SOURCE DETECTION PADA KASUS PLAGIARISME DOKUMEN MENGGUNAKAN METODE *BIWORD WINNOWING* DAN *RETRIEVAL* BERBASIS OKAPI BM25

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik
Pada Jurusan Teknik Informatika

Oleh

SYARIF HIDAYATULLAH

11051101864

FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU
2014

LEMBAR PERSETUJUAN

SOURCE DETECTION PADA KASUS PLAGIARISME DOKUMEN MENGGUNAKAN METODE *BIWORD WINNOWING* DAN *RETRIEVAL* BERBASIS OKAPI BM25

TUGAS AKHIR

Oleh

SYARIF HIDAYATULLAH
11051101864

Telah diperiksa dan disetujui sebagai laporan tugas akhir
di Pekanbaru, pada tanggal 18 Februari 2014

Pembimbing,

Surya Agustian, ST, M.Kom
NIP. 19760830 201101 1 003

LEMBAR PENGESAHAN

SOURCE DETECTION PADA KASUS PLAGIARISME DOKUMEN MENGGUNAKAN METODE *BIWORD WINNOWING* DAN *RETRIEVAL* BERBASIS OKAPI BM25

TUGAS AKHIR

Oleh

SYARIF HIDAYATULLAH

11051101864

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar sarjana Teknik Informatika
Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau
di Pekanbaru, pada tanggal 18 Februari 2014

Pekanbaru, 18 Februari 2014

Mengesahkan,

Dekan,

Dra. Hj. Yenita Morena, M.Si

NIP. 19601125 198503 2 002

Ketua Jurusan,

Elin Haerani, S.T. M. Kom

NIP. 19810523 200710 2 003

DEWAN PENGUJI

Ketua : Surya Agustian, S.T, M.Kom

Sekretaris : Surya Agustian, S.T, M.Kom

Penguji I : Elvia Budianita, S.T, M.Cs

Penguji II : Rahmad Abdillah, M.T

SOURCE DETECTION PADA KASUS PLAGIARISME DOKUMEN MENGGUNAKAN METODE *BIWORD WINNOWING* DAN *RETRIEVAL* BERBASIS OKAPI BM25

SYARIF HIDAYATULLAH
11051101864

Tanggal Sidang: 18 Februari 2014

Periode Wisuda: Juni 2014

Jurusan Teknik Informatika

Fakultas Sains dan Teknologi

Universitas Islam Negeri Sultan Syarif Kasim Riau

ABSTRAK

Tindakan plagiarisme merupakan salah satu dampak negatif kemajuan teknologi informasi dan komunikasi. Untuk mengatasi plagiarisme maka perlu adanya suatu aplikasi yang dapat mendeteksi plagiarisme. Saat ini sudah banyak penelitian yang membangun aplikasi pendeteksi plagiarisme, namun aplikasi tersebut hanya sebatas membandingkan kemiripan antara dua dokumen saja, tidak termasuk mendeteksi kemiripan dokumen dengan banyak dokumen dan mendeteksi dimana dokumen sumbernya. Pada penelitian ini akan dibangun sebuah aplikasi yang dapat mendeteksi dokumen sumber dari sebuah dokumen uji dan dapat membandingkan kemiripannya terhadap dokumen sumber yang diperoleh. Dalam mendeteksi dokumen sumber, isi dokumen uji akan dibentuk menjadi *query* dengan metode *fingerprint biword winnowing*. Q*uery* yang dibentuk terbagi menjadi dua jenis yakni *query* dengan *stemming* dan tanpa *stemming*. Setiap jenis query terbagi lagi menjadi 3 yakni *query* dengan n frekuensi *fingerprint* tertinggi, tengah dan terendah. Kemudian *query* tersebut akan dimasukkan kedalam sistem *information retrieval* (IR) Model Okapi BM25 untuk dicari dokumen sumbernya. Dokumen sumber yang diperoleh kemudian akan dibandingkan tingkat kemiripannya terhadap dokumen dengan menggunakan algoritma *biword winnowing*. Pada penelitian ini dilakukan pengujian hasil pembentukan *query* terhadap dokumen sumber yang diperoleh dan pengujian kemiripan dokumen uji dengan dokumen sumber yang diperoleh. *Output* pengujian hasil pembentukan *query* dimana n adalah 5, rekomendasi jenis *query* yang terbaik adalah *query stemming* dan tanpa stemming 5 frekuensi *fingerprint* tengah. sedangkan pengujian kemiripan dokumen disimpulkan bahwa dokumen sumber yang diperoleh benar terdapat kemiripan isi dengan dokumen uji dengan kemiripan 7,36-65,32%.

Kata Kunci: *biword winnowing*, *information retrieval*, *jaccard coefficient*, Okapi BM25, *query*.

SOURCE DETECTION OF PLAGIARISM CASE DOCUMENT USING BIWORD WINNOWING AND RETRIEVAL BASED ON OKAPI BM25 MODEL

SYARIF HIDAYATULLAH
11051101864

Date of Final Exam: February 18th, 2014
Graduation Ceremony Period: June 2014

Informatics Engineering Departement
Faculty of Sciences and Technology
State Islamic University of Sultan Syarif Kasim Riau

ABSTRACT

Plagiarism is one of the negative impact of information and communication technology advances. To overcome the plagiarism, an application that can detect plagiarism is needed. Currently, there are many studies that establish plagiarism detection application, but the application is limited only to compare the similarity between two documents, it is not including detecting similarity of a document with many documents and detect where the source documents. This research will build an application that can detect the source documents of a test document and be able to compare the similarity of the source documents were obtained. In detecting the source document, the contents of the test document will be formed into a query with Winnowing fingerprint biword method. Formed query is divided into two types, they are query by stemming and no stemming. Each type of query is divided into three namely queries with highest fingerprint frequency, middle and low. Then the query will be incorporated into the system of information retrieval (IR) models Okapi BM25 to look for the source document. Similarity level of source documents obtained will be compared to the document by using biword Winnowing algorithm. In this research, the test queries formation result to source document obtained and the similarity of test document to source document will be done. . Output test results query formation where n is 5, the best recommendation is query by stemming and query without stemming with 5 fingerprint middle frequency. While the similarity test concluded that the source documents obtained correctly has similarities with the contents of the test document from 7.36 to 65.32%.

***Keywords** : biword winnowing, information retrieval, jaccard coefficient, Okapi BM25, query.*

KATA PENGANTAR

Assalammu'alaikum wa rahmatullahi wa barakatuh.

Alhamdulillahi rabbil'alamin, tak henti-hentinya penulis ucapkan kehadirat Tuhan yang tiada Tuhan selain Dia, Allah SWT, yang dengan rahmat dan hidayahNya penulis mampu menyelesaikan Tugas Akhir ini dengan baik. Tidak lupa dan tak akan pernah lupa bershalawat kepada Nabi dan RasulNya, Muhammad SAW yang hanya menginginkan keimanan dan keselamatan bagi umatnya dan sangat belas kasihan lagi penyayang kepada orang-orang mukmin.

Tugas Akhir ini disusun sebagai salah satu syarat untuk mendapatkan gelar kesarjanaan pada jurusan Teknik Informatika Universitas Islam Negeri Sultan Syarif Kasim Riau. Banyak sekali pihak yang telah membantu penulis dalam penyusunan laporan ini, baik berupa bantuan materi ataupun berupa motivasi dan dukungan kepada penulis. Semua itu tentu terlalu banyak bagi penulis untuk membalasnya, namun pada kesempatan ini penulis hanya dapat mengucapkan terima kasih kepada :

1. ALLAH SWT yang karena hidayah dan rahmat nya, sehingga semua kebaikan mendekati saya dan saya bisa melakukan hal yang terbaik yang dapat saya lakukan.
2. Rasulullah SAW, seorang pimpinan umat islam sekaligus panutan umat islam yang tiada henti-hentinya berjuang demi menyebarkan agama islam.
3. Terima kasih kepada Kedua orang Tua Penulis, Ibu dan Bapak, serta adek yang tiada hentinya memanjatkan doa, memberikan dukungan dan semangat untuk kesuksesan penulis.
4. Bapak Prof. Dr. H. M. Nazir, selaku Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
5. Ibu Dra. Hj. Yenita Morena, M.Si, selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
6. Ibu Elin Haerani, S.T, M. Kom selaku Ketua Jurusan Teknik Informatika Fakultas Sains dan Teknologi UIN SUSKA RIAU.

7. Surya Agustian, S.T, M. Kom Selaku dosen pembimbing tugas akhir. Terimakasih pak untuk waktu yang selalu bapak luangkan untuk penulis, ilmu, semangat, dan motivasinya yang luar biasa. Terimakasih banyak pak.
8. Elvia Budianita, S.T,M.Cs, selaku dosen penguji 1 yang banyak membantu dan memberi masukan penulis dalam penyempurnaan Laporan Tugas Akhir ini, untuk ilmu-ilmunya.
9. Rahmad Abdillah, M.T, selaku dosen penguji 2, terimakasih pak untuk ilmu-ilmunya, saran-sarannya, perbaikan-perbaikannya, dan masukannya.
10. Muhammad Affandes, M.T, sebagai koordinator tugas akhir yang telah memberi masukan-masukan untuk penyelesaian tugas akhir ini, dan sangat sabar membantu penulis dalam mempersiapkan semua kebutuhan penulis dalam penyelesaian Tugas Akhir ini.
11. Bapak dan Ibu Dosen Jurusan Teknik Informatika Fakultas Sains dan Teknologi UIN SUSKA RIAU yang telah banyak memberikan ilmunya kepada penulis.
12. Teman-teman seperjuangan TIF B angkatan 10, Agung, Rizky, Denanda, Ardy, Alif, Rolli, Zul, Yessi, Fe, Nuraisyah dan yang lainnya yang tidak bisa disebutkan satu persatu. Tetap semangat dan tetap berjaya.

Penulis menyadari bahwa dalam penulisan laporan ini masih banyak kesalahan dan kekurangan, oleh karena itu kritik dan saran yang sifatnya membangun sangat penulis harapkan untuk kesempurnaan laporan ini. Akhirnya penulis berharap semoga laporan ini dapat memberikan sesuatu yang bermanfaat bagi siapa saja yang membacanya. Amin.

Wassalamu'alaikum wa rahmatullahi wa barakatuh

Pekanbaru, 18 Februari 2014

Penulis

DAFTAR ISI

DAFTAR GAMBAR

Gambar **Halaman**

DAFTAR TABEL

Tabel **Halaman**

DAFTAR RUMUS