

Sistem Pendeteksi Plagiarisme Pada Dokumen Berbasis *Fuzzy IR*

Plagiarism Detection System on Document Based on Fuzzy IR

Fery Kun Widi Yudantyo¹, Jondri², Bambang Ari Wahyudi³

^{1,3}Prodi S1 Teknik Informatika, Fakultas Informatika, Universitas Telkom

²Prodi S1 Ilmu Komputasi, Fakultas Informatika, Universitas Telkom

ferykun@windowslive.com, jondri@telkomuniversity.ac.id, bambang.ari@gmail.com

Abstrak

Berbagai metode telah dikembangkan dalam bidang pendeteksian plagiarisme. Salah satu metode yang dapat digunakan *Fuzzy Information Retrieval (IR)*. Kelebihan yang dimiliki metode ini adalah mampu mendeteksi bentuk plagiarisme baik *copy-paste* maupun *paraphrase*. Masalah yang muncul pada penggunaan *fuzzy IR* adalah bagaimana menentukan nilai *threshold* optimal yang ada pada *rule*, sehingga *fuzzy IR* dapat dengan baik mengidentifikasi kalimat yang melakukan plagiarisme. Oleh karena itu pada penelitian ini digunakan metode *Fuzzy IR* dengan beberapa kombinasi nilai *threshold* untuk mendeteksi tindakan plagiarisme dalam sebuah kalimat pada sebuah dokumen berbahasa Indonesia. Dari pengujian dan analisis yang telah dilakukan didapatkan kombinasi *threshold* terbaik adalah 0.725 untuk *permission threshold* dan 0.275 untuk *variance threshold*. Dari hasil pengujian dan analisis juga diketahui bahwa semakin besar nilai yang diberikan untuk *permission threshold* dan semakin kecil untuk nilai *variance threshold* maka toleransi *fuzzy IR* terhadap perubahan struktur kalimat juga semakin rendah, hal ini menyebabkan semakin sulit untuk mendeteksi plagiarisme dengan banyak perubahan pada struktur kalimat.

Kata kunci : Plagiarisme, Pendeteksian Plagiarisme, *Fuzzy Information Retrieval*, *Permission Threshold*, *Variance Threshold*

Abstract

There are some methods which have been developed in plagiarism detection field. One of them is *Fuzzy Information Retrieval (IR)*. The advantages of this method is it can detect copy paste or paraphrasing plagiarism. An issue we face when we use this method is how to determine an optimal value for threshold inside the rule, so it can determine which sentence is plagiarize well. Therefore in this research, *fuzzy IR* method is used with some several value of threshold to determine a plagiarize sentence from an Indonesian document. Based on test and analysis that had done before, an optimal value for permission threshold is 0.725 and 0.275 for variance threshold. Based on the analysis result too, known that the larger value used for permission, and the smaller value used for variance threshold, so the *Fuzzy IR* tolerance for changing of the sentence's structure get smaller. So it can make more difficult for the system to determine which sentence is plagiarize.

Keywords: Plagiarism, *Fuzzy Information Retrieval*, Web Based Application

1. Pendahuluan

Pada bidang penulisan, plagiarisme adalah penggunaan tulisan orang lain dalam tulisan sendiri tanpa memberikan kredit terhadap penulis asli, sehingga menyebabkan pembaca menganggap bahwa tulisan tersebut merupakan hasil pemikiran asli dan bukan kutipan dari orang lain [1]. Tindakan plagiarisme merupakan salah satu pelanggaran berat di bidang akademik dan juga bidang lainnya. Oleh karena itu diperlukan sebuah sistem yang dapat membantu dalam melakukan pendeteksian tindakan plagiarisme. Beberapa metode telah dikembangkan untuk mendeteksi tindakan plagiarisme pada suatu dokumen [5]. Mendeteksi plagiarisme dengan menggunakan metode *Char-Based (CNG)*, *Vector-Based (VEC)*, *Syntax-Based (SYN)* memiliki keunggulan dalam proses identifikasi plagiarisme yang cepat dan efisien, namun kelemahannya adalah metode tersebut hanya dapat digunakan untuk mendeteksi plagiarisme yang merupakan hasil *copy and paste* tanpa ada perubahan pada struktur kalimat. Sedangkan jika menggunakan *Structural-Based (STRUC)* dan *Stylometric-Based (STYLE)* tingkat

akurasi untuk mendeteksi plagiat yang dilakukan dengan paraphrase masih perlu diuji lebih lanjut. Oleh karena itu pada penelitian ini digunakan metode *Fuzzy-Based (FUZZY)*, yang memiliki tingkat akurasi baik dalam mendeteksi plagiat baik berupa copy and paste maupun paraphrase. *Fuzzy* yang digunakan pada penelitian ini adalah *Fuzzy IR (Information Retrieval)*.

Beberapa penelitian yang menerapkan *Fuzzy IR* dalam usaha untuk mendeteksi tindak plagiarisme telah dilakukan [7] [8]. Pada penelitian tersebut *Fuzzy IR* digunakan untuk menentukan tingkat kemiripan antara 2 buah dokumen. Penelitian yang telah dilakukan sebelumnya [7], menggunakan dokumen yang menggunakan bahasa arab sebagai data pengujian, sedangkan penelitian lain [8], menggunakan dokumen berbahasa inggris. Sedangkan penelitian untuk pendeteksi plagiarisme menggunakan dokumen berbahasa indonesia masih terbatas, oleh karena itu pada penelitian kali ini dokumen yang diuji adalah dokumen yang menggunakan bahasa indonesia.

Penerapan *Fuzzy IR (Information Retrieval)* dalam studi kasus pendeteksian tindak plagiarisme pada dokumen melibatkan beberapa nilai *threshold*. Nilai *threshold* tersebut digunakan sebagai *rule* untuk menentukan plagiarisme pada sebuah kalimat dalam masing-masing dokumen yang berbeda. Pada penelitian ini dicoba beberapa kombinasi dari nilai *threshold* tersebut, kemudian dilakukan analisis bagaimana performansi yang ditunjukkan sistem pada tahap pengujian dengan masing-masing kombinasi *threshold*.

2. Dasar Teori dan Perancangan

2.1 Plagiarisme

Plagiarisme atau sering disebut plagiat adalah penjiplakan atau pengambilan karangan, tulisan, pendapat, dan sebagainya dari orang lain dan menjadikannya seolah karangan dan pendapat sendiri [1]. [5] mengategorikan tindakan plagiarisme menjadi beberapa kategori sebagai berikut.

1. *Literal Plagiarism*

Literal Plagiarism merupakan bentuk plagiarisme yang sering ditemukan, dan tidak membutuhkan waktu yang lama dalam mengenali tindak plagiarisme tersebut [5]. Beberapa tindak plagiarisme yang termasuk di dalam kategori literal plagiarism adalah *copy-paste* (baik secara keseluruhan atau sebagian) dan *modified copy* (menyalin dengan mengubah struktur kalimat).

2. *Intelligence Plagiarism*

Intelligence Plagiarism merupakan tindak plagiarisme dimana plagiator (orang yang melakukan tindakan plagiat) berusaha untuk menutupi tindak plagiarisme yang ia lakukan, sehingga pembaca akan menyangka bahwa tulisan tersebut orisinal [5]. Beberapa cara yang dapat digunakan plagiator dalam usaha untuk menutupi tindakan plagiarisme dapat dilakukan adalah sebagai berikut.

a. *Text Manipulation*

Text manipulation adalah usaha untuk memanipulasi teks dan mengubahnya sehingga tidak menyerupai teks yang orisinal [5]. Cara ini dapat dilakukan dengan melakukan *syntactical paraphrasing* (mengubah beberapa kata menjadi sinonim/antonimnya), atau menambahkan beberapa kata yang tidak mengubah makna dari kalimat tersebut [5].

b. *Translating*

Translating adalah usaha menerjemahkan dari kalimat dari suatu bahasa ke dalam bahasa lain, sehingga secara kasat teks yang diambil tidak sama dengan teks orisinal [5].

c. *Idea Adoption*

Idea adoption adalah usaha untuk menyalin teks dari sumber dengan kalimat yang berbeda namun tanpa mengubah makna [5]. Bentuk plagiarisme dapat dilakukan dengan cara mengambil beberapa kata dari teks asli yang merupakan kata kunci, atau merangkum dari teks asli [5].

2.2 *Stemming*

Stemming merupakan suatu proses untuk mengubah suatu kata menjadi bentuk kata dasar dengan cara menghilangkan semua bentuk imbuhan (*affixes*) yang ada pada kata tersebut. Tujuan dilakukan *stemming* adalah agar suatu kata sesuai dengan struktur morfologi Bahasa Indonesia yang baik dan benar, sehingga dapat diproses dengan lebih baik.

Imbuhan yang mungkin muncul pada dokumen Bahasa Indonesia terdiri dari awalan (*prefix*), akhiran (*suffix*), dan sisipan (*infix*).

2.3 *Stopword*

Stopword merupakan kata yang umum ditemukan pada sebuah teks dokumen dalam jumlah besar dan dianggap tidak memiliki makna atau tidak memiliki relevansi dengan informasi yang terdapat dalam dokumen. Pada implementasi *information retrieval*, *stopword* akan dihilangkan dari teks dokumen untuk mengurangi beban

komputasi.

Beberapa kata yang termasuk ke dalam kategori *stopword* diantaranya adalah kata ganti orang (seperti aku, dia, kamu, mereka) dan kata penghubung (seperti dan, atau, akan tetapi, namun).

2.5 Information Retrieval

ISO 2382/1 mendefinisikan *Information Retrieval (IR)* sebagai tindakan, metode dan prosedur untuk menemukan kembali data yang tersimpan, kemudian menyediakan informasi mengenai subyek yang dibutuhkan. Tindakan tersebut mencakup *text indexing*, *inquiry analysis*, dan *relevance analysis* [10]. Atau secara umum information retrieval didefinisikan sebagai usaha untuk menemukan dokumen berdasarkan teks / isi yang memenuhi kebutuhan informasi user dari sekumpulan koleksi dokumen.

2.6 Fuzzy Information Retrieval

Fuzzy Information Retrieval adalah salah satu *classical model* pada bidang *information retrieval*, bersama dengan *Boolean IR*. Pada implementasi *Fuzzy IR* dalam mendeteksi plagiarisme terdapat beberapa langkah [7] yang harus dilakukan, yaitu.

1. Preprocessing

Pada tahap *preprocessing* ini dibuat sebuah *rule* untuk menentukan nilai *correlation factor* antara sebuah *term* dengan *term* yang lain. Nilai *correlation factor* didapatkan berdasarkan kesamaan arti dari sepasang *term*. *Rule* yang digunakan untuk menentukan *correlation factor* adalah sebagai berikut.

$$c_{i,j} \begin{cases} 1 & \text{if } k_i = k_j \\ 0.5 & \text{if } k_i \text{ and } k_j \text{ is a synonym} \\ 0.001 & \text{otherwise} \end{cases} \quad (1)$$

2. Fuzzification

Pada penelitian ini fuzzification ditujukan agar teks yang ada pada dokumen, baik pada dokumen corpus maupun dokumen query dapat diproses oleh fuzzy IR. Hal ini dilakukan dengan memberikan bobot nilai untuk teks yang ada pada dokumen tersebut dengan menggunakan sebuah membership function seperti berikut.

$$\mu_{i,j} = 1 - \prod_{k \in S_j} (1 - c_{i,k}) \quad (2)$$

Membership function tersebut digunakan \neg untuk menentukan bobot nilai dari term-to-sentence correlation ($u_{i,j}$) yang menunjukkan tingkat kedekatan atau keterkaitan antara sebuah term k_i dengan kalimat S_j .

3. Defuzzification

Pada tahap *defuzzification* ini, *fuzzy IR* yang digunakan akan melakukan justifikasi terhadap dokumen yang diberikan, apakah dokumen *query* yang digunakan melakukan tindakan plagiarisme terhadap dokumen *corpus* yang diberikan atau tidak. Justifikasi diberikan dengan melihat bobot nilai yang diberikan untuk tingkat kemiripan atau kedekatan antara tiap kalimat pada dokumen *query* dengan kalimat yang ada pada dokumen *corpus* yang disebut *degree of similarity*. Sebelum melakukan justifikasi terhadap sepasang kalimat, terlebih dahulu dihitung nilai *degree of similarity* antara 2 buah kalimat dengan menggunakan persamaan berikut.

$$Sim(S_i, S_j) = \frac{\sum_{i=1}^n \mu_{i,j}}{n} \quad (3)$$

Setelah didapatkan nilai *degree of similarity* dari sepasang kalimat tersebut kemudian dapat dilakukan justifikasi apakah kalimat yang terdapat pada dokumen *query* melakukan plagiarisme terhadap kalimat yang ada pada dokumen *corpus*. Justifikasi terhadap sepasang kalimat dilakukan dengan menggunakan aturan berikut.

$$EQ(S_i, S_j) = \begin{cases} 1, & \text{if } \text{MIN}(Sim(S_i, S_j), Sim(S_j, S_i)) \geq p \cap \\ & |Sim(S_i, S_j) - Sim(S_j, S_i)| \leq v \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Pada penerapan aturan diatas dilibatkan 2 nilai threshold, yakni permission (p) dan variance (v). Permission threshold digunakan sebagai ambang batas minimum nilai degree of similarity dari sepasang kalimat untuk dapat dinyatakan sebagai kalimat yang melakukan tindakan plagiarisme. Sedangkan variance threshold digunakan sebagai ambang batas maksimal dari perbedaan nilai degree of similarity dari sepasang kalimat. Dengan menggunakan aturan diatas didapatkan nilai $EQ(S_i, S_j)$ adalah antara 1 atau 0. Jika nilai dari $EQ(S_i, S_j)$ adalah 1 maka hasil justifikasi yang dilakukan adalah bahwa kalimat S_i pada dokumen query melakukan plagiarisme terhadap kalimat S_j pada dokumen corpus. Sedangkan ketika nilai yang didapatkan adalah 0 maka justifikasi yang diberikan adalah kalimat tersebut tidak melakukan plagiarisme.

2.7 Precision dan Recall

Precision adalah perbandingan antara jumlah data yang relevan yang berhasil teridentifikasi oleh sistem terhadap jumlah data keseluruhan yang teridentifikasi, sedangkan *recall* adalah perbandingan antara jumlah data yang relevan yang berhasil teridentifikasi dengan jumlah data relevan yang seharusnya teridentifikasi. Pada penerapan dalam *plagiarism detection*, data yang relevan adalah kalimat / bagian pada dokumen yang secara aktual melakukan plagiarisme. Untuk menghitung nilai *precision* dapat menggunakan persamaan berikut.

$$Precision = \frac{tp}{tp + fp} \quad (5)$$

Sedangkan, untuk menghitung nilai recall dapat menggunakan persamaan dibawah ini.

$$Recall = \frac{tp}{tp + fn} \quad (6)$$

tp adalah jumlah data *true positive*, yaitu jumlah data relevan yang berhasil teridentifikasi oleh sistem. Sementara fp adalah jumlah data yang tidak relevan namun ikut teridentifikasi oleh sistem (*false positive*) dan fn adalah jumlah data yang relevan namun tidak berhasil teridentifikasi oleh sistem (*false negative*).

2.8 Lexical Database

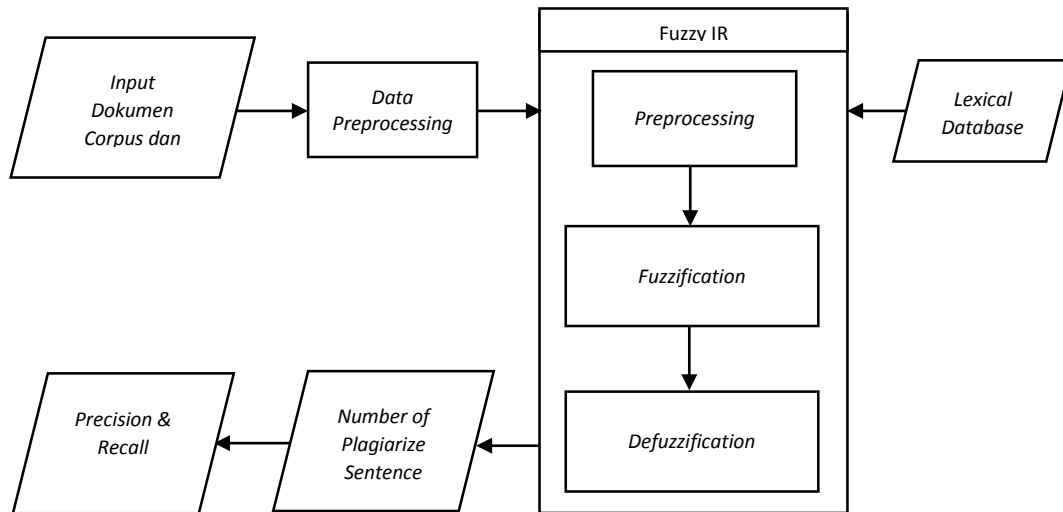
Lexical database merupakan sebuah *database* yang digunakan sebagai referensi dalam menentukan nilai *correlation factor* dari sebuah term terhadap term lain. *Lexical database* ini digunakan untuk menampung daftar pasangan kata yang merupakan sinonim dalam bahasa indonesia. Daftar pasangan kata yang merupakan sinonim didapatkan berdasarkan sebuah kamus bahasa indonesia sebagai referensi [12]. Berikut ini adalah contoh *lexical database* yang digunakan pada penelitian ini.

Tabel 1 Contoh Lexical Database

<i>Id_term</i>	<i>Term</i>	<i>Sinonim</i>
1	Absah	Sah
2	Absolut	Mutlak
3	Ahli	Pakar
4	Andal	Tangguh
5	Bahari	Laut
6	Dampak	Akibat
7	Frekuensi	Sinyal
8	Forum	Lembaga

2.9 Perancangan Sistem

Berikut ini adalah rancangan sistem yang dibangun pada penelitian ini.



Gambar 2 Alur Proses Sistem

1. Input Dokumen *Corpus* dan *Query*

Pada tahap ini user menginputkan 2 buah dokumen file pdf, masing-masing 1 dokumen *corpus* dan dokumen *query*.

2. *Data Preprocessing*

Pada *preprocessing* ini dilakukan 3 tahap proses yaitu, membaca teks dari dokumen, *split* teks menjadi kalimat, *stopword removal*, *stemming*. Pada proses *stemming* yang dilakukan pada penelitian ini dilakukan dengan menggunakan bantuan library PHP yaitu *sastrawi*. Sementara untuk membaca teks dari dokumen menggunakan bantuan *PDFParser*.

3. *Preprocessing*

Pada tahap ini tiap *term* / kata yang terdapat pada dokumen *query* dan tidak termasuk kedalam *stopword list* dicari nilai *correlation factor*-nya terhadap tiap *term* pada dokumen *corpus* yang juga tidak termasuk ke dalam *stopword list*. Untuk menentukan nilai *correlation factor* dari tiap *term* tersebut digunakan *rule* dari persamaan (1). Pada tahap ini juga, *lexical database* yang telah dirancang, digunakan untuk membantu menentukan nilai *correlation factor* dari sebuah *term* terhadap *term* lain. Berdasarkan *rule* dari persamaan (1) maka nilai *correlation factor* antara 2 buah *term* bernilai 1 jika kedua *term* tersebut adalah *term* yang sama, 0.5 jika *query* terhadap *lexical database* yang dijalankan mengembalikan hasil yang tidak *null*, dan 0.001 jika tidak memenuhi salah satu diantara kedua *rule*.

4. *Fuzzification*

Pada tahap ini dihitung nilai keanggotaan dari *fuzzy IR* yang digunakan. Nilai keanggotaan tersebut disebut juga *term-to-sentence correlation*. *Term-to-Sentence correlation* adalah sebuah nilai yang menunjukkan relevansi antara sebuah *term* dengan sebuah kalimat. Semakin tinggi nilai *term-to-sentence correlation* maka semakin tinggi juga relevansi *term* tersebut dengan sebuah kalimat dan sebaliknya. *Term-to-sentence correlation* dapat dihitung dengan menggunakan persamaan (2).

Relevansi antar kalimat tidak hanya dilihat dengan ada atau tidaknya *term* tersebut pada kalimat yang ingin dicari relevansinya. Melainkan juga dilihat relevansi dari kesamaan semantik, atau dalam hal ini adalah kata – kata yang merupakan sinonim. Pada suatu kondisi nilai keanggotaan bisa mendekati 1 tanpa ada kemunculan *term* tersebut, kondisi tersebut dapat ditemui jika terdapat banyak kemunculan kata sinonimnya atau kalimat yang diperiksa terdiri dari *term* yang banyak. Disinilah peran *fuzzy IR* yang digunakan untuk memberikan bobot nilai dari sebuah *term* terhadap kalimat.

5. *Defuzzification*

Setelah didapatkan bobot nilai keanggotaan untuk tiap *term* pada kedua dokumen (dokumen *corpus* dan dokumen *query*), langkah berikutnya adalah melakukan *defuzzifikasi* untuk menentukan tingkat kemiripan antara 2 buah kalimat, masing-masing 1 pada dokumen *corpus* dan 1 pada dokumen *query* (*degree of similarity*) untuk kemudian dilakukan *justifikasi* terhadap 2 kalimat tersebut apakah kalimat tersebut merupakan kalimat yang plagiarisme atau tidak.

Untuk menentukan tingkat kemiripan antara 2 buah kalimat digunakan persamaan (3) yang melibatkan nilai keanggotaan dari tiap *term* yang terdapat pada kalimat tersebut. Namun hasil dari persamaan (3) bukan merupakan hasil akhir, dari nilai hasil persamaan tersebut kemudian dilakukan justifikasi menggunakan aturan dari persamaan (4) untuk mengetahui apakah kalimat tersebut melakukan tindakan plagiarisme.

6. *Number of Plagiarize Sentence*

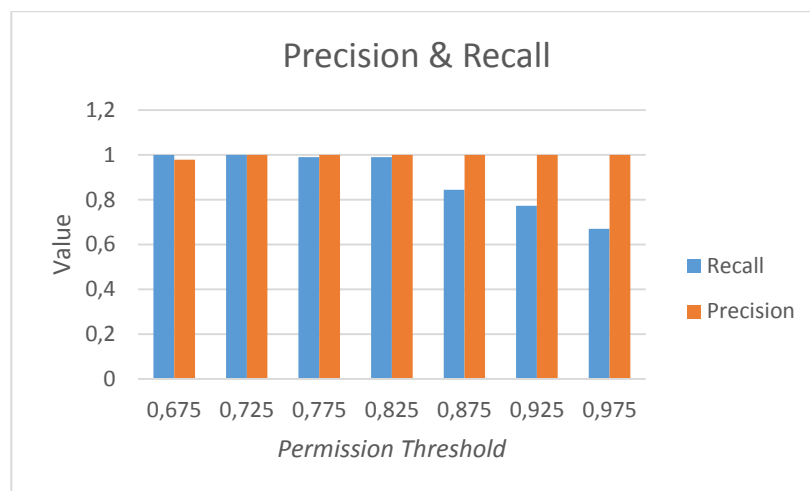
Pada tahap ini dihitung jumlah kalimat dari dokumen *query* yang setelah tahap justifikasi dinyatakan melakukan plagiarisme oleh sistem terhadap salah satu kalimat pada dokumen *corpus*.

7. *Precision dan Recall*

Pada tahap ini dihitung nilai *precision* dan *recall* dari pengujian yang dilakukan dengan menggunakan persamaan (5) dan (6) untuk dianalisis bagaimana rancangan dari sistem yang telah dibangun.

3. Pembahasan

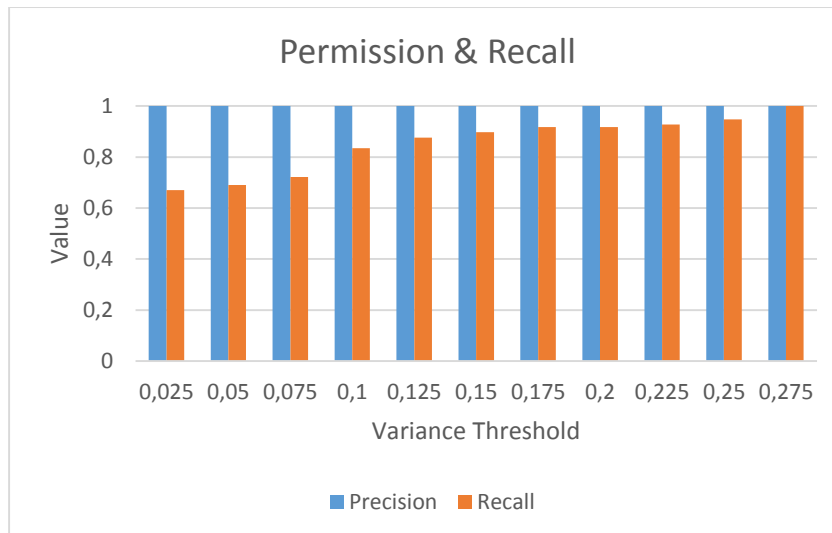
Dari pengujian yang telah dilakukan terhadap beberapa nilai *permission threshold* didapatkan nilai *precision* dan *recall* dari masing-masing pengujian seperti pada gambar dibawah ini.



Gambar 3 Precision dan Recall Pengujian Permission Threshold

Dari grafik *precision* dan *recall* hasil pengujian terhadap beberapa nilai *permission threshold* dapat dilihat bahwa semakin besar nilai *permission threshold* yang digunakan semakin baik nilai *precision* yang didapatkan. Hal ini disebabkan karena dengan semakin besar nilai dari *permission threshold*, maka toleransi terhadap perubahan struktur kalimat akan semakin rendah, sehingga semakin baik dalam mendeteksi kalimat non plagiat. Namun dilain pihak dengan semakin besar nilai *permission threshold* yang digunakan, maka plagiarisme yang banyak mengubah struktur kalimat dari kalimat orisinal, seperti pada *text manipulation* dan *idea adoption*, akan semakin sulit untuk diidentifikasi dengan baik oleh sistem.

Dari hasil pengujian seperti yang terlihat pada gambar diatas, dapat dilihat bahwa nilai $p = 0.725$ memiliki nilai *precision* dan *recall* yang paling baik dibandingkan dengan nilai lain yang telah diuji. Oleh karena itu dalam pengujian untuk menentukan nilai optimal untuk *variance threshold*, nilai *permission threshold* yang digunakan adalah 0.725.



Gambar 4 Precision dan Recall Pengujian Variance Threshold

Dari gambar hasil pengujian diatas dapat dilihat bahwa nilai *recall* mengalami peningkatan yang cukup signifikan, yaitu 0.11 pada saat pengujian nilai *variance threshold* di-set dengan nilai 0.1. Hal tersebut terjadi dikarenakan dengan semakin besar nilai *variance threshold* yang digunakan maka pada saat justifikasi dengan *fuzzy IR* dilakukan toleransi yang dapat diterima oleh sistem terhadap perbedaan struktur kalimat akan meningkat sehingga sistem akan memiliki performansi yang lebih baik dalam mengidentifikasi jenis plagiarisme *text manipulation* dan *idea adoption*. Hal ini juga menunjukkan bahwa terdapat cukup banyak kalimat yang melakukan plagiarisme pada *dataset* yang digunakan yang memiliki perbedaan nilai *degree of similarity* antara 0.075 hingga 0.10.

Dari gambar hasil pengujian diatas juga dapat dilihat bahwa nilai untuk *variance threshold* yang paling optimal adalah 0.275. Hal ini ditunjukkan dengan nilai *recall* dan *precision* yang lebih baik dari nilai lain yang diuji.

4. Kesimpulan

Berdasarkan analisis dan hasil pengujian yang dilakukan pada penelitian ini, dapat disimpulkan :

1. Kombinasi nilai *threshold* memiliki tingkat akurasi paling tinggi dalam penelitian ini adalah $p = 0.725$ dan $v = 0.275$, dengan nilai *precision* dan *recall* yang mencapai 100%.
2. Semakin besar nilai untuk *threshold permission*, atau semakin kecil nilai untuk *threshold variance* maka sistem akan semakin sulit untuk mendeteksi kalimat plagiat yang mengubah banyak struktur kalimat seperti *text manipulation* dan *idea adoption*. Hal ini disebabkan karena toleransi dari sistem terhadap perubahan struktur kalimat yang semakin berkurang.

Daftar Pustaka:

- [1] Ismail., Yunarso, Eka Widhi. 2014. *Aplikasi Berbasis Web Pendeteksi Plagiarisme Menggunakan Algoritma Himpunan Kata*. Jurnal Infotel : Indonesia
- [2] Alzahrani, Salha Mohammed., Salim, Naomie. 2008. *Plagiarism Detection In Arabic Scripts Using Fuzzy Information Retrieval*. Malaysia : Springer
- [3] Ismail, Heba M. 2014. *Using Concept Maps and Fuzzy Set Information Retrieval Model To Dynamically Personalize RSS Feeds*. UAE: IJCSNS.
- [4] Osman, Ahmed Hamza., Salim, Naomie., Kumar, Yogan Jaya., and Abuobieda, Albara. 2012. *Fuzzy Semantic Plagiarism Detection*. Germany: Springer.
- [5] Alzahrani, Salha M., Salim, Naomie., Abraham, Ajith. 2012. *Understanding Plagiarism Linguistic Patterns, Textual Features, and Detection Method*. Saudia Arabia: IEEE.

- [6] Alzahrani, Salha M., Salim, Naomie., Kent, Chow Kok., Binwahlan, Mohammed Salem., Suanmali, Ladda. 2010. *The Development of Cross-Language Plagiarism Detection Tool Utilising Fuzzy Swarm-Based Summarisation*. Saudia Arabia: IEEE.
- [7] Alzahrani, Salha M., Salim, Naomie. 2009. *On the Use of Fuzzy Information Retrieval for Gauging Similarity of Arabic Documents*. Saudi Arabia : IEEE.
- [8] Yerra, Rajiv., Ng, Yiu-Kai. 2005. *A Sentence-Based Copy Detection Approach for Web Documents*. Germany: Springer.
- [9] IndoDic.com, *Forming Indonesian Words & Using Indonesian Affixes*. 12 Juni 2015. <<http://indodic.com/affixeng.html> >
- [10] *Fuzzy Information Retrieval.docx* <<http://bethanurinasari.files.wordpress.com/2013/12/fuzzy-information-retrieval.docx>> didownload pada 15 November 2014.
- [11] Tala, F. Z. 2003. *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. M.S. thesis. M.Sc. Thesis. Master of Logic Project. Institute for Logic, Language and Computation. Universiteti van Amsterdam The Netherlands.
- [12] Alexander, Bonafacio., Wulandari, Yettik. 2014. *Kamus Poket Bahasa Indonesia*. Yogyakarta : Aksara Sukses.