Pengembangan Sistem Berbasis Komputer untuk Pembangunan Stemming pada Al-Quran Menggunakan Algoritma Shereen Khoja Stemmer

Aditya Hanif Utama[1], Moch. Arif Bijaksana[2], Arief Fatchul Huda[3]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[1]adityahanif@students.telkomuniversity.ac.id, [2]arifbijaksana@telkomuniversity.ac.id, [3]afhuda@gmail.com

Abstrak

Saat ini banyak ahli dalam bidang teknologi informasi telah merancang dan mengembangkan algoritma untuk memecahkan masalah *stemming*, khususnya dalam bahasa arab. Namun dari sekian banyak analisa *stemming* dalam bahasa arab, belum ada standardisasi algoritma *stemming* yang baik dalam menganalisa akurasi teks pada Al-Quran. Pembangunan *stemming* pada Al-Quran merupakan suatu pekerjaan yang penting karena mendukung klasifikasi *sharaf* dalam Al-Quran guna memahami arti dari setiap kata pada Al-Qur‟an. Salah satu *stemmer* atau algoritma *stemming* untuk mencari bentuk dasar dari suatu kata dalam bahasa arab ialah algoritma Khoja Stemmer. Cara kerja dari Khoja Stemmer ialah dengan mencoba untuk mencari akar pada suatu kata dalam bahasa arab dengan menghilangkan awalan terpanjang dan akhiran terpanjang suatu kata, lalu mencoba untuk menentukan akar dari kata yang tersisa menggunakan kamus akar kata. Dalam penelitian kali ini, Khoja Stemmer yang dibangun mampu menghitung rata-rata *stemming* pada Al-Quran sebesar 95,295%. Akan tetapi akar kata yang dihasilkan oleh Khoja Stemmer apabila di periksa secara manual masih terdapat beberapa kesalahan. Dengan demikian, dibutuhkan suatu kamus Al-Quran untuk menganalisa setiap hasil *stemming* yang dilakukan oleh Khoja stemmer dalam melakukan *stemming* pada Al-Quran.

Kata kunci: al-quran, stemming, khoja stemmer, sharaf, bahasa arab

Abstract

Today many experts in the field of information technology have been designing and developing algorithms to solve stemming problems, especially in Arabic. But, from many stemming analysis in Arabic, there is no standardization of good stemming algorithm in analyzing the accuracy of the text in the Quran. The construction of stemming in the Quran is an important work because it supports the classification of *sharaf* in the Quran to understand the meaning of every word in the Quran. One stemmer or stemming algorithm to find the basic form of a word in Arabic is the Khoja Stemmer algorithm. The workings of Khoja Stemmer is to try to search root in a word of Arabic by removing the longest prefix and longest suffix of a word, then trying to determine the root of the remaining word using the root dictionary. In this research, the built of Khoja Stemmer is able to calculate the average stemming in the Quran of 95.295%. However, the roots produced by Khoja Stemmer are still found some errors when manually checked. Thus, it takes a dictionary of the Quran to analyze every result of stemming done by Khoja Stemmer in order to stemming the Quran.

Keywords: quran, stemming, khoja stemmer, sharaf, arabic

1. Pendahuluan

Semakin luas penyebaran ajaran islam didunia ini, maka akan semakin banyak orang-orang yang mempelajari pedoman dari ajaran tersebut, yaitu salah satunya ialah Al-Qur‟an Al-Karim. Al-Qur‟an merupakan kitab suci umat Islam yang merupakan kumpulan firman-firman Allah yang diturunkan kepada Nabi Muhammad Sallahu „Alaihi Wasallam. Tujuan utama diturunkan Al-Qur‟an adalah untuk menjadi pedoman manusia dalam menata kehidupan supaya memperoleh kebahagiaan di dunia dan di akhirat [1].

Karena Al-Qur‟an diturunkan kepada Rasulullah menggunakan bahasa arab (QS. Yusuf [12] : 2), maka dibuatlah Al-Qur‟an dalam berbagai versi terjemahan. Walaupun begitu, banyak orang yang bukan berasal dari arab bahkan ingin mempelajari Al-Qur‟an dari bahasa asal nya guna mengetahui makna yang terkandung didalam bahasa tersebut. Namun untuk orang-orang yang baru mempelajari bahasa arab harus paham arti dari setiap kata yang ada pada Al-Qur‟an dan hal itu membutuhkan ilmu *al-sharf*, yaitu ilmu yang mempelajari seluk-beluk bentuk kata dalam bahasa Arab [2].

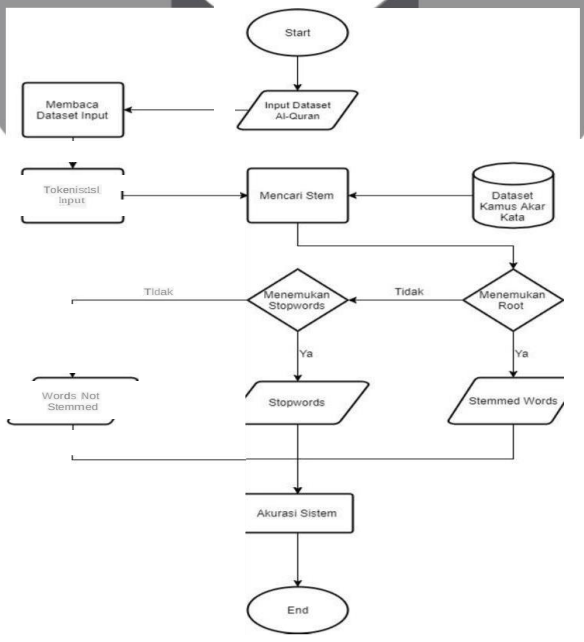
Dengan adanya perkembangan teknologi, dapat dibuat klasifikasi ayat-ayat pada Al-quran dengan *text classification*. *Text classification* adalah sub bidang dari *text mining* yang bertujuan untuk mengklasifikasikan kumpulan teks dari beberapa dokumen ke dalam kategori-kategori tertentu. Sebelum teks tersebut dapat diklasifikasi maka perlu dilakukan *pre-processing* terlebih dahulu, yang meliputi : *tokenization*, *filtering*, *stemming*, dan *weighting* [3]. Dalam hal ini, penulis secara khusus mencoba meneliti proses *stemming* karena *stemming* merupakan langkah penting dalam pemrosesan kata dalam bahasa Arab, terutama untuk Al-Quran. Disamping itu juga dibutuhkan algoritma yang tepat untuk mendapatkan akurasi hasil *stemming* yang baik guna mempermudah orang-orang dalam mempelajari Al-Quran.

Algoritma Khoja Stemmer adalah salah satu *stemmer* bahasa Arab yang terkenal dan banyak digunakan. Cara kerja dari algoritma Khoja Stemmer ialah dengan menghilangkan akhiran terpanjang dan awalan terpanjang dari sebuah kata. Kemudian mencocokkan kata yang tersisa dengan pola verbal dan kata benda (*noun*), untuk mendapatkan akar dari kata [4]. Oleh karena itu, peneliti membangun aplikasi *stemming* pada Al-quran menggunakan algoritma Shereen Khoja Stemmer untuk mengetahui apakah algoritma tersebut cocok digunakan untuk melakukan *stemming* pada Al-Quran sehingga mempermudah pelajar dalam memahami morfologi atau *sharaf* pada Al-Quran.

2. Sistem yang Dibangun

Sistem yang dibangun bertujuan untuk mencari bentuk dasar dari suatu kata pada setiap kata dalam Al-Qur‟an. Input berupa dataset Al-Qur‟an yang terdiri dari 30 juz yang dimuat ke dalam aplikasi. Kemudian aplikasi akan membaca dataset input dan dilakukan proses tokenisasi terhadap setiap baris input. Setelah semua baris input terbentuk menjadi token, maka aplikasi akan melakukan *stemming* pada setiap token dengan menghasilkan sebuah *root* atau akar kata dari setiap token. Suatu *root* dihasilkan dari dataset kamus akar kata yang dihubungkan dengan aplikasi. Setelah itu dihasilkan *output* berupa hasil *stemming* dari setiap kata pada Al- Qur‟an dan akurasi sistem dalam melakukan proses *stemming*.

Berikut ini adalah *flowchart* tentang gambaran umum sistem yang dibangun.

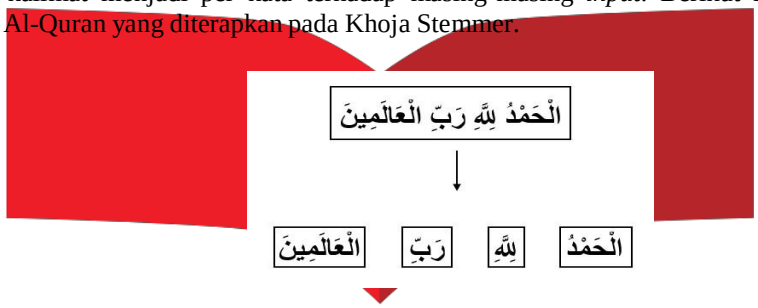


2.1 Membaca Dataset Input

Input adalah dataset Al-Quran berformat .txt yang terdiri dari 30 Juz. Input dilakukan sebanyak 2 kali yang terbagi menjadi 2 bagian, yaitu juz 1-15 pada bagian pertama dan juz 16-30 pada bagian kedua guna meminimalisasi waktu eksekusi. Pada tahap ini aplikasi akan **membaca dan menampilkan hasil input dataset Al-Quran**. Selanjutnya aplikasi akan melakukan proses *stem*. Berikut akan dijabarkan proses *stem* dalam membangun aplikasi *stemming* Al-Quran menggunakan algoritma Khoja Stemmer.

2.2 Tokenisasi

Pada tahap ini input dijabarkan secara perbaris guna dilakukan proses tokenisasi, satu baris terdiri dari satu ayat Al-Quran. Sistem melakukan proses tokenisasi terhadap *input* dataset Al-Quran dengan memecah *input* kalimat menjadi per kata terhadap masing-masing *input*. Berikut ini contoh tokenisasi dalam ayat pada Al-Quran yang diterapkan pada Khoja Stemmer.



2.3 Mencari Stem / Root

Setelah semua input kalimat diubah menjadi per kata dalam tahap sebelumnya, maka sistem akan mencari *root* atau akar dari kata tersebut. Berikut langkah-langkahnya:

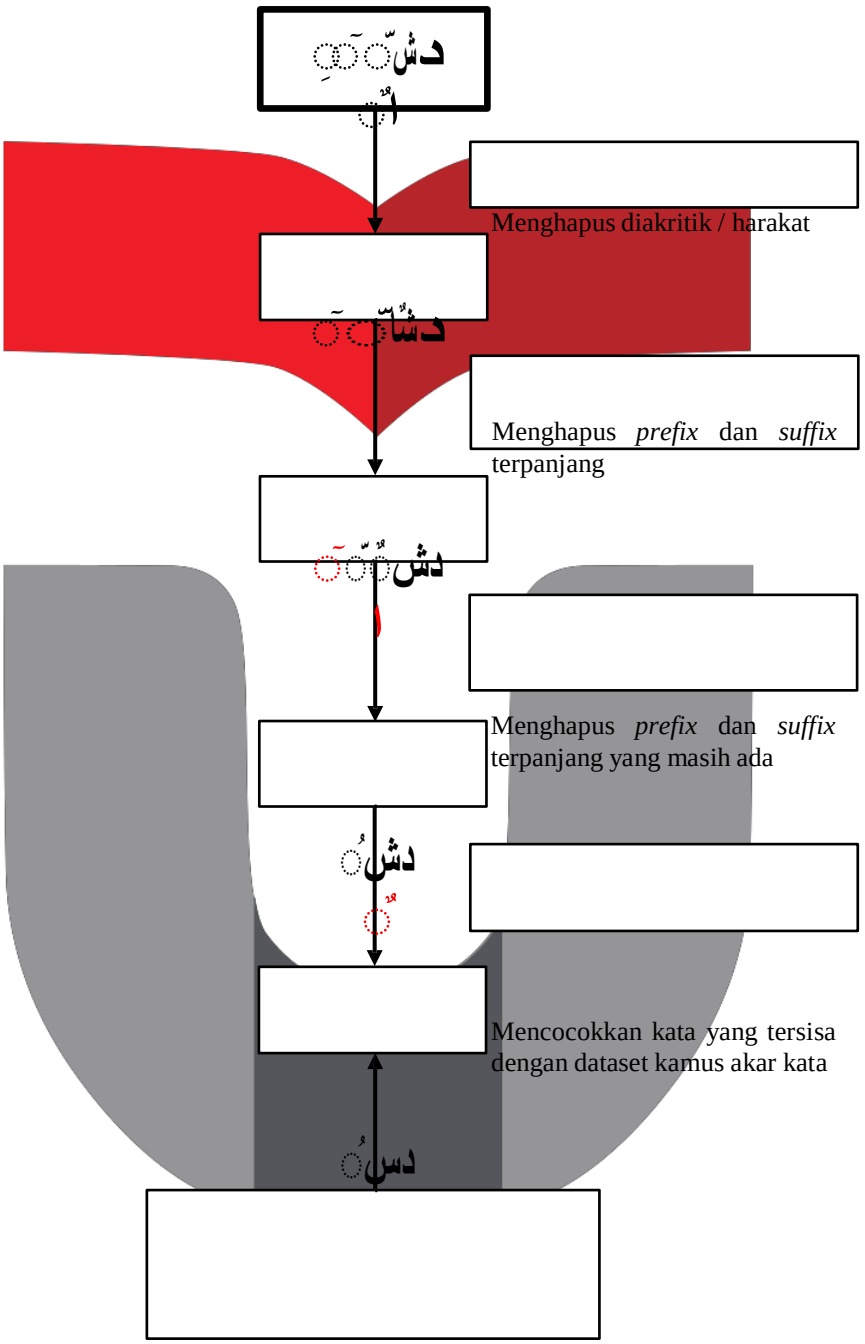
- Menghapus semua diakritik atau harakat dari hasil kata yang telah ditokenisasi.
- Menentukan setiap tanda *waqof* kedalam *Non Letter Words*.
- Melakukan normalisasi pada kata yang telah ditokenisasi, yang terdiri dari:
  - Mengubah إ atau أ menjadi ا
  - Mengubah آ menjadi ا
  - Mengubah ىء menjadi ئ
  - Mengubah ى menjadi ي
  - Mengubah ة menjadi ه
- Mencocokkan setiap kata hasil tokenisasi dengan **dataset kamus akar kata** guna diambil *root* atau akar kata nya dengan menghapus setiap *prefix* (awalan kata) dan *suffix* (akhiran kata) terpanjang yang terdapat pada kata.
- Selanjutnya ditentukan proses penghapusan *affixes* (*prefix* dan *suffix*).
  - Penghapusan *prefix* terdiri dari:
    - *Prefix* yang telah ditentukan pada dataset *prefix*.
    - Artikula yang telah ditentukan pada dataset artikula.
  - Penghapusan *suffix* terdiri dari kata-kata yang telah ditentukan pada dataset *suffix*.

Setelah semua *prefix* dan *suffix* dihilangkan, maka kata yang tersisa akan dicocokkan dengan dataset kamus akar kata untuk ditentukan akar kata nya. Berikut aturan pencarian *root* yang terdapat pada algoritma khoja stemmer:

- Apabila root yang dihasilkan terdiri dari 3 huruf, maka sistem akan mengeluarkan *output* yang telah ditentukan pada dataset *Tri roots*.
- Apabila root yang dihasilkan terdiri dari 4 huruf, maka sistem akan mengeluarkan *output* yang telah ditentukan pada dataset *Quad roots*.

Apabila sistem berhasil menemukan *root* atau akar kata dari sebuah kata, maka sistem akan mengeluarkan *output* berupa **Stemmed Words** yaitu semua kata yang berhasil di *stemming*. Apabila sistem

tidak menemukan *root* dari kata tersebut, maka sistem akan melakukan pengecekan apakah kata tersebut merupakan *stopwords* atau bukan. Apabila kata tersebut merupakan *stopwords*, maka sistem akan mengelurkan *output* berupa **Stopwords** yaitu semua kata yang termasuk ke dalam kata umum dan apabila kata tersebut bukan merupakan *stopwords*, maka sistem akan mengeluarkan *output* berupa **Words Not Stemmed** yaitu semua kata yang tidak berhasil di *stemming*. Apabila sistem membaca tanda *waqof* pada *input* data, maka sistem akan memasukkannya kedalam kategori **Non Letter Words.** Berikut ini adalah alur Khoja Stemmer dalam menghasilkan *root* atau *stem* pada suatu kata dalam Al-Quran.



Apabila kata yang tersisa cocok dengan daftar akar kata yang ada pada dataset kamus akar kata, maka output yang dihasilkan berupa kata tersebut.

2.4 Akurasi Sistem

Pada tahap ini, sistem akan menghasilkan akurasi nya secara keseluruhan dari hasil kata yang berhasil di *stemming* yang masuk dalam kategori *stemmed words*, dan *stopwords*.

3. Evaluasi

3.1 Stemmed Words

Pengujian ini ialah berupa semua kata yang berhasil di *stemming* yang menunjukkan bahwa aplikasi *stemmer* yang menerapkan algoritma Shereen Khoja Stemmer berjalan dengan baik. Pada tabel 4.2 dijabarkan contoh kata yang berhasil di *stemming* oleh aplikasi.

Stemmed Words

Tabel 4.1: Kata yang berhasil di *stemming*

Pada tabel diatas, setiap ayat yang dijabarkan perbarisnya dapat ditemukan *root* atau akar kata nya dan setiap *root* nya di munculkan berdasarkan urutan kata nya dalam setiap ayat Al-Quran yang diinput.
Contoh dalam ayat pertama pada surat Al-Fatihah yaitu pada kalimat:

Kata *bismi* /  , *arrahman* / dan *arrahim* / berhasil dicari *root* nya serta dihilangkan semua harakatnya oleh *stemmer* sehingga kata-kata tersebut dimasukkan kedalam *output* “Stemmed Word”. *Stemmer* memasukkannya kedalam *output* “stemmed word” karena seluruh kata yang *root* nya berhasil ditemukan cocok dengan dataset kamus akar kata dan *processing* yang dilakukan oleh algoritma Shereen Khoja Stemmer dalam melakukan proses *stemming* pada bahasa arab.

Berikut ini penjabaran hasil analisis lainnya yang penulis lakukan terhadap pengujian *Stemmed Words* dengan membandingkan *output* dari Khoja Stemmer dengan kamus Al-Quran secara manual.

TOKENISASI	OUTPUT KHOJA STEMMER	KAMUS AL-QURAN	KETERANGAN *ROOT*
			SALAH
			**BENAR**
			BENAR
			BENAR
			SALAH
			BENAR
			BENAR
			**BENAR**
			**BENAR**
			SALAH
			BENAR
			SALAH
			BENAR

امنسّ	لَ	قِ	**BENAR**

- Kata-kata yang masuk kedalam kategori “stemmed words” juga masuk kedalam kategori *fi’il madhi* (kata kerja lampau) dan *isim musytaq* (kata benda yang memiliki akar kata), contoh: ذجع (sembah) dan عِ (alam semesta).

- Apabila sistem menghasilkan *root* yang terdiri dari 2 huruf, maka akan timbul beberapa kasus:
  - Terjadi duplikasi kata pada huruf yang terakhir. Contoh:]دد : ردّ[, [ّ:[.
  - Terjadi penambahan huruf *'illat* atau huruf lemah (ا , و , ي) diawal, ditengah, atau diakhir kata. Contoh:]ّ:أ[, [ّل[.
- Apabila sistem menghasilkan *root* yang terdiri dari 3 huruf, maka akan timbul beberapa kasus:
  - Apabila ada huruf „*illat* atau huruf lemah (ا , و , ي) setelah huruf pertama, maka huruf „*illat* tersebut diubah menjadi ـ. Contoh:]غاش : غش[,]ذا : ذ[.
  - Apabila ada huruf yang disukun dan huruf *illat* atau huruf lemah (ا , ي) diakhir, maka huruf „*illat* tersebut diubah menjadi ـ. Contoh:]خا : خ[.
  - Apabila ada huruf yang disukun dan huruf *illat* atau huruf lemah (ـ , ي) diakhir, maka huruf „*illat* tersebut diubah menjadi ا. Contoh:]ذ : أ[, [ش : أشج[.
  - Hal diatas terjadi karena *stemmer* tidak memiliki pola yang baku terhadap penurunan huruf *'illat* pada *root* yang dihasilkan.
- *Stemmer* tidak dapat mengeluarkan *root* yang tepat apabila dalam kata tersebut tidak mengandung huruf yang menjadi *root* nya. Contoh:]و : م[dimana *root* yang tepat dari kata tersebut ialah ا.
- *Stemmer* tidak dapat mengeluarkan *root* yang tepat apabila dalam kata tersebut salah satu huruf dari *root* nya ialah termasuk *prefix* atau *suffix* (dimana didalam Khoja Stemmer harus dihilangkan) sehingga sistem akan mengeluarkan *root* berupa kata yang mendekati kata yang di *stem* tersebut. Contoh:]تجيس : جس[dimana *root* yang tepat dari kata tersebut ialah شجي. Huruf ت disini

  merupakan salah satu huruf *prefix* sehingga huruf tersebut tidak dapat menjadi *root*.

3.2 Words Not Stemmed

Pengujian ini ialah berupa semua kata yang tidak berhasil dicari *root* nya oleh khoja stemmer. Pada tabel 4.3 dijabarkan contoh kata yang tidak berhasil di *stemming* oleh aplikasi.

Words Not Stemmed

ا أ عيو يبو شيو ضبا ا صو تءيضا و ذ ا ب ح بشغ
خلق ٢ ب ملس ضبا استع
ب فع وب ا ج أ ا شا إ ا ا شا إ ا ي ب إ تثن ب ي ب شي ي ب ث ز ا
ض ض ا م ا ع ا
ب ب ب ب ب ب خز ب ت ن ج ذ ب ب ب ب ع
فبا ي ش ب ظ ي ي ن يب خ ب ذ ء
ز ج و تذا ب ع ذ ش ا خ ل ط ش ذ ن ز ئ ت ج
١٢ ذ ي ب أ
ش ج ج و ع ج و ج و ج ع ب ع ش ي ع ج ع ج ع و أ أ ذ ب
ا شا أ ذ

Tabel 4.2: Kata yang tidak berhasil di *stemming*

Pada tabel diatas, setiap kata yang tidak berhasil dicari *root* nya oleh *stemmer* di munculkan berdasarkan urutan kata nya dalam setiap ayat Al-Quran yang di *input*. Apabila *stemmer* tidak mengidentifikasi kata yang di *input* memiliki *root* atau sebagai *stopwords*, maka *stemmer* akan mengembalikan kata tersebut dan memasukkannya kedalam *output* "Words not Stemmed". Hal tersebut menunjukkan kekurangan dari algoritma Khoja Stemmer sendiri dalam melakukan *stemming* pada Al-Quran serta kurangnya kosakata pada kamus akar kata Khoja Stemmer. Berikut ini penjabaran hasil analisis yang penulis lakukan terhadap pengujian *Words Not Stemmed* dengan membandingkan *output* dari Khoja Stemmer dengan kamus Al-Quran secara manual disertai alasan kenapa *stemmer* tidak dapat membaca kata-kata tersebut.

TOKENISASI	OUTPUT KHOJA STEMMER	KAMUS AL-QURAN
اعيس	اعيس	ع
ابضا	ابض	ي ض
بلسم	بلسم	ق س

س ِ ر ِ ا	ح ش خ ٢ ب ث ّ	ح ش خ ٢ ب ث ّ ِ
ش ِ غ ِ ّ ّ	ح ّ ب ش غ	ح ّ ب ش غ ِ
ب ئ ا	ب ئ ا	ب ئ ا
د ِ ّ ّ ّ ّ	، ذ ن ّ ّ ِ	، ذ ن ّ ّ ّ ِ

ثَوَبٌ	ثَوَبَ	يَثُوبُ
تَعِيضُ	تَعِيضُ	عَوَضِ
تَاصِوِ	تَاصِو	ةَ
عِيضُ	عِيضُ	عَوَضِ

Tabel 4.3: pengujian *words not stemmed* secara manual

- Semua huruf nya termasuk kedalam *prefix* dan *suffix*, sehingga kata nya tidak dapat dibaca oleh *stemmer*. Contoh: بِ , بِأ.
- *Root* nya termasuk *prefix* dan atau *suffix* serta kata yang tersisa tidak dihimpun kedalam *dataset* kamus akar kata. Contoh: ذَنْبِ yang memiliki *root* yaitu دِ dimana huruf ialah sebuah *suffix* dan kata yang tersisa yaitu دِ tidak dihimpun kedalam *dataset* kamus akar kata, sehingga *stemmer* tidak dapat membaca kata tersebut. Contoh lainnya ialah تَعِيضُأ, عَسْ , ثِشَخْ.
- Semua huruf nya tidak termasuk kedalam *prefix* dan atau *suffix* serta kata yang tersisa tidak dihimpun kedalam *dataset* kamus akar kata. Contoh: ثَوَبٌ dimana kata tersebut seharusnya memiliki *root* yaitu يَثُ. Karena huruf ن bukan merupakan *prefix*, maka huruf tersebut tidak dapat dihilangkan, sehingga *stemmer* akan mengembalikan kata tersebut.
- Kata-kata yang termasuk kedalam kategori "Words Not Stemmed" juga termasuk kedalam kategori *isim jamak mudzakkar saalim* yaitu *isim* yang menunjukkan makna banyak bagi laki-laki, contoh: ذَنْبِ , ضَا , عَسْ.

3.3 Stopwords

Pengujian ini ialah berupa semua kata yang masuk kedalam *stopwords* yang telah dihimpun pada dataset kamus akar kata yang digunakan oleh khoja stemmer. Pada tabel 4.4 dijabarkan contoh kata yang termasuk kedalam *stopword* oleh aplikasi.

Stopwords
]لِلَّهِ[ ]عَلَى:عَلَيْ[ ]عَلَى:عَلَيْهِمْ[ ]ذَا:ذَا[ ]بِهِ:بِهِ[ ]ا:لَا[ ]لَا:لَا[ ]مِنْ[ ]فِي:فِي[ ]لِلَّهِ:لِلَّهِ[
]ه:هُم[ ]إِ[ ]بِ:بِ[ ]بِهِ:بِهِ[ ]ذَا:ذَا[ ]ذَا:ذَا[ ]الَّذِي:الَّذِي[ ]لَا[ ]هُم:هُم[ ]لَ:لَ[
]لِلَّهِ:لِلَّهِ[ ]لَ:لَ[ ]إِ[ ]عَ:عَ[ ]ذَا:ذَا[ ]أَ:أَ[ ]إِ[ ]عَلَ:عَلَ[ ]عَلَيْ:عَلَيْ[ ]ا[ ]بِهَا:بِهَا[ ]إِ[ ]إِ[ ]إِ[ ]إِ[ ]عَلَيْ:عَلَيْ[ ]عَلَيْ:عَلَيْ[ ]عَ:عَ[
]بِهِ:بِهِ[ ]إِ[ ]بِ:بِ[ ]لِلَّهِ:لِلَّهِ[ ]إِ[ ]لَ:لَ[ ]ذَا:ذَا[ ]لِ:لِ[ ]أَ:أَ[ ]إِ[ ]لِلَّهِ:لِلَّهِ[ ]ا:ا[
]إِ[ ]يَ:يَ[ ]لَ:لَ[ ]إِ[ ]أ:أ[

Tabel 4.4: *Stopwords*

Pada tabel diatas, setiap ayat yang dijabarkan perbarisnya dihasilkan *stopwords* nya dan setiap *stopwords* di munculkan berdasarkan urutan kata nya dalam setiap ayat Al-Quran yang di *input*. Contoh diatas ialah beberapa *stopwords* yang dihasilkan oleh *stemmer*.

Stopwords juga dihasilkan dengan proses *stem* dari kata yang di *input*. Apabila sistem mengidentifikasi *root* dari kata yang di *input* sebagai kata umum, maka sistem akan memasukannya kedalam *output* "Stopwords". Pada pengujian diatas, *stemmer* menghasilkan kesalahan dalam membaca beberapa lafadz Allah / لِلَّهِ. *Stemmer* menghasilkan *root* dari kata Allah / لِلَّهِ menjadi *lah* / لَهُ.

Kesalahan tersebut terjadi karena dalam kamus akar kata yang digunakan oleh khoja stemmer, kata Allah / لِلَّهِ tidak dihimpun kedalam dataset kata-kata yang termasuk *stopwords*, sedangkan kata *lah* / لَهُ termasuk kata yang dihimpun kedalam dataset *stopwords*, sehingga *stemmer* hanya dapat menghasilkan *stopwords* yang mendekati hasil dari kata yang dibaca tersebut. Berikut ini penjabaran hasil analisis lainnya yang penulis lakukan terhadap pengujian *stopwords* dengan membandingkan *output* dari Khoja Stemmer dengan kamus Al-Quran secara manual.

TOKENISASI	OUTPUT KHOJA STEMMER	KAMUS AL-QURAN	KETERANGAN *ROOT*
لِلَّهِ	لَهُ	يَأَلَهُ	SALAH
فِي	فِي	فِي	BENAR
بِهِ	ا	ا	SALAH

SALAH	اِصْلَحَ	اَ	نَبَإِ
SALAH	رِصْيِصَةَ	زَكَاَ	زَكَاتِهِ

اِعَمُ	عِ	اِعَمُ	SALAH
هُرِ	هُر	نِيِر	BENAR
لَ	لَ	لَ	BENAR
اف	اف	اف	BENAR
بّ	بِ	بّ	SALAH
بّث	بِ	بّث	SALAH

Tabel 4.5: Pengujian *stopwords* secara manual

- Apabila akar kata yang dihasilkan salah, maka akar kata yang tepat tersebut tidak dihimpun dalam *dataset* kamus akar kata khoja stemmer. Sehingga *stemmer* hanya dapat menghasilkan *stopwords* yang dihimpun dalam dataset kamus akar kata yang mendekati hasil dari kata yang dibaca tersebut. Contoh: نابٍ dan نابًا yang memiliki akar kata menurut *output* dari *stemmer* dimana seharusnya akar dari kata tersebut ialah .
- Apabila huruf berdiri sendiri, maka huruf tersebut merupakan *stopwords*. Tetapi apabila huruf sebagai *wawu athaf* atau huruf yang tidak dapat berdiri sendiri tanpa diikuti oleh kata lainnya, maka huruf tersebut akan dianggap sebagai *prefix* dan dihilangkan. Contoh: بِ yang memiliki akar kata بِ, dimana seharusnya akar kata dari kata tersebut ialah kata itu sendiri yaitu بّ.
- Semua kata yang termasuk kedalam *stopwords* pada khoja stemmer memiliki ciri-ciri sebagai berikut:
  - Kata sambung atau kata depan (kelompok *harf*) yaitu kata yang tidak dapat berdiri sendiri tanpa diikuti oleh kata lainnya, contoh: لَ (tidak ada), اف (didalamnya).
  - Kata yang menunjukkan keterangan waktu, contoh: (hari), (sebelum).
  - Kata yang menunjukkan keterangan tempat, contoh: عِ (atas).

3.4 Akurasi Sistem

Berikut ini tabel rincian dalam angka dari hasil pengujian aplikasi.

Data Uji	Stemmed Words	Words Not Stemmed	Stopwords	Non Letter Words (Tanda Waqof)	Total Words	Akurasi (%)
Juz 1-15	23.278	1.898	13.754	2.556	41.486	95,42
Juz 16-30	24.191	1.984	13.140	1.823	41.138	95,17
						95,295

Tabel 4.6: Hasil pengujian

Dari tabel diatas, data uji dibagi menjadi 2 bagian untuk meminimalisasi waktu eksekusi aplikasi. Pada data uji pertama, yaitu juz 1-15 dihasilkan **stemmed words** berjumlah 23.278, **words not stemmed** berjumlah 1.898, **stopwords** berjumlah 13.754 kata, dan **non letter words** berjumlah 2.556 sehingga seluruh kata yang di proses pada data uji pertama berjumlah 41.486. Akurasi *stemming* yang dihasilkan pada data uji pertama sebesar 95,42%.

Pada data uji kedua, yaitu juz 16-30 dihasilkan **stemmed words** berjumlah 24.191, **words not stemmed** berjumlah 1.984, **stopwords** berjumlah 13.140 kata, dan **non letter words** berjumlah 1.823 sehingga seluruh kata yang di proses pada data uji kedua berjumlah 41.138. Akurasi *stemming* yang dihasilkan pada data uji kedua sebesar 95,17%. Jadi total akurasi *stemming* yang dihasilkan dari seluruh data uji ialah sebesar 95,295%.

4. Kesimpulan

Walaupun Algoritma Khoja Stemmer memiliki akurasi dalam melakukan *stemming* pada Al-Quran sebesar 95,295%, tetapi *root* yang dihasilkan oleh *stemmer* apabila diperiksa secara manual dan dibandingkan dengan kamus Al-Quran ternyata masih banyak melakukan kesalahan. Sehingga *stemmer* ini kurang cocok untuk diterapkan pada aplikasi *stemming* untuk Al-Quran.

5. Daftar Pustaka

[1] A. Nurdin, dalam *Quranic Society: Menelusuri Konsep Masyarakat Ideal dalam Al-Qur'an*, Jakarta, Erlangga, 2006, p. 1.

[2] M. Zaenuddin, "Morfologi Arab," *Direktori File UPI.*

[3] A. Firmanto, S. Widowati dan A. Rakhmatsyah, "Implemetasi Principal Component Analysis dan Backpropagation Neural Network dalam Pengklasifikasian Terjamahan Ayat-ayat Ilmu Pengetahuan dalam Al-Quran," 2011.

[4] M. N. Al-Kabia, S. A. Kazakzehb, B. M. A. Atab, S. A. Al-Rababahc dan I. M. Alsmadid, "A novel root based Arabic stemmer," *Journal of King Saud University - Computer and Information Sciences ,* vol. 27, no. 2, pp. 94-103, 2015.

[5] [Online]. Available: http://repository.usu.ac.id/bitstream/123456789/53366/4/Chapter%20II.pdf.

[6] N. Thabet, "Stemming the Qur'an," *Proceedings of the Workshop on Computational Approaches to Arabic Script-based Languages,* pp. 85-88, 2004.

[7] M. Sawalha dan E. Atwell, "Comparative Evaluation of Arabic Language Morphological Analysers and Stemmers," *White Horse Research Online,* pp. 107-110, 2008.

[8] S. Khoja, "Sheereen Khoja Associate Professor of Computer Science," Pacific University, [Online]. Available: http://zeus.cs.pacificu.edu/shereen/research.htm. [Diakses 22 Februari 2018].

[9] "Belajar Bahasa Al-Quran, Metoda "Belajar Aktif Kata PerKata Lewat Intra/Internet"," [Online]. Available: http://quran.bblm.go.id/. [Diakses 20 Februari 2018].

[10] T. M. T. Sembok dan B. A. Ata, "Arabic Word Stemming Algorithms and Retrieval Effectiveness," *Proceedings of the World Congress on Engineering ,* vol. III, pp. 3-5, 2013.

[11] L. S. Larkey dan M. E. Connell, "Arabic Information Retrieval at UMass in TREC-10," *TREC,* pp. 562-570, 2001.