Uji Coba *Stemming* ECS (*Enhance Confix Stripping*) Ayat-Ayat Al Qur'an Dan Hadist Terjemahan Bahasa Indonesia

Tristyanti Yusnitasari[1], Lily Wulandari[2], Irfan Humaini[3], Diana Ikasari[4].
Universitas Gunadarma (Jl. Margonda Raya 100, Depok, Jawa Barat, INDONESIA – 16424)
tyusnita@staff.gunadarma.ac.id[1], lily@staff.gunadarma.ac.id[2], irfan_humaini@staff.gunadarma.ac.id[3], d_ikasari@staff.gunadarma.ac.id[4]

Abstract- **The Quran is a holy book which is the main guidance and source of the Muslim law, after the Qur'an another source which is also an important reference is the Hadith, because the Hadith will explain in more detail what has been described in the Qur'an. Research conducted is information retrieval system with information retrieval technique which is technique in searching and tracing relevant information. Information retrieval techniques are used to find similarities between the keywords entered with documents stored in the database. The database consists of documents of the Qur'an and Hadith. The developed system can produce relevant information with a high degree of precision in the process of searching documents of verses of the Qur'an and Hadith. The focus of this research is the stemming of ECS (Enhance Confix Stripping) and the creation of the Al Quran and Hadist corpus index based on the meaning or synonymous equations of the words in Al Quran and Hadist of Indonesian translation.**

Keywords: Information Retrieval, Corpus Index, ECS

Abstrak— Al Quran adalah kitab suci yang merupakan pedoman dan sumber hukum utama umat Islam, setelah Al Quran sumber lain yang juga menjadi acuan penting adalah Hadist, karena Hadist akan menjelaskan lebih rinci mengenai apa yang telah dijelaskan dalam Al Quran. Penelitian yang dilakukan adalah sistem pencarian informasi dengan teknik temu kembali informasi yang merupakan teknik dalam pencarian dan penelusuran informasi yang relevan. Teknik temu kembali informasi *(information retrieval)* digunakan untuk mencari kemiripan antara kata kunci yang dimasukkan dengan dokumen yang tersimpan dalam basis data. Basis data terdiri dari dokumen Al Qur'an dan Hadits. Sistem yang dikembangkan dapat menghasilkan informasi yang relevan dengan tingkat presisi tinggi pada proses pencarian dokumen ayat-ayat Al Qur'an dan Hadits. Fokus dari penelitian ini adalah uji coba stemming ECS (*Enhance Confix Stripping)* dan pembuatan indeks korpus Al Quran dan Hadist berdasarkan persamaan makna atau sinonim dari kata-kata dalam Al Quran dan Hadist terjemahan bahasa Indonesia.

Kata Kunci: *Information Retrieval, Indeks Korpus, ECS.*

I. PENDAHULUAN

Al Qur'an adalah kitab suci umat Islam, hal-hal yang terkandung di dalam Al Qur'an berhubungan dengan keimanan, ilmu pengetahuan, hukum, peraturan-peraturan yang mengatur tingkah laku dan tata cara hidup manusia, kisah-kisah umat sebelumnya, ibadah serta tauhid (pengesaan Allah).

Hadist juga termasuk pedoman hidup dalam ajaran Islam. Hadits menurut istilah ahli hadits adalah apa yang disandarkan kepada Nabi Muhammad Shalallahu alaihi wa sallam, baik berupa ucapan, perbuatan, penetapan, sifat, atau *sirah* beliau, baik sebelum kenabian ataupun sesudahnya. Sudah merupakan kewajiban umat Islam (muslim) untuk mengimplementasikan kehidupan sehari-hari berdasarkan petunjuk Al Qur'an dan Hadist supaya mendapatkan kehidupan yang baik di dunia dan di akherat. Sebelum mengimplementasikannya tentu dipelajari terlebih dahulu hal-hal yang terkandung dalam Al Qur'an dan Hadist.

Sebagai seorang muslim mungkin hampir semua mengetahui apa saja yang dilarang atau diharamkan dan apa saja yang diperbolehkan sesuai dengan Al Qur'an dan Hadist. Tidak sedikit yang mengetahui hal tersebut, hanya sekedar mendengar bahwa apa saja yang dilarang atau apa saja yang tidak dilarang tanpa mengetahui secara benar bahwa sesungguhnya yang diperdengarkan adalah tertulis di dalam Al Qur'an dan Hadist. Seperti contoh hampir semua umat muslim mengetahui bahwa babi diharamkan untuk dikonsumsi oleh muslim, tetapi banyak yang tidak mengetahui secara pasti bahwa larangan tersebut ada di dalam surat dan ayat berapa pada Al Qur'an serta Hadist yang mengemukakan mengenai larangan mengkonsumsi babi. Banyak lagi contoh lain seperti makanan apa saja yang diharamkan atau minuman yang diharamkan, keutamaan sholat dan lain sebagainya. Keterbatasan waktu untuk mencari informasi mengenai hal tersebut adalah salah satu alasan dan kesulitan mencari kata yang diinginkan untuk dicari di dalam Al Qur'an dan Hadist karena Al Qur'an terdiri dari 30 Juz, 114 Surat dan 6326 Ayat sehingga untuk mencari kata yang sesuai dengan tema yang diinginkan akan sulit sekali. Sudah ada beberapa terjemahan yang melakukan index terhadap isi dari Al Qur'an, dan sudah banyak pengembang perangkat lunak megembangkan *digital Qur'an* dan *digital Hadist.* Pada beberapa perangkat lunak yang ada, pencarian informasi seperti mencari kata babi maka hasil pencarian adalah nama ayat dan surat mengenai kata babi, sementara jika mencari kata makanan yang diharamkan ayat atau surat yang mengandung kata babi bukan merupakan bagian dari hasil pencarian.

Penelitian yang dilakukan di sini adalah sistem pencarian informasi dengan teknik temu kembali informasi yang merupakan teknik yang digunakan dalam pencarian dan penelusuran informasi yang relevan. Teknik temu kembali informasi *(information retrieval)* digunakan untuk mencari kemiripan antara kata kunci yang dimasukkan dengan dokumen yang tersimpan di dalam basis data. Basis data terdiri dari dokumen Al Qur'an dan Hadits. Sistem yang dihasilkan dapat mempercepat proses pencarian pada dokumen Al Qur'an dan Hadits, serta menghasilkan informasi yang relevan.

Pada teknik Temu Kembali Informasi, terdapat beberapa pemodelan yang biasa digunakan, yaitu model *boolean, Vector Space model*, model *probabilistic* dan lain-lain. Pada pencarian Al Qur'an dan hadits, Temu kembali informasi digunakan untuk menampilkan beberapa ayat Al Qur'an dan Hadits yang berhubungan dengan kata kunci yang dicari sesuai kriteria tertentu. Data yang digunakan pada penelitian ini bersumber dari terjemahan kementrian Agama Republik Indonesia. Selanjutnya *pre-processing* dilakukan pada data tersebut. Proses *pre-processing* meliputi *tokenizing, filtering*, pembentukan *Inverted Index* dan *stemming*. Metode *stemming* yang digunakan pada penelitian ini adalah *stemming* ECS (*Enhanced Confix Stripping*). Sedangkan proses pencariannya akan mengembangkan salah satu pemodelan untuk meningkatkan presisi agar temu kembali informasi menjadi lebih relevan.

II. TINJAUAN PUSTAKA

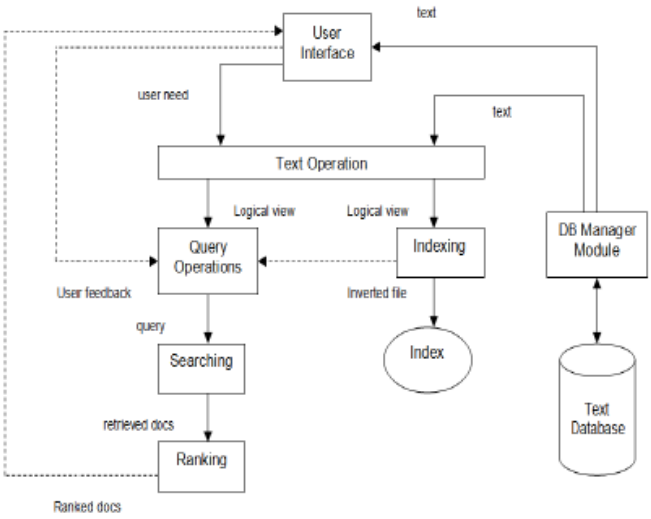
A. Temu Kembali Informasi *(Information Retriveal)*

Information retrieval (IR) *system* digunakan untuk menemukan kembali (*retrieve*) informasi-informasi yang relevan terhadap kebutuhan pengguna dari suatu kumpulan informasi secara otomatis [1][2]. Beberapa ahli mendefinisikan *Information Retrieval* sebagai berikut:

1. Manning, mendefinisikan bahwa *Information Retrieval* adalah proses menemukan material (dokumen) dari lingkungan bersifat tidak terstruktur (biasanya teks) yang memenuhi kebutuhan informasi dari dalam koleksi yang berukuran besar (biasanya pada komputer)[7].
2. Menurut Baeza-Yates, *Information Retrieval* adalah bagian dari ilmu komputer yang mempelajari tentang pengumpulan data dan temu kembali dokumen[3].
3. Greengrass, menyatakan bahwa *Information Retrieval* adalah sebuah disiplin ilmu yang berhubungan dengan pencarian data tidak terstruktur, khususnya dokumen-dokumen tekstual, dalam respon terhadap sebuah *query* atau pernyataan topik.[6]

Salah satu aplikasi umum dari *information retrieval system* adalah *search engine* atau mesin pencarian yang terdapat pada jaringan internet. Sebagai suatu sistem, *information retrieval system* memiliki beberapa bagian yang membangun sistem secara keseluruhan. *Information Retrieval* merupakan bagian dari *computer science* yang berhubungan dengan pengambilan informasi dari dokumen-dokumen yang didasarkan pada isi dan konteks dari dokumen-dokumen itu sendiri. Berdasarkan referensi dijelaskan bahwa *information retrieval* merupakan suatu pencarian informasi yang didasarkan pada suatu query yang diharapkan dapat memenuhi keinginan user dari kumpulan dokumen yang ada. Prinsip kerja sistem temu kembali informasi jika ada sebuah kumpulan dokumen dan seorang user yang memformulasikan sebuah pertanyaan (*request* atau *query*). Jawaban dari pertanyaan tersebut adalah sekumpulan dokumen yang relevan dan membuang dokumen yang tidak relevan [5].

Bagian-bagian Temu Kembali Informasi



Gambar 1 Proses Temu Kembali Informasi

Berdasarkan proses temu kembali informasi pada gambar 1 memiliki bagian-bagian *information retrieval.* Baeza-Yates membagi *information retrieval* ke dalam lima bagian [3];

B. Text Operations / Preprocessing

Text operations (pengoperasian teks) adalah proses transformasi dokumen dan *query* menjadi kata-kata indeks. Dalam suatu dokumen terdapat beberapa kata yang memiliki makna lebih penting dibandingkan kata-kata lainnya, sehingga *preprocessing* terhadap teks dalam suatu koleksi dokumen dianggap perlu dalam menentukan kata yang akan digunakan sebagai *index terms.* Pada tahap *preprocessing* ini juga mencakup pengoperasian teks seperti penghapusan *markup*, penghapusan *stopwords*, dan *stemming* (pembentukan kata dasar).

a. *Tokenizing*
   Tokenizing atau tokenisasi yaitu proses penguraian deskripsi yang semula berupa kalimat-kalimat menjadi kata-kata dan menghilangkan delimiter-delimiter dan karakter yang ada pada kata tersebut.

Tujuan dari proses ini adalah untuk membentuk *token* yaitu setiap kata dalam *string*

b. Korpus

Proses information retrieval membutuhkan database yang terdapat satu atau beberapa tabel sebagai tempat penyimpanan data yang akan diolah pada saat proses pencarian. Database memakai korpus untuk proses pembuatan tabel pendukungnya. Pada prinsipnya, setiap koleksi lebih dari satu teks dapat disebut dengan korpus, istilah korpus dalam bahasa latin berarti body, maka korpus dapat didefinisikan sebagai isi setiap teks. Tapi istilah korpus ketika digunakan dalam konteks linguistic modern memiliki konotasi yang lebih spesifik. Ada empat karakteristik dari korpus.

c. Penghapusan *Stopwords*

Kata-kata yang sering muncul dalam dokumen pada suatu koleksi *(corpus)* yang tidak mewakili indeks untuk melakukan pencarian disebut dengan *stopwords*. Kata-kata yang termasuk dalam kelompok *stopwords* ini biasanya akan disaring dari kelompok kata yang akan dijadikan sebagai indeks. Dalam bahasa inggris yang merupakan bagian dari *stopwords* adalah artikel*(articles)*, kata depan *(prepositions)*, dan kata penghubung*(conjunctions)*. Dengan membuang stopwords, ukuran struktur *indexing* dapat dikurangi dan dapat mengurangi *recall* dokumen yang tidak sesuai.

c. *Stemming*

Stem adalah bagian kata yang tinggal setelah menghilangkan imbuhannya (awalan dan akhiran). Seorang *user* sering kali memasukkan kata dalam *query* tetapi hanya beberapa dari kata-kata tersebut yang muncul dalam dokumen yang relevan. Bentuk jamak, bentuk kata kerja berakhiran –ing, dan akhiran dalam bentuk masa lampau adalah contoh-contoh dari jenis-jenis sintaksis kata yang akan menghalangi kecocokan antara query dengan dokumen yang berhubungan. Masalah ini dapat diselesaikan dengan substitusi kata dengan masing-masing bentuk asalnya.

2. ECS (*Enhance Confix Stripping*)

Algoritme *ECS Stemmer* ini merupakan Algoritme perbaikan dari Algoritme *Confix Stripping (CS) Stemmer*. Perbaikan yang dilakukan oleh *ECS Stemmer* adalah perbaikan beberapa aturan pada table acuan pemenggalan imbuhan. Selain itu, Algoritme *ECS Stemmer* juga menambahkan langkah pengembalian akhiran jika terjadi penghilangan akhiran yang seharusnya tidak dilakukan[1].

III. METODE PENELITIAN

1. Langkah Pertama dalam penelitian ini adalah studi literatur mengenai information retrieval dan studi literatur tentang Alquran dan Hadist.

2. Langkah Kedua Melakukan Pengamatan Terhadap Terjemahan Alquran Dan Hadist (Menyiapkan basis data).

- Pertama yang dilakukan adalah mendapatkan basis data Alquran terjemahan bahasa Indonesia versi Kementrian Agama Republik Indonesia dan Hadist Shahih Bukhori.
- Kedua dilakukan *pre-process* teks terjemahan Alquran dan Hadist yang terdiri dari *tokenizing*, *stopword removal* dan *stemming*.
- Ketiga dilakukan pembentukan *corpus* basis data yang sudah diolah dalam *preprocessing.*

Kerangka Kerja Sistem

Gambar 2. Kerangka Kerja Sistem

IV. PEMBAHASAN

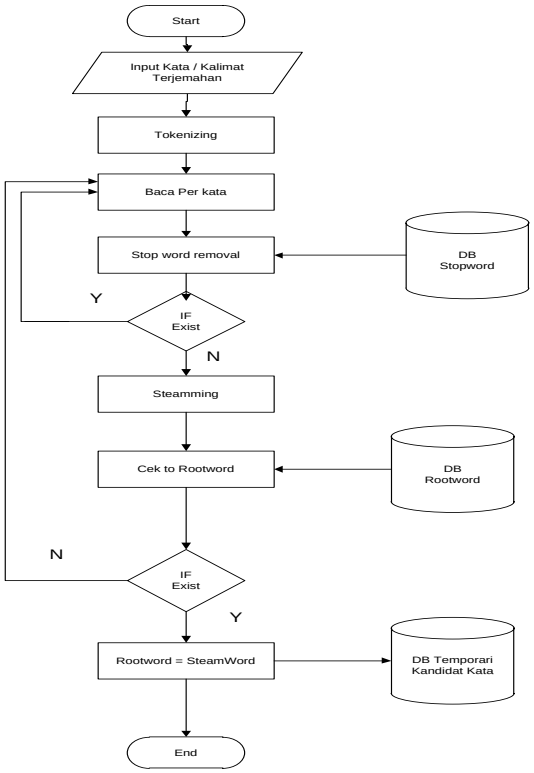
Pembentukan atau pembuatan indeks korpus merupakan hal yang paling krusial pada penelitian ini karena akan menjadi acuan utama dalam proses informatian retrieval pencarian ayat-ayat Al Quran dan Hadist. Proses ini masuk dalam *Preprocessing.*

- Pertama yang dilakukan adalah mendapatkan basis data Alquran terjemahan bahasa Indonesia versi Kementrian Agama Republik Indonesia dan Hadist Shahih Bukhori.
- Kedua dilakukan *pre-process* teks terjemahan Alquran dan Hadist yang terdiri dari *tokenizing, stopword removal* dan *stemming.*

Penjelasan skema sistem preprocessing pada gambar 4.1 adalah sebagai berikut:

- Proses awal pada preprocessing adalah melakukan input terhadap data ayat Alquran dan Hadist yang didapat dalam bentuk kalimat.
- Tokenizing
  Tokenizing merupakan proses pemisahan sebuah teks menjadi kata, frasa, symbol atau elemen bermakna lain yang disebut *Token* .

Proses Preprocessing



Gambar 3. Skema Sistem *Preprocessing*

Proses untuk melakukan *Tokenizing* adalah sebagai berikut :

a. Setelah input kalimat diterima, sistem akan mengubah semua karakter huruf besar menjadi huruf kecil.
b. Sistem selanjutnya akan menghilangkan tanda baca yang ada di dalam kalimat.
c. Akan dihasilkan kumpulan kata penyusun kalimat atau *Terms.*
d. Proses *Tokenizing* selesai.
   Contoh :
   a. Inputan kalimat Ayat Alquran sebagai berikut :

Dengan Nama Allah Yang Maha Pemurah Lagi Maha Penyayang.

   b. Proses perubahan menjadi huruf kecil, dan kalimat inputan berubah menjadi huruf kecil seluruhnya.

dengan nama allah yang maha pemurah lagi maha penyayang.

   c. Proses penghilangan semua tanda baca, menghasilkan pernyataan tanpa adanya tanda baca

dengan nama allah yang maha pemurah lagi maha penyayang

- *Stopword Removal*
  Biasa dikenal dengan *Filtering* adalah proses yang bertujuan untuk penghapusan, penghilangan atau pembuangan kata-kata yang tidak penting dan tidak memiliki makna pada kalimat input.
  Langkah-langkah untuk melakukan proses ini adalah :
  1. Output kata-kata dari proses *Tokenizing* akan digunakan sebagai input.
  2. Selanjutnya sistem akan membandingkan setiap kata-kata tersebut dengan setiap stopword yang terdapat didalam database.
  3. Apabila terjadi kesamaan kata dengan kata yang terdapat di dalam stopword database, maka kata tersebut dihilangkan. Apabila berbeda, maka kata tersebut akan disimpan dan untuk selanjutnya akan digunakan sebagai input pada proses selanjutnya.
  4. Proses *stopword removal* selesai.
     Contoh hasil proses *stopword removal* menggunakan hasil dari tokenizing

allah maha pemurah maha penyayang

- *Stemming*
  Stemming merupakan suatu proses untuk menemukan kata dasar dari sebuah kata *(root word)*. Dengan menghilangkan semua imbuhan (affixes) baik yang terdiri dari awalan (prefixes), sisipan (infixes), akhiran (suffixes) dan confixes (kombinasi dari awalan dan akhiran) pada kata turunan. Stemming digunakan untuk mengganti bentuk dari suatu kata menjadi kata dasar dari kata tersebut yang sesuai dengan struktur morfologi Bahasa Indonesia yang baik dan benar.
  Pada dasarnya, Algoritme *stemming* mengelompokkan imbuhan kedalam beberapa kategori sebagai berikut:

1. *Inflection Suffixes* yakni kelompok-kelompok akhiran yang tidak mengubah

bentuk kata dasar. Kelompok ini dapat dibagi menjadi dua:

- *Particle* (P) atau partikel, termasuk di dalamnya adalah partikel "-lah", kah", "-tah", dan "-pun".
  Contoh: jika kata "benarkah" maka akan ter-*stemming* kata "-kah" menjadi kata "benar" maka kata "benar" adalah termasuk kata dasar.
- *Possessive Pronoun* (PP) atau kata ganti kepunyaan, termasuk di dalamnya adalah "-ku" , "-mu", dan "-nya".
  Contoh: jika kata "miliknya" maka akan ter-*stemming* kata "-nya" menjadi kata "milik" maka kata "milik" adalah termasuk kata dasar.

2. *Derivation Suffixes* (DS) yakni kumpulan akhiran yang secara langsung dapat ditambahkan pada kata dasar. Termasuk di dalam tipe ini adalah akhiran "-i", "-kan", dan "-an".
   Contoh: jika kata "campuri" maka akan ter-*stemming* akhiran "-i" menjadi kata "campur" maka kata "campur" adalah termasuk kata dasar.
3. *Derivation Prefixes* (DP) yakni kumpulan awalan yang dapat langsung diberikan pada kata dasar murni, atau pada kata dasar yang sudah mendapatkan penambahan sampai dengan dua awalan. Termasuk di dalamnya adalah awalan yang dapat bermorfologi ("me-", "be-", "pe-", dan "te-") dan awalan yang tidak bermorfologi ("di-", "ke-" dan "se ").
   Contoh: jika kata "menyanyi" maka proses *stemming*nya yaitu menghilangkan kata "me-" dan terbentuklah kata "nyanyi" maka kata "nyanyi" adalah kata dasar.
   - Jika kata "dirumah" maka proses *stemming*nya yaitu menghilangkan awalan kata "di-" dan terbentuklah kata "rumah" maka kata "rumah" adalah termasuk kata dasar.

Algoritme ECS *stemmer* bekerja sebagai berikut:

1. Kata yang akan di-*stemming* dibandingkan ke dalam *database* kata dasar. Jika ditemukan, maka kata tersebut adalah kata dasar dan Algoritme berhenti. Jika kata tidak sesuai dengan kata dalam kamus kata dasar, maka lanjut ke langkah 2
2. Kata yang akan di *stemming* minimal terdiri dari kurang dari 3 huruf, jika kata terdiri kurang dari sama dengan 3 huruf maka kata tersebut termasuk kata dasar, jika tidak maka lanjut ke langkah 3.
3. Kata yang mengandung perulangan maka akan ter-*stemming* menjadi kedalam 1 kata atau tidak ada reduplikasi, kemudian kata tersebut di cek dengan *database* kata dasar, jika ditemukan maka kata tersebut adalah kata dasar, jika kata tidak terdapat dalam *database* kata dasar maka lanjut ke langkah 3
4. Cek *Rule Precedence* yaitu jika kata memiliki pasangan awalan-akhiran "be-lah","be-an","me-i","di-i","pe-i", atau "te-i" maka langkah *stemming* selanjutnya adalah 8, 5, 6, 7, 8, 9 tetapi jika kata yang di-input tidak memiliki pasangan awalan-akhiran tersebut, langkah *stemming* berjalan normal yaitu 5, 6, 7, 6, 8, 9
   Sebagai contoh: Jika kata yang diinput adalah "merencanakan" karena kata tersebut bukan termasuk kedalam aturan yang tidak diperbolehkan maka proses selanjutnya yaitu langsung ke langkah 5, 6, 7, 8, 9
5. Hilangkan partikel (P) dan kata ganti kepunyaan (PP). Pertama hilangkan partikel (P) ("-lah", "-kah", "-tah", "-pun"). Setelah itu hilangkan juga kata ganti kepunyaan (PP) ("-ku", "-mu", atau "-nya"). Sesuai dengan model imbuhan, menjadi : [[[DP+]DP+]DP+] Kata Dasar [+DS] Kata "merencanakan" maka proses pada aturan 3 ini tidak ada karena kata "-kan" termasuk kedalam *Derrivation Suffixes*.
6. Identifikasi kata yang mengandung kombinasi awalan dan akhiran yang tidak diperbolehkan yang ada pada tabel 2.1, jika terdapat kata yang mengandung kombinasi awalan dan akhiran yang tidak diperbolehkan, maka kata tersebut dianggap sebagai kata dasar dan Algoritme berhenti. Jika tidak adavkata yang mengandung kombinasi awalan dan akhiran yang dilarang maka lanjut ke langkah 6
7. Hilangkan juga akhiran (DS) ("-i","-an", dan "-kan"), sesuai dengan model imbuhan, maka menjadi: [[[DP+]DP+]DP+] Kata Dasar, Selanjutnya pada proses kata "merencanakan" akan membuang kata "-kan" karena kata "-kan" termasuk *derrivation suffixes* atau akhiran. Sehingga kata yang didapat adalah "merencana". Karena kata "merencana" bukan termasuk kata dasar, maka proses selanjutnya yaitu ke langkah 5.
8. Penghilangan awalan (DP) ("di-", "ke-", "se-", "te-", "be-", "me-", dan "pe-") mengikuti langkah-langkah berikut:
   a. Algoritme akan berhenti jika:
      i. Terdapat kombinasi aturan yang tidak diperbolehkan
      ii. Awalan yang dideteksi saat ini sama dengan awalan yang dihilangkan sebelumnya.
      iii. Kata tersebut sudah tidak memiliki awalan.

   b. Identifikasi jenis awalan dan akhiran, yaitu:

i. Jika awalan dari kata adalah "di-", "ke-", atau "se-" maka awalan dapat langsung dihilangkan.
ii. Hapus awalan "te-", "me-", "be-", atau "pe-" yang menggunakan aturan peluruhan yaitu aturan pemenggalan Nazief Adrian modifikasi *Enhanced Confix Stripping*.
Selanjutnya kata "merencana" akan ter-*stemming* sesuai pada proses langkah kelima yaitu menghilangkan awalan atau *Derivation Prefixes*. Karena kata "me-" termasuk awalan maka kata "me-" akan ter-*stemming*. Maka kata "rencana" adalah kata yang didapat dari proses ini. Selanjutnya kata "rencana" di cek kedalam *database* kata dasar, karena kata "rencana" termasuk kata dasar, maka proses *stemming* berhenti sampai disini.

9. Jika semua langkah gagal, maka awalan- awalan yang telah diikembalikan lagi dan kata tersebut dianggap sebagai kata dasar.
Dalam menginput data dibedakan menjadi dua buah kategori yaitu data latih dan data uji. Memisahkan data menjadi data latih dan data uji dimaksudkan agar model yang diperoleh nantinya memiliki kemampuan generalisasi yang baik dalam melakukan klasifikasi data. Tidak jarang sebuah model klasifikasi dapat melakukan klasifikasi data dengan sangat baik pada data latih, tetapi sangat buruk dalam melakukan klasifikasi data yang baru dan belum pernah ada, hal ini dinamakan *overfitting*. Dalam *data mining* klasifikasi bisa digunakan untuk memprediksi kelas data dari data baru yang berdasarkan kelas yang sudah ditentukan (*predetermined class*) dari data yang sudah ada.

- Ketiga dilakukan pembentukan *corpus* basis data yang sudah diolah dalam *preprocessing. Corpus* dilakukan dengan cara membuat basis data kata-kata atau potongan kalimat yang memiliki makna yang sama. sebagai contoh kata korupsi tidak ada dalam terjemahan
Alquran dan Hadist bahasa Indonesia, sehingga pada inputan pencarian ayat menggunakan kata korupsi jika pencarian hanya berdasarkan persamaan kata maka ayat-ayat yang bermakna korupsi tidak akan muncul. Sedangkan jika pencarian kata / kalimat berdasarkan makna ayat-ayat tentang korupsi akan muncul, karena kata korupsi sudah dibuat indeks yang memiliki makna yang sama yaitu : memakan harta, merampas, dan lain-lain. Contoh Pembentukan Indeks Korpus :

TABEL I. KORPUS SINONIM

No	Kata Kunci	KBBI	Makna Sama dalam Al Quran dan Hadist
1	Korupsi	penyelewengan atau penyalahgunaan uang negara (perusahaan dsb) untuk keuntungan pribadi atau orang lain	1. Memakan harta 2. Harta Rampasan
2	Gosip	Obrolan tentang orang-orang lain; Cerita negatif tentang seseorang; pergunjingan	1. Membicarakan Orang 2. Menggunjing 3. kan Orang
3	Bir	Minuman mengandung alkohol yang dibuat dengan peragian lambat	1. Khamar 2. Minuman Keras 3. Mabuk

Tahapan selanjutnya implementasi metode *information retrieval* ayat-ayat Al Qur'an dan Hadist. Ujicoba metodologi hasil dari penelitian *information retrieval* ayat-ayat Al Qur'an dan Hadist untuk mengatahui hasilnya apakah sudah sesuai dengan yang diinginkan, pada tahap ini ujicoba juga dilakukan bersama pakar Alquran dan Hadist yang merupakan akademisi dibidang tersebut, hal ini untuk memastikan apakah keluarannya sudah sesuai dan benar berdasarkan makna dan secara keilmuan tidak menyalahi.

Tahap akhir yang dikerjakan selanjutnya dari penelitian ini yang merupakan *Output* dari *Information retrieval* ayat-ayat Al Qur'an dan Hadist. Pengujian kemampuan sistem *information retrieval* dilakukan dengan menghitung nilai *precision* dan *recall* berdasarkan kerelevanan sistem menampilkan dokumen sesuai dengan *query*. Nilai *precision* adalah keakurasian atau kecocokan antara permintaan informasi dengan jawaban terhadap perrmintaan itu yang hasilnya dapat dipertanggung jawabkan karena sudah berdasarkan penelitian ilmiah yang melibatkan pakar-pakar dibidang terkait.

Contoh pada Tabel II, ditampilkan hasil uji coba information retrieval dengan uji coba menggunakan kata kunci "Korupsi".

TABEL II. HASIL IR DENGAN KATA KUNCI KORUPSI

No	Surat/ Ayat	ISI AYAT ALQURAN
1	Al Baqarah 2. (188)	Dan janganlah sebahagian kamu ***memakan harta*** sebahagian yang lain di antara kamu dengan jalan yang bathil dan (janganlah) kamu membawa (urusan) harta itu kepada hakim, supaya kamu dapat memakan sebahagian daripada harta benda orang lain itu dengan (jalan berbuat) dosa, padahal kamu mengetahui.
2	An Nisa 4. (6)	Dan ujilah anak yatim itu sampai mereka cukup umur untuk kawin. Kemudian jika menurut pendapatmu mereka telah cerdas (pandai memelihara harta), maka serahkanlah kepada mereka harta-hartanya. Dan janganlah kamu ***makan harta*** anak yatim lebih dari batas kepatutan dan (janganlah kamu) tergesa-gesa (membelanjakannya) sebelum mereka dewasa. Barang siapa (di antara pemelihara itu) mampu, maka hendaklah ia menahan diri (dari memakan harta anak yatim itu) dan barangsiapa yang miskin, maka bolehlah ia makan harta itu menurut yang patut. Kemudian apabila kamu menyerahkan harta kepada mereka, maka hendaklah kamu adakan saksi-saksi (tentang penyerahan itu) bagi mereka. Dan cukuplah Allah sebagai Pengawas (atas persaksian itu).
3	An Nisa 4. (10)	Sesungguhnya orang-orang yang ***memakan harta*** anak yatim secara zalim, sebenarnya mereka itu menelan api sepenuh perutnya dan mereka akan masuk ke dalam api yang menyala-nyala (neraka).
4	An Nisa 4. (29)	Hai orang-orang yang beriman, janganlah kamu saling ***memakan harta*** sesamamu dengan jalan yang batil, kecuali dengan jalan perniagaan yang berlaku dengan suka sama-suka di antara kamu. Dan janganlah kamu membunuh dirimu; sesungguhnya Allah adalah Maha Penyayang kepadamu.
5	An Nisa 4. (161)	dan disebabkan mereka memakan riba, padahal sesungguhnya mereka telah dilarang daripadanya, dan karena mereka memakan harta benda orang dengan jalan yang batil. Kami telah menyediakan untuk orang-orang yang kafir di antara mereka itu siksa yang pedih.
6	At Taubah 9. (34)	Hai orang-orang yang beriman, sesungguhnya sebahagian besar dari orang-orang alim Yahudi dan rahib-rahib Nasrani benar-benar memakan harta orang dengan jalan batil dan mereka menghalang-halangi (manusia) dari jalan Allah. Dan orang-orang yang menyimpan emas dan perak dan tidak menafkahkannya pada jalan Allah, maka beritahukanlah kepada mereka, (bahwa mereka akan mendapat) siksa yang pedih,
7	Al-Fajr 89. (19)	dan kamu memakan harta pusaka dengan cara mencampur baurkan (yang halal dan yang bathil),
8	Ali Imron 3. (161)	Tidak mungkin seorang nabi berkhianat dalam urusan harta rampasan perang. Barangsiapa yang berkhianat dalam urusan rampasan perang itu, maka pada hari kiamat ia akan datang membawa apa yang dikhianatkannya itu, kemudian tiap-tiap diri akan diberi pembalasan tentang apa yang ia kerjakan dengan (pembalasan) setimpal, sedang mereka tidak dianiaya.
9	Al Anfal 8. (1)	Mereka menanyakan kepadamu tentang (pembagian) harta rampasan perang. Katakanlah: "Harta rampasan perang kepunyaan Allah dan Rasul, oleh sebab itu bertakwalah kepada Allah dan perbaikilah perhubungan di antara sesamamu; dan taatlah kepada Allah dan Rasul-Nya jika kamu adalah orang-orang yang beriman

V. KESIMPULAN

Secara umum, uji coba *stemming* dengan menggunakan metode ECS pada *preprocessing* sudah berhasil dilakukan dengan baik dan akurat karena metode ECS pada prosesnya melakukan pengecekan pada kamus kata dasar. *Information retrieval* yang dibangun dengan pembentukan indeks korpus dapat memberikan hasil yang lebih presisi dan lebih relevan, dengan pembentukan indeks sinonim korpus kata-kata yang sebelumnya tidak terjaring dalam proses kueri pencarian dengan dibangunnya korpus sinonim sudah termasuk dari hasil pencarian hal ini dilihat dari ujicoba kueri yang diuji secara manual.

VI. DAFTAR PUSTAKA

[1] Agusta, Ledy. *Perbandingan Algortima Stemming Porter Dengan Algoritme Nazief & Adriani Untuk Stemming Dokumen Teks Bahasa Indonesia*. Universitas Kristen Satya Wacana. 2009.

[2] Akram Roshdi, Akram Roohparvar, 2015. *Review: Information Retrieval Techniques and Applications*, International Journal of Computer Networks and Communications Security, VOL. 3, NO. 9, SEPTEMBER 2015, 373–377

[3] Baeza R.Y., Neto R., 1999. Modern Information Retrieval, *Addison Wesley*-Pearson international edition, Boston. USA.

[4] Broto Poernomo T.P , Ir. Gunawan, 2015. *Sistem Information Retrieval Pencarian Kesamaan Ayat Terjemahan AlQur'an berbahasa Indonesia dengan Query Expansion dari tafsirnya* IDeaTech 2015, ISSN: 2089-1121

[5] Fatkhul Amin, 2012 *Sistem Temu Kembali Informasi dengan Metode Vector Space Model*, Jurnal Sistem Informasi Bisnis 02(2012)

[6] Jasman Pardede, Mira M Barmawi, Wildan D Pramono, 2013. *Implementasi Metode Generalized Vector Space Model Pada Aplikasi Information Retrieval,* No.1, Vol. 4, Januari – April 2013 ISSN : 2008-5266

[7] Manning, Christopher D., Prabhakar Raghavan, dkk, 2009. *Introduction to Information Retrieval*. Cambridge University Press, Cambridge, England.

[8] Nesdi E. Rozanda, Arif Marsal, Kiki Iswanti, 2014. *Rancang Bangun Sistem Informasi Hadist Menggunakan Teknik Temu Kembali Informasi Model Ruang Vektor, ejournal.uin-suska.ac.id.*

[9] Subari, Ferdinandus, 2017, *Sistem Information Retrieval Layanan Kesehatan Untuk Berobat Dengan Metode Vektor Space Model (VSM) Berbasis WebGis*, Snatika 2015, ISSN 2089-1083

[10] Surya Agustian, Imelda Sukma Wulandari, 2013. *Sistem Qur'an Retrieval Terjemahan Bahasa Indonesia berbasis Web dengan reorganisasi Korpus*, KNSI 2013, ISBN 978-602-17488-0.