat bantu untuk anotasi entitas manusia secara otomatis pada teks Al-Qur'an ini berhasil dibuat dan diterima oleh peserta uji, pekerjaan anotasi dapat dilakukan dengan cara membuka alat bantu lalu memilih surat serta teks yang akan dianotasi dan menganotasinya dengan menggunakan tombol annotate yang ada di dalam alat bantu. Metode Rule based berhasil diimplementasikan dengan cara mengambil pattern entitas manusia yang ada pada dataset entitas manusia secara manual dan digunakan sebagai rule dalam pemrosesan sistem yang dibuat sendiri dengan algoritma rule based dan menghasilkan dataset entitas manusia secara otomatis oleh sistem dengan performansi

precision sebesar 50.71% recall sebesar 100% dan f-measure sebesar 67.29%. Hasil dataset secara otomatis oleh sistem ini selanjutnya digunakan sebagai database pada alat bantu untuk menambahkan fitur saran pada alat bantu sehingga alat bantu yang telah dibuat mempunyai fitur saran anotasi entitas manusia secara otomatis yang dapat digunakan dengan cara melakukan check pada checkbox yang ada di samping teks Show Human Entity Suggestion yang ada pada alat bantu sehingga akan ditampilkan teks-teks yang merupakan entitas manusia yang ditandai dengan tanda kurung berwarna hijau diantara teks-teks tersebut.

Karena hanya memakai pattern kategori sintaksisnya saja hasil saran yang dihasilkan masih belum sempurna karena rule yang didapat untuk pemrosesan pengenalan entitas manusianya masih kurang kompleks sehingga masih terdapat banyak kesalahan dalam pengenalan entitasnya. Masih kurangnya kompleksitas dari rule yang dibuat ini menyebabkan hasil saran masih belum maksimal sehingga diperlukan pengetahuan mengenai grammar bahasa Arab atau nahwu shorof agar mengetahui aturan yang lebih akurat dalam penentuan entitas dalam teks Al-Qur'an atau bisa dilakukan analisis lebih mendalam dari dataset yang telah dikumpulkan dengan pendekatan pada variabel yang lain. Source code alat bantu hasil penelitian ini pun dapat digunakan kembali oleh peneliti lain yang ingin memperbaiki maupun meningkatkan hasil yang telah ada, apabila kedepannya ingin dibangun alat bantu yang dapat dijalankan di smartphone bisa dengan mudah melakukannya dengan menggunakan teknologi react native yaitu teknologi yang sangat mirip dengan react dan digunakan untuk membangun aplikasi mobile sehingga pekerjaannya menjadi lebih mudah dan lebih cepat karena dibangun dari teknologi yang memiliki bahasa dasar yang sama yaitu javascript.



Daftar Pustaka

- [1] S. A. Dayinta Warih Wulandari, Putra Pandu Adikara. Named entity recognition (ner) pada dokumen biologi menggunakan rule based dan naive bayes classifier. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2:4555–4563, 2018.
- [2] M. Diab. Second generation amira tools for arabic processing: Fast and robust tokenization, pos tagging, and base phrase chunking. *2nd International Conference on Arabic Language Resources and Tools*, 110, 2009.
- [3] K. Dukes. The arabic quranic corpus. <http://corpus.quran.com/>, 2017. Online; Accessed 28 June 2020.
- [4] A. N. El-ghobashy, G. M. Attiya, and H. M. Kelash. A proposed framework for arabic semantic annotation tool. *International Journal of Computing and Digital Systems*, 2014.
- [5] Grishman. The nyu system for muc-6 or where's the syntax. *proceedings of Sixth Message Understanding Conference (MUC-6)*, 1995.
- [6] E. E. I Gede Bagus Artha Suryawan, Y. Sigit Purnomo WP. Pembangunan aplikasi alat bantu proses anotasi menggunakan progressive web apps. *Jurnal Buana Informatika*, 10, 2019.
- [7] A. H. A. Kabir. Analisis sentimen data kritik dan saran pelatihan aplikasi teknologi informasi (pati) menggunakan algoritma support vector machine. Final project, Universitas Muhammadiyah Malang, 2017.
- [8] R. S. MAK, M. A. Bijaksana, and A. F. Huda. Person entity recognition for the indonesian qur'an translation with the approach hidden markov model-viterbi. *Procedia Computer Science* 157, pages 214–220, 2019.
- [9] G. L. Marina Sokolova. A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, pages 427–437, 2009.
- [10] D. W. Mesri Silalah. Computer Based Information System *Journal PERBANDINGAN PERFORMANSI DA-TABASE MONGODB DAN MYSQL DALAM APLIKASI FILE MULTIMEDIA BERBASIS WEB*, volume 1. 2018.
- [11] D. J. Mona Diab, Kadri Hacioglu. Automatic tagging of arabic text: From raw text to base phrase chunks. *HLT-NAACL 2004: Short Papers*, 2004.
- [12] A. F. H. Teguh Ikhlas Ramadhan, Moch Arif Bijaksana. Rule based pattern type of verb identification algorithm for the holy qur'an. *Procedia Computer Science*, 157, 2019.
- [13] Y. Wahyuningsih. Pengembangan aplikasi test online dengan menggunakan framework codeigniter di smk darul ma'wa plumpang pada mata pelajaran jaringan dasar kelas x. *It-Edu*, 2, 2017.
- [14] M. L. Wibisono, Yudi Khodra. Pengenalan entitas bernama otomatis untuk bahasa indonesia dengan pende-katan pembelajaran mesin. *Seminar Tahunan Linguistik 2018*, 2018.