Pengaruh Algoritma *Semantic Suffix Tree Clustering* Terhadap Tingkat Akurasi Sistem Tanya Jawab Bahasa Indonesia

Diajukan sebagai Syarat untuk
Menyelesaikan Pendidikan Program Strata-1 pada
Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Oleh:

Dininta Isnurthina
09021181320028

JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2018

LEMBAR PENGESAHAN TUGAS AKHIR

PENGARUH ALGORITMA SEMANTIC SUFFIX TREE CLUSTERING TERHADAP TINGKAT AKURASI SISTEM TANYA JAWAB BAHASA INDONESIA

Oleh :

DININTA ISNURTHINA

NIM : 09021181320028

Palembang, Januari 2018

Pembimbing I,

Ir. M. Ihsan Jambak, M.Sc.
NIP. 19680405201308201

Pembimbing II,

Novi Yusliani, S.Kom., M.T.
NIP. 198211082012122001

Mengetahui,

Ketua Jurusan Teknik Informatika

Rikie Primartha, M.T
NIP. 197706012009121004

TANDA LULUS SIDANG TUGAS AKHIR

Pada hari Jumat, 12 Januari 2018 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Dininta Isnurthina

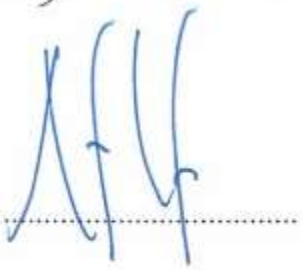
NIM : 09021181320028

Judul : Pengaruh Algoritma Semantic Suffix Tree Clustering Terhadap Tingkat Akurasi Sistem Tanya Jawab Bahasa Indonesia

1. Ketua Penguji

Ir. M. Ihsan Jambak, M.Sc.
NIP. 19680405201308201

2. Sekretaris Penguji



Novi Yusliani, S.Kom., M.T.
NIP. 198211082012122001

3. Penguji I

M. Fachrurrozi, S.Si, M.T
NIP. 198005222008121002

4. Penguji II

Rusdi Effendi, M.Kom.
NIP. 1671140201820005

Mengetahui,
Ketua Jurusan Teknik Informatika

Rikie Pnmartha, M.T
NIP. 197706012009121004

HALAMAN PERNYATAAN BEBAS PLAGIAT

Yang bertanda tangan di bawah ini :

Nama : Dininta Isnurthina
NIM : 09021181320028
Program Studi : Teknik Informatika
Judul Skripsi : Pengaruh Algoritma *Semantic Suffix Tree Clustering* Terhadap Tingkat Akurasi Sistem Tanya Jawab Bahasa Indonesia
Hasil Pengecekan Software *iThenticate/Turnitin* : 15 %

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.

Palembang, 02 Januari 2018

Dininta Isnurthina
NIM. 09021181320028

SIC • PARVIS • MAGNA

"Greatness from small beginnings"

- Sir Francis Drake (1540-1596)

PENGARUH ALGORITMA SEMANTIC SUFFIX TREE CLUSTERING TERHADAP TINGKAT AKURASI SISTEM TANYA JAWAB BAHASA INDONESIA

Oleh:
Dininta Isnurthina
09021181320028

ABSTRAK

Dalam Sistem Tanya Jawab Bahasa Indonesia untuk pertanyaan *factoid* dan *non-factoid* diuji perbandingan algoritma *Semantic Suffix Tree Clustering* (SSTC) dan *Suffix Tree Clustering* (STC) dalam mengelompokkan dokumen sumber jawaban. Hasil perbandingan menunjukkan bahwa tingkat akurasi Sistem Tanya Jawab setelah dokumen dikelompokkan dengan SSTC lebih rendah dibandingkan dengan setelah dikelompokkan dengan STC. Perbedaan tingkat akurasi tersebut ditemukan pada hampir semua kategori pertanyaan, kecuali kategori definisi. Secara rata-rata, tingkat akurasi yang mampu diraih Sistem Tanya Jawab Bahasa Indonesia dengan algoritma SSTC hanya sebesar 23,31%. Sedangkan tingkat akurasi Sistem Tanya Jawab dengan STC adalah sebesar 83%. Penurunan tingkat akurasi yang signifikan ini menunjukkan bahwa algoritma *Semantic Suffix Tree Clustering* tidak sesuai digunakan dalam konteks pengelompokan dokumen sumber jawaban pada Sistem Tanya Jawab Bahasa Indonesia.

Kata kunci: *factoid*, *non-factoid*, *semantic suffix tree clustering*, sistem tanya jawab, *suffix tree clustering*

Indralaya, 16 Januari 2018

Pembimbing I,

Ir. M. Ihsan Jambak, M.Sc.
NIP. 19680405201308201

Pembimbing II,

Novi Yusliani, S.Kom., M.T.
NIP. 198211082012122001

Mengetahui,
Ketua Jurusan Teknik Informatika

Rifkie Primartha, M.T
NIP. 197706012009121004

THE IMPACT OF SEMANTIC SUFFIX TREE CLUSTERING ALGORITHM ON THE ACCURACY RATE OF AN INDONESIAN QUESTION ANSWERING SYSTEM

By:
Dininta Isnurthina
09021181320028

ABSTRACT

This research analyzes the comparison between Semantic Suffix Tree Clustering (SSTC) algorithm and Suffix Tree Clustering (STC) algorithm in clustering documents on Indonesian Question Answering System. Comparison result shows that the accuracy rate of Indonesian Question Answering System after the documents is clustered by SSTC is lower than by STC. The accuracy rate degradation occurred in almost every question category, except definition category. In average, the accuracy rate obtained by Indonesian Question Answering System with SSTC is only 23,31%, while Indonesian Question Answering System with STC is able to obtain 83% accuracy rate. This significant difference shows that Semantic Suffix Tree Clustering algorithm is not suitable in the context of document clustering on Indonesian Question Answering System.

Keywords: factoid, non-factoid, question answering system, semantic suffix tree clustering, suffix tree clustering

Indralaya, 16 Januari 2018

Pembimbing I,

Ir. M. Ihsan Jambak, M.Sc.
NIP. 19680405201308201

Pembimbing II,

Novi Yusliani, S.Kom., M.T.
NIP. 198211082012122001

Mengetahui,
Ketua Jurusan Teknik Informatika

Rifkie Primartha, M.T
NIP. 197706012009121004

KATA PENGANTAR

Puji syukur kepada Allah swt. atas berkat dan rahmat-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas akhir ini disusun untuk memenuhi salah satu syarat guna menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini banyak pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun secara tidak langsung. Untuk itu Penulis ingin menyampaikan rasa terima kasih kepada:

1. Orang tuaku, Prof. Dr. Ir. Andy Mulyana, M.Sc. dan Ibu Dr. Ir. Lifianthi, M.Si., saudaraku, Indah Fitri Nurdianti dan Rahmadinda Nurfiana, kakekku Saad Nasuhim (alm) dan nenekku Farlina. Serta seluruh keluarga besar yang selalu mendokan serta memberikan dukungan baik moril maupun materil
2. Adik-adikku, Angga, Irdan, Fina, Dhila, Dhira, Defan, Davin, dan Adia yang menjadi sumber inspirasi bagi Penulis untuk meraih prestasi agar dapat membanggakan orang tua dan keluarga
3. Bapak Jaidan Jauhari, S.Pd., M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya
4. Bapak Rifkie Primartha, S.T., M.T. selaku Ketua Jurusan Teknik Informatika
5. Bapak Ir. M. Ihsan Jambak, M.Sc. selaku dosen pembimbing I dan Ibu Novi Yusliani, S.Kom., M.T. selaku dosen pembimbing II yang telah memberikan arahan serta dukungan dalam proses pengerjaan Tugas Akhir

6. Bapak M. Fachrurrozi, S.Si, M.T selaku dosen penguji I dan Bapak Rusdi Effendi, M.T selaku dosen penguji II yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir
7. Bapak Syamsuryadi, S.Si., M.Kom., Ph.D. selaku dosen pembimbing akademik
8. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya
9. Seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan
10. EC Technology Euphoria 2015, Novita Hidayati, S.Kom., Tiara Windri Apriani, S.Kom., Clara Fin Badillah, S.Kom., Choirunnisa Qanitah, S.Si., Priscillia Lupita, S.Si., Ningrum Kartika A., S.Si., dan Rahmatullah, S.Si., yang selalu mengerti, mendukung, dan mewarnai hidup Penulis
11. Anak-Anak Gaul (AAG), Nadia Kamila, Dwi Tiara Kurnila Sari, Amanda Farrah Merrynda, Selly Monica, Syifa Luthfia, Mitha Claudia Elsivia, dan Tiffany Putri Alamanda yang telah berteman dengan Penulis sejak SMA dan setia memberikan semangat
12. Tim PKPA, Latifah Alhaura, S.Kom., Clara Fin Badillah, S.Kom., Alvin Tamaarsa, S.Kom., Novita Hidayati, S.Kom., Muhammad Niudandri, S.Kom., dan Suwarto yang telah memotivasi dan berjuang bersama Penulis selama masa perkuliahan
13. Keluargaku BPH HMIF angkatan 2014 dan 2015, Programming Club, Ilkom Developer, Srivijaya, Kak Arief, Kak Ade, Kak Dian, Kak Daniel, Kak Satria, kak Agus, dan Latifah yang telah memberikan ruang bagi Penulis untuk berprestasi dan berkarya

14. Keluargaku, Himpunan Mahasiswa Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya, khususnya angkatan 2013 (INFINITY) yang menghabiskan waktu dan menorehkan kenangan bersama Penulis semasa kuliah
15. Semua pihak yang tidak bisa Penulis sebutkan satu persatu yang telah membantu dan berperan dalam Tugas Akhir ini.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya.

Akhir kata dengan segala kerendahan hati, semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Indralaya, Januari 2018

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis *cluster* atau *clustering* adalah teknik pengelompokan objek data hanya berbasis pada informasi yang ditemukan dalam data yang mendeskripsikan objek-objek serta relasi diantaranya (Tan, Steinbach, & Kumar, 2005). Tujuan yang ingin dicapai dari sebuah proses *clustering* adalah objek-objek yang ada dalam satu grup, saling berelasi atau mirip antara satu dan lainnya, tetapi berbeda dengan objek-objek dalam grup lain. Semakin tinggi tingkat kemiripan tiap objek dalam suatu grup dan semakin tinggi tingkat perbedaan tiap grup, maka dapat dinyatakan semakin baik hasil *clustering*-nya. Pada beberapa kasus, *clustering* hanya berperan sebagai titik awal yang dapat dimanfaatkan untuk berbagai tujuan, contohnya peringkasan data. *Clustering* sejak lama telah berperan penting dalam berbagai bidang: psikologi, biologi, statistik, pengenalan pola, *document retrieval*, *machine learning*, dan *data mining*.

Sistem Tanya Jawab adalah sebuah kajian dalam bidang Pemrosesan Bahasa Alami yang secara otomatis memberikan jawaban untuk pertanyaan berbahasa alami (Zulen & Purwarianti, 2011). Sistem Tanya Jawab dapat menggunakan basis data atau koleksi dokumen (lokal atau *web*) sebagai sumber untuk jawabannya. Sebuah sistem tanya jawab biasanya terdiri dari tiga komponen utama: *question analyzer*, *passage retriever*, dan *answer finder* (Harabagiu, Paşca, & Maiorano,

2000). Zulen dan Purwarianti (2011) mengemukakan bahwa komponen *question analyzer* bertujuan untuk mengklasifikasikan pertanyaan sesuai dengan *Expected Answer Type* (EAT) serta mengekstrak kata kunci dari pertanyaan tersebut. *Passage retriever* adalah komponen yang memiliki fungsi yang sama dengan mesin pencari, yaitu mengumpulkan informasi yang diinginkan. Pengumpulan informasi dibagi menjadi dua proses. Pertama, pengumpulan dokumen yang dilakukan oleh komponen *document retriever*, yang menggunakan *query* dari kata kunci hasil proses *question analyzer*. Kedua, pengumpulan paragraf yang dilakukan oleh komponen *passage retriever*. Komponen *answer finder* mencari jawaban dari kandidat dokumen atau paragraf yang telah ditemukan sebelumnya.

Penelitian mengenai Sistem Tanya Jawab Bahasa Indonesia telah banyak dilakukan, namun belum banyak yang mempertimbangkan penambahan fungsi *clustering* dokumen di dalamnya. Purwarianti, Tsuchiya, dan Nakagawa (2007) melakukan penelitian tentang Sistem Tanya Jawab Bahasa Indonesia untuk pertanyaan *factoid* yang menerapkan algoritma *Support Vector Machine* pada komponen *question analyzer* dan *answer finder*. Pada komponen *passage retriever* menggunakan *inverse document frequency* (idf). Hasil percobaan pada *passage retriever* menunjukkan bahwa penggunaan idf lebih cocok daripada *term frequency-inverse document frequency* (tf-idf). Purwarianti dan Yusliani (2012) melakukan penelitian tentang Sistem Tanya Jawab untuk pertanyaan *non-factoid* (definisi, alasan, dan metode) Bahasa Indonesia. Pada komponen *question analyzer* analisa pertanyaan didasarkan pada kata tanya yang digunakan. *Document retriever* meggunakan tf-idf dan perhitungan *cosine similarity*. *Answer finder* menggunakan

aturan *surface expression*. Dengan menggunakan 90 pertanyaan yang dikumpulkan dari 10 orang Indonesia dan 61 dokumen sumber, diperoleh nilai MRR 0.7689, 0.5925, dan 0.5704 untuk tipe pertanyaan definisi, alasan, dan metode secara berurutan. Algoritma *Suffix Tree Clustering* yang dikembangkan oleh Zamir dan Etzioni (1998) diimplementasikan oleh Rahmansyah (2015) pada Sistem Tanya Jawab Bahasa Indonesia untuk pertanyaan *factoid* dan *non-factoid* dengan menerapkan fungsi *clustering* dalam *document retriever*. Penelitian ini menghasilkan tingkat akurasi jawaban yang diberikan sebesar 83%.

Suffix Tree Clustering adalah algoritma pengelompokan informasi yang didasarkan pada frase-frase yang sering muncul pada sekumpulan dokumen. *Suffix Tree Clustering* terbukti sangat efisien namun memiliki kekurangan (Janruang & Guha, 2011). Berbagai pendekatan telah dilakukan untuk memecahkan masalah dari algoritma *Suffix Tree Clustering*. Algoritma NSTC digunakan dengan *vector space model* untuk menghitung kemiripan pasangan dokumen sebagai solusi dari *Suffix Tree Clustering* yang tingkat akurasinya rendah dalam mengembalikan *cluster* jumlah besar. *Suffix Tree Clustering* dengan teknik n-gram digunakan untuk menangani masalah terlalu panjangnya sebuah *suffix tree* yang dihasilkan. Namun kesulitannya, teknik n-gram membangkitkan label *cluster* yang terpotong apabila ukuran frase umumnya lebih besar. Untuk mengurangi penggunaan memori, dikembangkan *Suffix Tree Clustering* dengan teknik χ-gram untuk membentuk *suffix tree*-nya. Namun, *Suffix Tree Clustering* dengan teknik χ-gram tidak dapat menghasilkan *cluster* yang bersifat semantik. Walapun *Suffix Tree Clustering* memiliki performa yang baik, namun tidak ditambahkannya kata yang memiliki

kemiripan semantik untuk membentuk *suffix tree*, alhasil *tree* yang dibangkitkan menjadi besar dan kompleks sehingga sulit untuk diolah.

Tetapi apabila kata yang memiliki kemiripan semantik dikombinasikan dengan pencocokan *string* untuk membentuk *suffix tree*, jumlah simpulnya akan berkurang sehingga *tree* yang dibangkitkan mengecil. Janruang dan Guha (2011) mengembangkan algoritma *Semantic Suffix Tree Clustering* yang memanfaatkan hasil kemiripan semantik menggunakan sinonim kata yang diperoleh dari basis data WordNet dan dihitung ukuran kemiripan antar pasangan katanya. Sebuah algoritma *clustering* teks harus memanfaatkan makna dari kata-kata yang ada untuk meningkatkan tingkat akurasi *clustering*-nya. Konsep dasar algoritma *Semantic Suffix Tree Clustering* adalah mengelompokkan dokumen-dokumen yang mirip secara semantik ke dalam *cluster* yang sama. Hasil penelitian Janruang dan Guha (2011) menunjukkan nilai presisi *clustering* menggunakan *Semantic Suffix Tree Clustering* sebesar 0.81, sedangkan menggunakan *Suffix Tree Clustering* sebesar 0.68 dengan *dataset* berasal dari 26.890 hasil pencarian dari 10 queri dalam Dmoz.com.

Penelitian ini akan menguji apakah algoritma *Semantic Suffix Tree Clustering* dapat meningkatkan hasil *clustering* pada komponen *document retriever* sehingga menghasilkan Sistem Tanya Jawab yang akurat.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah apa pengaruh *Semantic Suffix Tree Clustering* terhadap tingkat akurasi sebuah Sistem Tanya Jawab Bahasa Indonesia jika diterapkan pada komponen *document clusterer*. Untuk menjawab rumusan masalah tersebut, diuraikan beberapa *research question* sebagai berikut:

1. Bagaimana cara kerja *Semantic Suffix Tree Clustering* dalam mengelompokkan dokumen pada Sistem Tanya Jawab Bahasa Indonesia?
2. Apakah penambahan komponen *document clusterer* memberikan pengaruh terhadap tingkat akurasi Sistem Tanya Jawab Bahasa Indonesia?
3. Adakah faktor-faktor yang mempengaruhi performa *Semantic Suffix Tree Clustering* ketika melakukan pengelompokan dokumen pada Sistem Tanya Jawab Bahasa Indonesia?
4. Bagaimana cara menguji pengaruh *Semantic Suffix Tree Clustering* terhadap tingkat akurasi Sistem Tanya Jawab Bahasa Indonesia?

1.3 Tujuan Penelitian

Tujuan dilakukannya penelitan ini adalah sebagai berikut:

1. Mengetahui cara kerja algoritma *Semantic Suffix Tree Clustering* dalam mengelompokkan dokumen pada Sistem Tanya Jawab Bahasa Indonesia.
2. Melihat pengaruh penambahan komponen *document clusterer* terhadap tingkat akurasi Sistem Tanya Jawab Bahasa Indonesia.

3. Mengetahui faktor-faktor yang mempengaruhi algoritma *Semantic Suffix Tree Clustering* ketika melakukan pengelompokan dokumen pada Sistem Tanya Jawab Bahasa Indonesia.
4. Mengetahui cara menguji pengaruh *Semantic Suffix Tree Clustering* terhadap tingkat akurasi Sistem Tanya Jawab Bahasa Indonesia.

1.4 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut

1. Menambah pengetahuan tentang algoritma *Semantic Sufix Tree Clustering* beserta keunggulan dan cara kerjanya.
2. Menghasilkan sebuah Sistem Tanya Jawab Bahasa Indonesia yang hasil *clustering* dokumennya diperoleh menggunakan *Semantic Suffix Tree Clustering*.
3. Bila ditemukan faktor-faktor yang mempengaruhi algoritma *Semantic Suffix Tree Clustering* dalam mengelompokkan dokumen pada Sistem Tanya Jawab Bahasa Indonesia, maka penelitian ini dapat menjadi referensi untuk penelitian terhadap faktor-faktor tersebut dalam rangka peningkatan performa *Semantic Suffix Tree Clustering*.
4. Menjadi referensi dalam pemilihan algoritma *Semantic Suffix Tree Clustering* sebagai solusi permasalahan *clustering* pada bidang yang lain.

1.5 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Algoritma *clustering* yang diteliti adalah *Semantic Suffix Tree Clustering*.
2. Komponen pada Sistem Tanya Jawab Bahasa Indonesia yang menjadi fokus penelitian ini hanya *document clusterer*.
3. Jenis pertanyaan yang akan menjadi data uji adalah pertanyaan-pertanyaan yang menanyakan organisasi, lokasi, tanggal/waktu, nama, orang, jumlah, definisi, alasan, dan metode/cara. Pertanyaan diawali dengan kata tanya berupa "apa" dan "apakah" untuk organisasi, nama dan definisi, "dimana" dan "dimanakah" untuk lokasi, "kapan" dan "kapankah" untuk tanggal/waktu, "siapa" dan "siapakah" untuk orang, "berapa" dan "berapakah" untuk jumlah, "mengapa" dan "kenapa" untuk alasan, serta "bagaimana" dan "bagaimanakah" untuk metode/cara.
4. Pengujian akan dilakukan dengan melihat persentase keakuratan hasil pencarian jawaban dan waktu proses dari Sistem Tanya Jawab dengan *Semantic Suffx Tree Clustering*, kemudian membandingkannya dengan hasil dan waktu proses dari *Suffix Tree Clustering*.

DAFTAR PUSTAKA

Alam, M., & Sadaf, K. (2013). A review on clustering of web search result *Advances in Computing and Information Technology* (pp. 153-159): Springer.

Arifin, A. Z., Darwanto, R., Navastara, D. A., & Ciptaningtyas, H. T. (2008). *Klasifikasi Online Dokumen Berita Dengan Menggunakan Algoritma Suffix Tree Clustering*. Paper presented at the Seminar Sistem Informasi Indonesia (SESINDO2008).

Harabagiu, S. M., Paşca, M. A., & Maiorano, S. J. (2000). *Experiments with open-domain textual question answering*. Paper presented at the Proceedings of the 18th conference on Computational linguistics-Volume 1.

Janruang, J., & Guha, S. (2011). *Semantic suffix tree clustering*. Paper presented at the First IRAST International Conference on Data Engineering and Internet Technology, DEIT.

Lin, J., & Katz, B. (2006). Building a reusable test collection for question answering. *Journal of the Association for Information Science and Technology, 57*(7), 851-861.

Nazief, B. A., & Adriani, M. (1996). Confix Stripping: Approach to Stemming Algorithm for Bahasa Indonesia. *Internal publication, Faculty of Computer Science, University of Indonesia, Depok, Jakarta, 41*.

Perera, R. (2012). *Ipedagogy: Question answering system based on web information clustering*. Paper presented at the Technology for Education (T4E), 2012 IEEE Fourth International Conference on.

Purwarianti, A., Tsuchiya, M., & Nakagawa, S. (2007). *A machine learning approach for indonesian question answering system*. Paper presented at the Artificial Intelligence and Applications.

Purwarianti, A., & Yusliani, N. (2012). SISTEM QUESTION ANSWERING BAHASA INDONESIA UNTUK PERTANYAAN NON-FACTOID. *Jurnal Ilmu Komputer dan Informasi, 4*(1), 10-14.

Rahmansyah, A. (2015). *IMPLEMENTASI SUFFIX TREE CLUSTERING PADA SISTEM TANYA JAWAB BAHASA INDONESIA UNTUK PERTANYAAN FACTOID DAN NON-FACTOID*. (S-1), Universitas Sriwijaya, Indralaya.

Sameh, A., & Kadray, A. (2010). Semantic web search results clustering using lingo and wordnet. *International Journal of Research and Reviews in Computer Science (IJRRCS), 1*(2), 71-76.

Shabbir, U., Kanwal, T., Malik, R., Khalid, S., & Khan, A. A. (2015). *Comparison between SSTC and LINGO Algorithms in Clustered Based Semantic Search for Browsing Scholarships.* Paper presented at the 2015 13th International Conference on Frontiers of Information Technology (FIT).

Tan, P.-N., Steinbach, M., & Kumar, V. (2005). *Introduction to Data Mining, (First Edition)*: Addison-Wesley Longman Publishing Co., Inc.

Wen, H., Xiao, N.-F., & Chen, Q. (2009). *Web snippets clustering based on an improved suffix tree algorithm*. Paper presented at the Fuzzy Systems and Knowledge Discovery, 2009. FSKD'09. Sixth International Conference on.

Yusliani, N. (2010). Sistem Tanya Jawab Bahasa Indonesia untuk Non Factoid Question. *Master, Program Studi Informatika, Institut Teknologi Bandung, Bandung*.

Zamir, O., & Etzioni, O. (1998). *Web document clustering: A feasibility demonstration.* Paper presented at the Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval.

Zulen, A. A., & Purwarianti, A. (2011). *Study and Implementation of Monolingual Approach on Indonesian Question Answering for Factoid and Non-Factoid Question.* Paper presented at the PACLIC.