

**EKSTRAKSI SITASI KUTIPAN BERBASIS ATURAN DAN
REGULAR EXPRESSION**

Skripsi



oleh
ELVIN ANDREAN
71130054

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA YOGYAKARTA

2017

Ekstraksi Sitasi Kutipan Berbasis Aturan dan Regular Expression

Skripsi



Diajukan kepada Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

ELVIN ANDREAN

71130054

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA

2017

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

EKSTRAKSI SITASI KUTIPAN BERBASIS ATURAN DAN REGULAR EXPRESSION

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 6 Juni 2017



Elvin Andrian

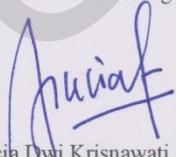
71130054

HALAMAN PERSETUJUAN

Judul Skripsi : EKSTRAKSI SITASI KUTIPAN BERBASIS ATURAN
DAN REGULAR EXPRESSION
Nama Mahasiswa : ELVIN ANDREAN
NIM : 71130054
Matakuliah : Skripsi (Tugas Akhir)
Kode : TIW276
Semester : Genap
Tahun akademik : 2016/2017

Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 6 Juni 2017

Dosen Pembimbing I


Lucia Dwi Krisnawati, Dr.

Dosen Pembimbing II


Willy Sudiarto Raharjo, S.Kom.,M.Cs.

HALAMAN PENGESAHAN

EKSTRAKSI SITASI KUTIPAN BERBASIS ATURAN DAN REGULAR EXPRESSION

Oleh: ELVIN ANDREAN / 71130054

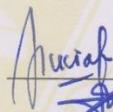
Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 31 Mei 2017

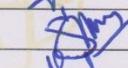
Pada tanggal 6 Juni 2017

Mengesahkan,

Dewan Penguji:

1. Lucia Dwi Krisnawati, Dr.
2. Willy Sudiarto Raharjo, S.Kom.,M.Cs.
3. Hendro Setiadi, ST, MM, MEngSc.
4. Laurentius Kuncoro Probo Saputra, S.T.,M.Eng.







Dekan

Ketua Program Studi


(Budi Susanto, S.Kom., M.T.)


(Gloria Virginia, Ph.D.)

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa karena atas rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan tugas akhir yang berjudul “EKSTRAKSI SITASI KUTIPAN BERBASIS ATURAN DAN *REGULAR EXPRESSION*”.

Dengan selesainya tugas akhir ini tidak lepas dari bantuan banyak pihak yang telah memberikan bantuan dan masukan-masukan kepada penulis. Untuk itu penulis mengucapkan banyak terimakasih.

Penulis menyadari bahwa laporan tugas akhir ini masih jauh dari kesempurnaan baik dari bentuk penyusunan maupun materinya. Oleh karena itu segala kritikan dan saran yang membangun akan penulis terima dengan baik. Akhir kata semoga laporan tugas akhir ini dapat memberikan manfaat kepada kita sekalian.

Yogyakarta, 8 Mei 2017

Penulis

UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa atas melimpahnya berkat dan anugerah-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul ” Ekstraksi Sitasi Kutipan Berbasis Aturan dan *Regular Expression*” dengan lancar dan tepat waktu.

Penulisan skripsi ini dimaksudkan untuk memenuhi salah satu syarat mencapai gelar Sarjana Komputer pada Fakultas Teknologi Informasi Jurusan Teknik Informatika Universitas Kristen Duta Wacana Yogyakarta. Selama pembuatan skripsi ini penulis juga mendapat bantuan dari berbagai pihak , maka pada kesempatan ini penulis menyampaikan penghargaan dan terima kasih kepada:

1. Tuhan Yesus Kristus yang selalu memberikan rahmat dan penyertaan-Nya selama menempuh kuliah hingga pembuatan skripsi, sehingga penulis mampu menyelesaikan dan mengerjakan skripsi ini dengan baik.
2. Ibu **Dr. Lucia Dwi Krisnawati**, selaku dosen pembimbing I dan Bapak **Willy Sudiarto Raharjo, S.Kom, M.Cs.** selaku dosen pembimbing II yang telah membimbing dan memberikan arahan serta masukan kepada penulis selama proses pembuatan skripsi dari awal hingga selesai.
3. Para dosen penguji yang telah memberikan arahan dan masukan kepada penulis selama pengujian skripsi.
4. Bapak dan Ibu dosen Fakultas Teknologi Informasi Universitas Kristen Duta Wacana Yogyakarta yang telah berjasa membagi ilmu dan mendidik penulis.
5. Fakultas Teknologi Informasi Jurusan Teknik Informatika Universitas Kristen Duta Wacana Yogyakarta sebagai tempat penulis menimba Ilmu.
6. Para staf admin Fakultas Teknologi Informasi Jurusan Teknik Informatika Universitas Kristen Duta Wacana Yogyakarta yang turut membantu penulis memperoleh informasi.
7. Orang tua yang telah banyak membantu dalam doa dan memberi semangat dan dukungan secara langsung.
8. Teman-teman TI angkatan 2013 yang telah berjuang bersama selama 4 tahun di kampus.

9. Teman-teman yang tidak bisa dipersebutkan satu persatu yang telah selalu memberi semangat dan dukungan saat mengerjakan skripsi.

Penulis menyadari banyak kekurangan dalam penyusunan skripsi ini, oleh karena itu sangat diharapkan masukan dari pembaca baik berupa saran maupun kritik. Semoga skripsi ini berguna dan bermanfaat bagi semua pihak.

Yogyakarta, 8 Mei 2017

Penulis

©UKDWN

INTISARI

ANALITIK TEKS UNTUK PENCARIAN SITASI KUTIPAN MENGUNAKAN REGULAR EXPRESSION

Sitasi menjadi salah satu hal yang cukup penting di dunia penulisan akademis. Sitasi adalah cara menunjukkan kepada pembaca bahwa bahan yang ada pada artikel berasal dari sumber orang lain. Sitasi juga memberikan informasi kepada pembaca tentang pengarang, judul, nama dan lokasi tempat pembulikasi, tanggal dan nomor halaman dari pemilik asli. Sitasi juga dapat digunakan untuk beberapa bidang pakar ilmu komputer seperti *Text Reuse Plagiarism Detection*, *Bibliographic Coupling*, dan menjadi salah satu komponen penilaian peringkat universitas dan artikel. Penelitian ini melakukan implementasi *Regular Expression* untuk mencari dan mengekstrak sitasi dari artikel yang menggunakan gaya penulisan *American Psychological Association* (APA) dan *Association for Computing Machinery* (ACM) yang berekstensi PDF dan DOCX. Sistem ini dibangun menggunakan bahasa *Hypertext Processor* (PHP), antarmuka dibuat menggunakan HTML5 dan bootstrap. Pengujian dilakukan menggunakan *precision*, *recall* dan F1 dengan sedikit memodifikasi perhitungan menggunakan 2 level granularitas yaitu uji kalimat dan uji unit. Hasil yang didapat untuk *precision*, *recall*, dan F1 dari 38 dokumen yakni di atas 75%.

Kata kunci : Sitasi, plagiarisme, *bibliographic coupling*, APA, ACM, *regular expression*, *precision* dan *recall*.

DAFTAR PUSTAKA

HALAMAN SAMBUNG DEPAN	
HALAMAN SAMBUNG DALAM	
PERNYATAAN KEASLIAN SKRIPSI.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
KATA PENGANTAR	vi
UCAPAN TERIMA KASIH.....	vii
INTISARI.....	ix
DAFTAR PUSTAKA	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN.....	xv
BAB I. PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah.....	2
1.4. Tujuan Penelitian	3
1.5. Metodologi Penelitian	3
1.6. Sistematika Penulisan	3
BAB II. TINJAUAN PUSTAKA DAN LANDASAN TEORI	5
2.1. Tinjauan Pustaka	5
2.2. Landasan Teori.....	6
2.2.1. Ekstraksi Informasi	6
2.2.2. Gaya Penulisan.....	7
2.2.2.1 Gaya Penulisan APA.....	7
2.2.2.2 Gaya Penulisan ACM.....	11
2.2.3. Regular Expressions.....	13
2.2.4. Precision (Ketepatan), Recall (Perolehan) dan F1	17
BAB III. ANALISIS DAN PERANCANGAN SISTEM	20

3.1. Analisis Kebutuhan	20
3.1.1. Perangkat Keras	20
3.1.2. Perangkat Lunak.....	20
3.1.3. Analisis Data	21
3.2. Rancangan Kerja Sistem	21
3.2.1. Use Case.....	21
3.3. Perancangan User Interface.....	24
3.4. Mekanisme Pengujian	28
BAB IV. IMPLEMENTASI DAN ANALISIS SISTEM	29
4.1. Implementasi Sistem	29
4.1.1. Pra-Pemrosesan	29
4.1.1.1 Konversi Berkas Format	29
4.1.1.2 Normalisasi Teks.....	29
4.1.2. Pembentukan Regular Expressions	30
4.1.2.1 Segmentasi Daftar Pustaka Untuk APA.....	30
4.1.2.2 Pencarian Label Terbesar ACM.....	32
4.1.2.3 Ekstraksi Sitasi	33
4.1.2.3.1 APA.....	33
4.1.2.3.2 ACM.....	36
4.1.2.4 Ekstraksi Kutipan	39
4.1.3. Antarmuka.....	39
4.1.3.1 Tampilan Halaman Utama.....	39
4.1.3.2 Halaman Sitasi	40
4.1.3.3 Halaman Tentang.....	41
4.1.3.3.1 Tentang Penulis	41
4.1.3.3.2 Tentang Aplikasi.....	41
4.1.3.4 Halaman Kontak	41
4.1.3.5 Tampilan Browse.....	42
4.1.3.6 Tampilan Kirim File	42
4.1.3.7 Halaman Hasil ACM	43

4.1.3.8 Halaman Hasil APA	44
4.2. Uji Validitas Sistem	44
4.2.1. Tabel Uji Kalimat.....	45
4.2.2. Tabel Uji Unit	48
4.3. Analisis Sistem.....	50
BAB V. KESIMPULAN DAN SARAN	53
5.1. Kesimpulan	53
5.2. Saran.....	53
REFERENSI	55

©UKDWN

DAFTAR TABEL

2.1 Tabel Penulisan sitasi berdasarkan jumlah pengarang.....	8
2.2 Tabel Contoh pembuatan pola string	13
2.3 Tabel Penggunaan operator “[]”	13
2.4 Tabel Penggunaan operator “-”	14
2.5 Tabel Penggunaan operator “^”	14
2.6 Tabel Penggunaan operator “?”	14
2.7 Tabel Operator lain	14
2.8 Tabel Operator modifiers	16
2.9 Tabel Operator assertion dan Operator modifiers.....	17
2.10 Tabel Matrik Recall dan Precision.....	18
4.1 Tabel Tiga Penulis.....	37
4.2 Tabel Contoh Lima Sitasi	37
4.3 Tabel Contoh pemecahan string “[2]”	38
4.4 Tabel Penjelasan atribut relevan dan lainnya.....	45
4.5 Tabel Halaman hasil sistasi APA ekstensi DOCX.....	46
4.6 Tabel Halaman hasil sitasi APA ekstensi PDF	46
4.7 Tabel Halaman hasil sitasi ACM ekstensi PDF.....	47
4.8 Tabel Halaman hasil kutipan APA ekstensi DOCX	48
4.9 Tabel Halaman hasil kutipan APA ekstensi PDF	48
4.10 Tabel Halaman hasil kutipan ACM ekstensi PDF	49
4.11 Tabel Uji Kalimat.....	50
4.12 Tabel Uji Unit	50

DAFTAR GAMBAR

3.1 Gambar Use Case Diagram User	22
3.2 Gambar Diagram Pencarian APA	23
3.3 Gambar Diagram Pencarian ACM	24
3.4 Gambar Tampilan Awal Sistem	25
3.5 Gambar Tampilan Menu Sitasi	25
3.6 Gambar Tampilan Menu Tentang Penulis	26
3.7 Gambar Tampilan Menu Tentang Aplikasi.....	26
3.8 Gambar Tampilan Kontak.....	26
3.9 Gambar Tampilan Output Akhir Gaya ACM	27
3.10 Gambar Tampilan Output Akhir Gaya APA.....	27
4.1 Gambar Pembagian Konten dan Daftar Pustaka.....	31
4.2 Gambar Proses Pengambilan Nama Belakang Pengarang	32
4.3 Gambar Pengambilan Label Nomor Terbesar.....	33
4.4 Gambar Pengecekan 2 Pengarang dalam 1 Sitasi	36
4.5 Gambar Tampilan Halaman Utama	40
4.6 Gambar Halaman Sitasi.....	40
4.7 Gambar Tentang Penulis.....	41
4.8 Gambar Tentang Aplikasi	41
4.9 Gambar Tampilan Menu Kontak	42
4.10 Gambar Tampilan Browse	42
4.11 Gambar Tampilan Kirim File.....	43
4.12 Gambar Halaman Hasil ACM.....	43
4.13 Gambar Halaman Hasil APA	44

DAFTAR LAMPIRAN

Source Code Regex.php dan Index.html	1
Kartu Konsultasi	40
Form Revisi Skripsi	42

©UKDWN

INTISARI

ANALITIK TEKS UNTUK PENCARIAN SITASI KUTIPAN MENGUNAKAN REGULAR EXPRESSION

Sitasi menjadi salah satu hal yang cukup penting di dunia penulisan akademis. Sitasi adalah cara menunjukkan kepada pembaca bahwa bahan yang ada pada artikel berasal dari sumber orang lain. Sitasi juga memberikan informasi kepada pembaca tentang pengarang, judul, nama dan lokasi tempat pembulikasi, tanggal dan nomor halaman dari pemilik asli. Sitasi juga dapat digunakan untuk beberapa bidang pakar ilmu komputer seperti *Text Reuse Plagiarism Detection*, *Bibliographic Coupling*, dan menjadi salah satu komponen penilaian peringkat universitas dan artikel. Penelitian ini melakukan implementasi *Regular Expression* untuk mencari dan mengekstrak sitasi dari artikel yang menggunakan gaya penulisan *American Psychological Association* (APA) dan *Association for Computing Machinery* (ACM) yang berekstensi PDF dan DOCX. Sistem ini dibangun menggunakan bahasa *Hypertext Processor* (PHP), antarmuka dibuat menggunakan HTML5 dan bootstrap. Pengujian dilakukan menggunakan *precision*, *recall* dan F1 dengan sedikit memodifikasi perhitungan menggunakan 2 level granularitas yaitu uji kalimat dan uji unit. Hasil yang didapat untuk *precision*, *recall*, dan F1 dari 38 dokumen yakni di atas 75%.

Kata kunci : Sitasi, plagiarisme, *bibliographic coupling*, APA, ACM, *regular expression*, *precision* dan *recall*.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Ketersediaan informasi yang melimpah di internet membuat penggunanya dengan mudah mendapatkan informasi yang diinginkan. Berbagai macam data dan informasi yang tersedia membuat pengguna lebih memanfaatkan internet dibandingkan dengan mencari data secara manual seperti ke perpustakaan atau membeli sebuah buku. Padahal, informasi yang tersedia di internet belum tentu benar karena tergantung pada website penyedia informasi tersebut apakah kredibilitasnya bisa dipercaya atau tidak. Ketersediaan informasi yang melimpah ini menimbulkan hal positif dan negatif dalam penggunaannya. Banyak website ataupun blog yang mengabaikan tentang *copyright violation article*, sehingga masih banyak web ataupun blog yang artikelnya memuat isi artikel milik orang lain alias *copy-paste* informasi milik orang lain. Secara hukum, hal ini menyalahi aturan dan dianggap sebagai plagiarisme, kemudian hal ini juga menyebabkan redundansi data.

Selain teks, salah satu indikasi plagiarisme adalah sitasi (Gipp & Msuschke, 2011). Sitasi cukup penting dalam penulisan sebuah artikel karena dapat berguna untuk menghindari sebuah artikel dari plagiarisme dan menunjukkan kepada pembaca bahwa informasi yang ditampilkan berasal dari sumber penelitian yang benar ("*Why citting is important?*", 2016). Selain sebagai indikasi plagiarisme, sitasi juga berfungsi untuk *bibliographic coupling* (Garfield, 2001). Ekstraksi informasi dan *regular expression* dapat menjadi salah satu alternatif dalam memecahkan masalah plagiarisme dan *bibliographic coupling*. Ekstraksi informasi adalah sebuah teknologi pengolahan bahasa alami (*Natural Language Processing - NLP*) yang berfungsi untuk memproses data tidak terstruktur, teks bahasa natural, untuk menemukan potongan informasi yang spesifik, atau fakta di dalam sebuah teks (Yangarber, 2001). *Regular expression* merupakan salah satu metode yang dapat diterapkan dalam teknologi ekstraksi informasi.

Berdasarkan semua latar belakang tersebut, maka penelitian ini akan membangun sebuah sistem ekstraksi informasi sitasi kutipan dari artikel yang berekstensi pdf atau docx menggunakan *regular expression*.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana cara membedakan antara sitasi dan informasi biasa?
Dari sini masalah yang penulis hadapi adalah bagaimana membuat sistem yang mampu membedakan antara sitasi (*Johnson, 2000*) dengan informasi biasa seperti singkatan KTP (kartu tanda penduduk).
2. Bagaimana cara mendefinisikan pola kedalam *regular expressions*?
Karena format gaya penulisan sitasi ada banyak, dalam kasus ini adalah *American Psychological Association (APA)* dan *Association for Computing Machinery (ACM)*, maka masalah yang penulis hadapi adalah cara menerapkan pola yang sesuai dengan format gaya yang berbeda-beda tersebut.
3. Bagaimana cara mendapatkan kutipan berdasarkan hasil sitasi yang dicari?

1.3 Batasan Masalah

Adapun batasan masalah yang akan dibuat dalam penelitian ini adalah sebagai berikut:

1. Artikel yang diekstrak menggunakan gaya penulisan dari APA dan ACM.
2. Artikel yang diekstrak menggunakan bahasa Indonesia.
3. Sitasi yang dicetak memiliki format nama pengarang dan tahun untuk gaya penulisan APA.
4. Konverter pdf ke teks menggunakan *library* yang diambil dari <http://pdfparser.org> dan konverter docx ke teks diambil dari https://www.phpclasses.org/package/8908-PHP-Convert-DOCX-DOC-PDF-to-plain-text.html#view_files. Template HTML didapat dari www.startbootstap.com

1.4 Tujuan Penelitian

Melalui penelitian ini, maka tujuan yang ingin dicapai oleh penulis adalah menghasilkan sistem yang dapat mengekstrak informasi sitasi kutipan sebagai sumber referensi secara otomatis.

1.5 Metodologi Penelitian

Metode yang digunakan dalam penelitian ini adalah:

1. Studi literatur

Pada tahap ini peneliti akan mengumpulkan berbagai literatur/pustaka yang berkaitan dengan metode yang berkaitan dengan ekstraksi informasi untuk sitasi kutipan dan *regular expression*.

2. Pengumpulan data

Peneliti akan melakukan pengumpulan data yang diperlukan untuk pembangunan sistem. Data yang akan dikumpulkan berupa artikel yang berbahasa Indonesia. Untuk gaya penulisan APA, penulis mengambil data dari jurnal informatika UKDW, kemudian untuk gaya penulisan ACM, penulis mengambil data dari situs online www.e-journal.com.

3. Pembangunan Sistem

Sistem ini akan dibuat menggunakan bahasa pemrograman php dan menggunakan *regular expression*.

4. Implementasi dan Testing

Tahap implementasi dan testing berisi proses pengujian terhadap program ekstraksi informasi sitasi kutipan dengan menampilkan hasil sitasi kutipan yang ada.

5. Analisis Hasil Percobaan dan Evaluasi

Tahap analisis dan evaluasi adalah tahap yang dilakukan untuk menganalisis dan mengevaluasi program yang telah dibangun.

1.6 Sistematika Penulisan

Sistematika penulisan yang digunakan penulis untuk mengerjakan laporan tugas akhir ini sesuai dengan ketentuan dari buku panduan penulisan tugas akhir.

Penulisan akan dibagi menjadi 5 bab, yaitu: Pendahuluan, Tinjauan Pustaka dan Landasan Teori, Analisis dan Perancangan Sistem, Implementasi dan Analisis Sistem, Kesimpulan dan Saran.

Pada bab pertama akan dibahas mengenai latar belakang masalah yang ada, fungsi sitasi, metode yang digunakan dalam ekstraksi informasi, dan contoh penelitian yang pernah dilakukan.

Pada bab kedua akan menjelaskan mengenai tinjauan pustaka dan landasan teori. Pada tinjauan pustaka akan berisi penelitian-penelitian yang sudah pernah dilakukan oleh orang lain berkaitan dengan penelitian yang akan dilakukan oleh penulis. Hasil dari penelitian-penelitian yang sudah pernah ada tersebut nantinya akan digunakan sebagai referensi dalam penelitian yang akan dilakukan oleh penulis. Sedangkan landasan teori akan berisi tentang dasar-dasar teori yang akan digunakan pada penelitian ini, termasuk metode-metode dan rumus-rumus yang akan digunakan.

Pada bab ketiga penulis membahas mengenai analisis dan perancangan sistem yang akan diteliti dan dibuat. Bab ini akan menjelaskan tentang use case, desain dan lain sebagainya untuk membangun sistem.

Pada bab keempat penulis akan membahas mengenai hasil implementasi yang telah dilakukan, antarmuka sistem, hasil uji ketepatan dan perolehan sistem.

Pada bab kelima penulis akan membahas mengenai kesimpulan dari hasil penelitian dan saran. Kesimpulan berisi tentang hasil yang telah didapatkan selama penelitian berlangsung. Sedangkan saran dapat berupa sesuatu hal yang dapat dilakukan berkaitan dengan pengembangan topik penelitian tersebut di masa mendatang serta metode yang dapat digunakan.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan maka dapat disimpulkan:

- a. Untuk pengujian kalimat, ekstraksi yang memiliki hasil paling tinggi adalah dokumen dengan gaya penulisan ACM yaitu *recall* sebesar 92% dan *precision* sebesar 93%. Untuk pengujian unit, dokumen dengan gaya penulisan ACM juga memiliki hasil paling tinggi yaitu *recall* 99% dan *precision* 98%.
- b. Untuk metode uji, uji unit lebih tinggi dibandingkan dengan uji kalimat karena memiliki granularitas yang lebih halus.
- c. Presentase *recall* paling rendah adalah 75% pada dokumen uji kalimat dengan gaya penulisan APA ekstensi PDF.
- d. Konverter mempunyai peranan penting dalam sistem, dalam 38 dokumen uji, terdapat 1 sitasi yang tidak dapat di konversi dengan baik yaitu pada gaya penulisan ACM dokumen id 4 sehingga membuat *recall* menjadi 80%.

5.2 Saran

Saran yang diberikan untuk pengembangan sistem selanjutnya adalah:

- a. Pengambilan kutipan masih berdasarkan tanda titik sebagai batas kalimat, sehingga masih terdapat bagian yang bukan kalimat kutipan seperti misalnya sub-bab terambil, contoh lainnya sumber gambar tidak lengkap terambil dan isi tabel terambil. Oleh karena itu disarankan pada tahap pra-pemrosesan dilakukan segmentasi terhadap judul dan sub-bab, melakukan deteksi gambar dan tabel sehingga data gambar dan tabel tidak terambil. Jika ingin dipertimbangkan, dapat menggunakan salah satu aplikasi, contohnya *Natural language Tool Kit* (NLTK) agar kalimat yang diinginkan dapat terambil seutuhnya.

- b. Terdapat beberapa sitasi pada gaya penulisan APA yang masih belum bisa terambil, misalnya sitasi dengan nomor halaman, nama organisasi, dan judul sehingga mungkin ke depannya nanti bisa dikembangkan sampai pada bagian-bagian tersebut. Salah satu saran untuk masalah ini adalah dengan melakukan *parsing* pada daftar pustaka agar sistem dapat membedakan nama orang, organisasi atau perusahaan dan judul.
- c. Untuk pengembangan selanjutnya, output dari sistem ini dapat dipakai untuk pra-pemrosesan *Bibliographic Coupling (BC)*, *Text Reuse Plagiarism Detection (TRPD)*, *Cited by* untuk ranking universitas dan *paper*.

©UKDW

REFERENSI

- Connan, J., & Omlin, C. W. (2000). Bibliography extraction with hidden markov models. Technical report, University of Stellenbosch.
- Day, M. Y., Tsai, T. H., Sung, C. L., Lee, C. W., Wu, S. H., Ong, C. S., & Hsu W. L. (2005). A knowledge-based approach to citation extraction. In Proceedings of the IEEE International Conference on Information Reuse and Integration, IRI-2005, Vegas, USA, pp. 50–55. IEEE Systems, Man, and Cybernetics Society.
- Fagerberg, A. (2014). Temporal Information Extraction using Regular Expressions. Retrieved 11 10 2016 from http://www.antonfagerberg.com/files/tempex_anton_fagerberg.pdf
- Garfield, E. (2001, November). From Bibliographic Coupling to Co-Citation Analysis via Algorithmic Historio-Bibliography. Retrieved 10 10 2016 from <http://www.garfield.library.upenn.edu/papers/drexelbelvergriffith92001.pdf>
- Geng, J., & Yang, J. (2004). Autobib: Automatic extraction of bibliographic information on the web. In *8th International Database Engineering and Applications Symposium (IDEAS 2004)*, pp. 193–204.
- Gipp, B., & Meuschke, N. (2011). Citation Pattern Matching Algorithms for Citation-based Plagiarism Detection: Greedy Citation Tiling, Citation Chunking and Longest Common Citation Sequence. In Proceedings of the 11th ACM symposium on Document engineering: September 19 - 22, 2011, Mountain View, California, USA / Matthew Hardy (Hrsg.). - New York, NY.
- Goyvaerts, J., & Levithan, S. (2012). *Regular Expressions Cookbook 2nd Edition*. USA: O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.
- Jurafsky, D., & Martin, H. J. (1999). Speech and language processing. New Jersey: Alan Apt Publisher

- Krisnawati, L. D. (2016). *Plagiarism Detection for Indonesian Texts*. Muenchen: UB-Maximilians Univeritat Muenchen
- Li, Y., Krishnamurthy, R., Raghavan, S., Vaithyanathan, S. (2008). Regular expression learning for information extraction. Department of EECS, University of Michigan.
- Peng, F., & McCallum, A. (2004). Accurate information extraction from research papers using conditional random fields. In Proceedings of Human Language Technology Conference and North American (HLT-NAACL) pp. 329–336.
- Powers, D.M.W. (2007). Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. School of Informatics and Engineering Flinders University of South Australia.
- Why citing so important?. (2016, September 20). Retrieved 10 10 2016 from <http://libguides.mit.edu/citing>
- Yangarber, R (2001). Scenario Customization for Information extraction. Ph.D. Dissertation, New York University.