

Pengembangan Sistem Deteksi Kesesuaian Dokumen Proposal Program Kreativitas Mahasiswa Dengan Metode *Extended Weighted Tree Similarity*

Wiwik Suharso¹⁾, Qurrota A'yun²⁾, Deni Arifianto³⁾

^{1,2,3)} Jurusan Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
Jl. Karimata No. 49 Jember Kode Pos 68121
Email : ¹⁾ wiwiksuharso@unmuhjember.ac.id

ABSTRAK

Dalam rangka meningkatkan kualitas proposal dan kuantitas perolehan dana hibah Program Kreativitas Mahasiswa (PKM), Tim seleksi internal di Perguruan Tinggi mengevaluasi kesesuaian dokumen digital proposal usulan mahasiswa dengan dokumen pedoman PKM yang berlaku. Evaluasi tersebut seringkali tidak optimal karena ketidakseimbangan jumlah anggota tim seleksi dan jumlah proposal yang ditangani serta perubahan ketentuan dalam setiap tahun pedoman baru. Proposal usulan mahasiswa umumnya memiliki kesalahan dalam sistematika penulisan, uraian isi, ketentuan kriteria dan persyaratan pengusul dalam skema yang dipilih sehingga upaya perbaikan membutuhkan banyak waktu, tenaga dan biaya. Oleh karena itu tujuan jangka panjang penelitian ini adalah tersedianya perangkat lunak sistem pendeteksian otomatis dokumen digital yang mampu mengidentifikasi kesalahan sistematika, isi dan ketentuan sesuai dokumen pedoman PKM yang berlaku. Target khusus yang akan dicapai adalah ditemukannya model pendeteksian dokumen digital format pdf yang memiliki tingkat presisi dan akurasi tinggi sehingga dapat digunakan oleh tim seleksi internal sebagai alat bantu koreksi awal. Pemodelan ini memanfaatkan operasi text mining untuk mengidentifikasi isi dokumen dengan melibatkan serangkaian metode yaitu tokenizing, stoplist/wordlist, stemming, pembobotan kata TF-IDF dan kemiripan antar bagian dokumen dalam struktur tree dengan metode cosine measure yang dinyatakan dalam bentuk nilai bobot label node tree. Perhitungan kemiripan total rata-rata antara dua buah tree dokumen dengan metode Extended Weighted Tree Similarity menggambarkan hasil kesesuaian sejauh mana proposal mahasiswa telah menerapkan pedoman PKM. Prototipe perangkat lunak telah menghasilkan nilai auto matching antar bagian (sub tree) yang sama dan scoring antara dokumen proposal dan dokumen pedoman PKM. Data pelatihan dari proposal yang lolos pendanaan tidak selalu memiliki nilai tertinggi dibandingkan dengan data pengujian, sehingga kepatuhan terhadap pedoman harus diikuti oleh kualitas dari proposal yang diusulkan. Sebagian besar data pelatihan memiliki nilai tertinggi pertama dan kedua, sehingga kepatuhan terhadap pedoman menjadi pertimbangan dalam menentukan proposal yang didanai.

Kata Kunci : Kesesuaian Dokumen, Proposal, Pedoman PKM, *Text Mining*, *Cosine Measure*, *Extended Weighted Tree Similarity*

1. PENDAHULUAN

Program Kreativitas Mahasiswa (PKM) merupakan salah satu program dari Direktorat Jenderal Pembelajaran dan Kemahasiswaan, Kementerian Riset, Teknologi, dan Pendidikan Tinggi untuk meningkatkan mutu mahasiswa di Perguruan Tinggi. Mutu mahasiswa

tersebut salah satunya dapat diukur dari kompetensi untuk menghasilkan karya kreatif dan inovatif berlandaskan IPTEKS (Pedoman PKM 2015). Kegiatan hibah PKM mendapat respon positif dari Perguruan Tinggi dan mahasiswa. Respon tersebut berupa pembentukan tim seleksi internal untuk meningkatkan

kualitas dan kuantitas proposal usulan mahasiswa dalam setiap tahun. Tim seleksi bertugas mengevaluasi kesesuaian dokumen digital format pdf proposal usulan mahasiswa dengan dokumen pedoman PKM yang berlaku. Evaluasi tersebut seringkali tidak optimal karena ketidakseimbangan jumlah anggota tim seleksi dan jumlah proposal usulan yang ditangani serta perubahan ketentuan dalam setiap tahun pedoman baru. Proposal usulan mahasiswa umumnya memiliki kesalahan dalam sistematika penulisan, uraian isi, ketentuan kriteria dan persyaratan pengusul dalam skema yang dimaksudkan sehingga upaya perbaikan membutuhkan banyak waktu, tenaga dan biaya. Hal ini didukung oleh data dan informasi tim seleksi PKM Universitas Muhammadiyah Jember (UM Jember) bahwa (1) Tim seleksi internal terdiri dari 5 orang sesuai SK Rektor Nomor : 0203/KEP/II.3.AU/F/2016, (2) Jumlah kelompok pengusul PKM tahun 2015 sebanyak 149 dan sejak tahun 2013 terjadi peningkatan rata-rata pertahun sebesar 24%, (3) Perubahan pedoman PKM terjadi setiap tahun seperti tahun 2013, 2014, 2015, dan perubahan tersebut rata-rata terjadi menjelang waktu pengunggahan proposal usulan baru PKM 5 bidang atau pelaksanaan PIMNAS, (4) Jumlah proposal diterima atau penerima hibah PKM menurun tajam dari 19 kelompok pada tahun 2015 menjadi 5 kelompok pada tahun 2016 atau turun sebesar 280%, (5) Perubahan kebijakan bidang kemahasiswaan UM Jember tahun 2016 memasukkan persyaratan pembuatan proposal usulan PKM bagi pelamar beasiswa mempengaruhi peningkatan jumlah proposal yang harus dikoreksi (Dokumen Tim PKM, 2016). Bahkan observasi *reviewer* terhadap skema PKM-KT menemukan fakta

kesalahan administratif dan atau substansi di atas 50% berakibat proposal PKM tidak lolos pada saat pra-evaluasi (Pedoman PKM 2015).

Untuk mengurangi kesalahan administratif dan substansi isi diperlukan perhatian dosen pembimbing dan tim seleksi internal dalam membantu mahasiswa. Sejalan dengan tugas tersebut ketersediaan perangkat lunak sistem pendeteksian otomatis dokumen digital proposal mahasiswa dapat membantu menemukan kesalahan sistematika, isi dan ketentuan sesuai pedoman PKM sehingga upaya perbaikan lebih mudah dan cepat. Oleh karena itu penelitian ini mengusulkan model pendeteksian kesesuaian dokumen digital pdf proposal PKM dengan metode *Extended Weighted Tree Similarity* (EWTS). Menurut Sulistyono (2008), Sarno (2008) algoritma EWTS memanfaatkan *text mining* dengan mempertimbangkan frekuensi suatu kata dalam suatu dokumen (*term frequency*), dan penyebaran suatu kata pada sekumpulan dokumen (*Inverse Document Frequency*) serta kemiripan antar dokumen dengan metode *cosine measure* dinyatakan dalam nilai bobot label node *tree*. Perhitungan kemiripan antara dua buah *tree* dengan metode EWTS akan menghasilkan nilai kesesuaian sejauh mana proposal mahasiswa telah menerapkan pedoman PKM. Dalam penelitian ini permasalahan yang akan diselesaikan adalah bagaimana caranya memodelkan sistem pendeteksian otomatis kesesuaian dokumen digital format pdf proposal usulan dan pedoman PKM untuk menilai tingkat kesesuaian sejauh mana proposal telah menerapkan pedoman, serta bagaimana membangun perangkat lunak sistem pendeteksian otomatis kesesuaian dokumen digital.

2. TINJAUAN PUSTAKA

Dokumen direpresentasikan dalam bentuk *tree* berdasarkan sistematika setiap skim PKM yang telah ditetapkan pada pedoman PKM yang berlaku. Setiap skim PKM memiliki 4 bagian yaitu bagian kriteria, bagian awal, bagian isi, dan bagian lampiran, dimana masing-masing bagian dapat memiliki sub bagian dan pada sub bagian tersebut terdapat dokumen atau penjelasan detail dari suatu bagian. Oleh sebab itu, pada sub bagian inilah fokus pengukuran kesesuaian dokumen dilakukan. Berdasarkan struktur tersebut maka pembentukan *tree* terbagi atas 4 lapis.

Contoh pembentukan *tree* pada skim PKMK digambarkan pada Gambar 1. terdiri dari bagian-bagian sebagai berikut.

- 1) Bagian Kriteria terdiri dari : Inti Kegiatan, Materi Kegiatan, Strata Pendidikan, Jumlah Anggota, Alokasi Pendanaan, Tahun Angkatan, Luaran, Jumlah Halaman, Format File.
- 2) Bagian awal terdiri dari : Halaman Sampul, Halaman Pengesahan, Daftar Isi, Ringkasan.
- 3) Bagian Isi terdiri dari : Bab 1 Pendahuluan, Bab 2 Gambaran Umum Rencana Usaha, Bab 3 Metode Pelaksanaan, Bab 4 Biaya dan Jadwal Kegiatan.
- 4) Bagian Lampiran terdiri dari : Lampiran 1 Biodata Ketua, Anggota dan Dosen Pembimbing, Lampiran 2 Justifikasi Anggaran Kegiatan, Lampiran 3 Susunan Organisasi Tim Kegiatan dan Pembagian Tugas, Lampiran 4 Surat Pernyataan Ketua Kegiatan.

Text mining didefinisikan sebagai proses penemuan kembali relasi dan fakta yang terkubur didalam teks dan tidak harus baru (Sulistyo, 2008). Dalam penelitian ini *text mining* meliputi *tokenizing*, *stoplist/wordlist*, *stemming*.

Dalam penelitian Suharso (2015) ketiga operasi *text mining* tersebut sebagai tahap persiapan untuk menyaring dan mengurangi jumlah informasi atau term-term yang tidak memiliki relevansi dalam pengukuran derajat kemiripan suatu dokumen. Proses selanjutnya adalah pembobotan kata dan pengukuran kemiripan dokumen. Masing-masing tahapan tersebut dijelaskan dengan merujuk pada penelitian Sulistyo (2008) (Suharso, 2015).

1) *Tokenizing*

Tokenizing adalah proses pemecahan dokumen atau paragraf atau kalimat menjadi daftar kata atau *token* yang berdiri sendiri. Fungsi *token* tersebut akan melakukan pengecekan terhadap jarak antar kata (spasi, tabulasi, enter) untuk membuat daftar kata pada semua kata yang terdapat dalam dokumen.

2) *Stoplist/Wordlist*

Stoplist adalah daftar kata yang tidak relevan dalam mengidentifikasi isi suatu dokumen seperti kata sambung, kata depan, kata ganti, simbol dan tanda baca. *Wordlist* adalah daftar kata yang relevan sebagai kata kunci sehingga akan digunakan dalam stemming. *Stoplist/Wordlist* adalah proses pembuangan atau penyaringan terhadap *stoplist* sehingga dihasilkan *wordlist*.

3) *Stemming*

Stemming bertujuan mengubah atau mengembalikan kata menjadi bentuk dasarnya dengan menghilangkan imbuhan-imbuhan pada kata dalam dokumen. Pembentukan kata dasar ini menggunakan algoritma porter dari Dr. Martin Porter yang disesuaikan atau dikembangkan dalam Bahasa Indonesia.

Sistem pencarian informasi secara otomatis dapat dilakukan dengan membandingkan *content identifier* berupa kata (*term*) yang terdapat pada teks (*document*) dan informasi yang diminta oleh user (*user information queries*). Dokumen dapat berupa dokumen, paragraf atau kalimat *D* dinyatakan dalam *term vectors*.

$$D = (t_i, t_j, \dots, t_p) \quad (1)$$

Dimana setiap t_k mengidentifikasi term yang terdapat pada dokumen *D*. Demikian juga pada *query* *Q* direpresentasikan dalam *term vectors*.

$$Q = (q_a, q_b, \dots, q_r) \quad (2)$$

Dimana setiap q_k mengidentifikasi term yang terdapat pada query *Q*. Sehingga bobot (*weight*) pada setiap term untuk membedakan term yang terdapat dalam dokumen *D* dan query *Q* dituliskan sebagai berikut.

$$D = (t_0, w_{d0}; t_1, w_{d1}; \dots; t_t, w_{dt}) \quad (3)$$

$$Q = (q_0, w_{q0}; q_1, w_{q1}; \dots; q_t, q_a, w_{qt}) \quad (4)$$

Dimana w_{dk} merupakan bobot dari term t_k dalam dokumen *D*, dan w_{qk} merupakan bobot term t_k dalam dokumen *Q*.

Metode *tf*idf* adalah cara untuk memberikan bobot hubungan suatu kata (*term*) terhadap dokumen. Proses pembobotan ini membentuk bobot $\omega_D(t_i)$ melibatkan tiga tahapan yaitu *TF*, *IDF*, dan *TF*IDF* dengan persamaan :

$$TF = (t_i, D) \quad (5)$$

$$IDF = \log \frac{N}{df(t_i, D)} + 1 \quad (6)$$

$$\omega_D(t_i) = \frac{tf(t_i, D) \times \log \frac{N}{df(t_i)} + 1}{\sqrt{\sum_t (tf(t_i, D) \times \log \frac{N}{df(t_i)} + 1)^2}} \quad (7)$$

Penilaian tingkat kemiripan *query-document* bisa didapatkan dengan membandingkan antara kedua vektor yang sesuai dengan persamaan 8.

$$Cos(Q, D) = \sum_{r=1}^M \omega_Q(t_i) \times \omega_D(t_i) \quad (8)$$

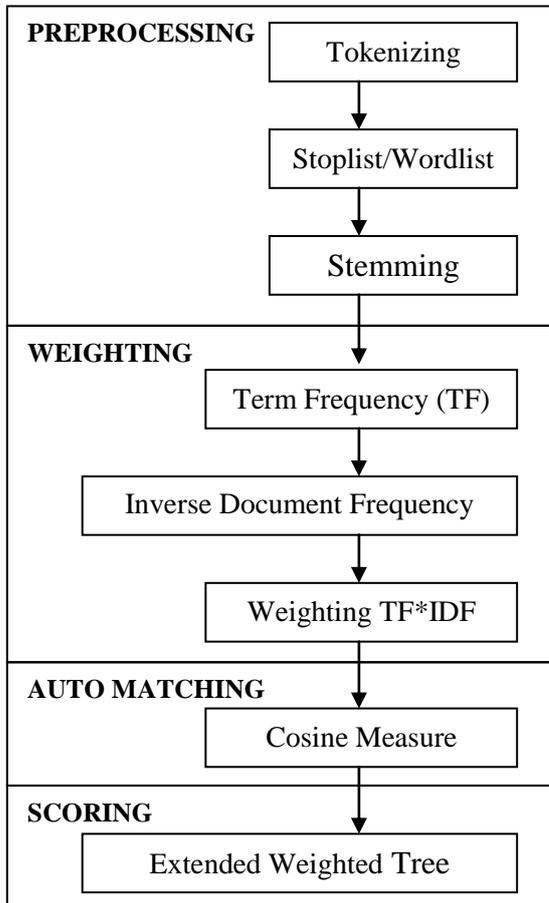
Algoritma *Extended Weighted-Tree Similarity* (disingkat *EWTS*) adalah algoritma yang digunakan untuk mengukur tingkat kemiripan (*similarity*) antara dua buah *tree*. Penentuan tingkat *similarity* dilakukan dengan menentukan *distance* atau level *similarity*. Dua buah obyek dikatakan memiliki tingkat kemiripan tinggi, ditandai dengan nilai *distance* yang mendekati 0 atau nilai *similarity* yang mendekati angka 1. Pada penelitian ini ditentukan tingkat *similarity* antara dua buah *tree* dengan algoritma *EWTS* dimana *similarity* akan bernilai 0 apabila keduanya tidak memiliki kesamaan sama sekali dan bernilai 1 apabila benar-benar memiliki kesamaan. *Tree* pada metode ini dibagi menjadi 3 bagian yaitu *node-labelled*, *arc-labelled*, *arc-weighted tree*. Dimana *node-labelled* adalah label atau nama atau identitas yang ada pada setiap *node*, *arc-labelled* adalah label atau nama atau identitas pada setiap cabang *node*, *arc-weighted* adalah bobot pada setiap cabang *node*. Menurut Sulisty (2008), Sarno (2008) *EWTS* memiliki 3 fungsi yaitu *treemap*, *treemap* dan *treeplicity*. Fungsi *treemap* secara rekursif membandingkan dua *list* *l* dan *l'*, masing-masing adalah himpunan *arc* pada setiap *level* yang *root*-nya, dimana rata-rata bobot dapat dihitung dengan persamaan 9.

$$\sum(A(S_i)(w_i + w'_i)/2) \quad (9)$$

3. METODE PENELITIAN

Preprocessing dilakukan dengan text mining untuk mendapatkan terms relevan, Weighting dilakukan dengan *TF*IDF* untuk

mendapatkan bobot dari term, Auto Matching dilakukan dengan Cosine Measure pada bagian (*sub tree*) dokumen pedoman dan proposal, Scoring dengan EWTS untuk menghitung kemiripan total dua buah tree.



Gambar 1. Metode Penelitian *Preprocessing, Weighting, Auto Matching, Scoring*

4. HASIL DAN PEMBAHASAN

Data yang digunakan terdiri dari data query (Q), data pelatihan (L) dan data pengujian (U). Data query berupa daftar term relevan dari masing-masing bagian dokumen digital sesuai skim PKM dari buku pedoman PKM 2015 dan 2016. Data pelatihan berupa dokumen proposal usulan mahasiswa yang lolos pendanaan dari Dirjen Belmawa Kemenristekdikti. Data pengujian berupa daftar dokumen proposal mahasiswa yang diunggah di

laman simbelmawa.ristekdikti.go.id. Penelitian ini menggunakan data pelatihan (L) sebanyak 8 proposal dan data pengujian (U) sebanyak 24 proposal dalam dua tahun terakhir (usulan tahun 2015 dan 2016) sehingga jumlah keseluruhan dokumen sebanyak 32 proposal ditunjukkan dalam Tabel 1. Dimana rasio jumlah data perbandingan menggunakan 1 : 3 (satu data pelatihan berbanding tiga data pengujian).

Tabel 1. Rekapitulasi Jumlah Data Pelatihan (L) dan Data Pengujian (U)

Tahun Usulan	PKMP		PKMK		PKMM		PKMT		Total	
	L	U	L	U	L	U	L	U	L	U
2015	2	6	1	3	1	3	1	3	5	15
2016	1	3	1	3	1	3	0	0	3	9
Total	3	9	2	6	2	6	1	3	8	24

Sumber Data :
Surat Penugasan PKM 5 Bidang Tahun 2015, 2016. Laman <http://simlitabmas.dikti.go.id/> dan <http://simbelmawa.ristekdikti.go.id>

Pengujian dilakukan dengan *auto matching* kemiripan antar bagian yang sama (*sub tree*) dari dokumen pedoman PKM (Query Q) dan dokumen proposal mahasiswa dalam setiap skim PKM (Dokumen D). Bagian (*sub tree*) dokumen meliputi **Daftar Isi, Ringkasan, Bab 1, Bab 2, Bab 3, dan Bab 4**. Nilai kemiripan *cosine* antar bagian yang sama akan dihitung total rata-ratanya untuk menentukan kemiripan antara dua dokumen. Perhitungan scoring total rata-rata tersebut dengan fungsi *treemap* dari algoritma Extended Weighted Tree Similarity.

Prototipe perangkat lunak sistem pendeteksian dokumen digital menerima input berupa data format teks (*.txt) dari bagian-bagian (*sub tree*) dari dokumen proposal dan dokumen pedoman PKM. Dimana file dokumen format pdf dikonversi dalam format text (*.txt) menggunakan laman pdftotext.com/id.

Hasil dari *auto matching* antar bagian (*sub tree*) dokumen dari data penelitian dan pedoman PKM menghasilkan nilai kemiripan ditunjukkan dalam Tabel 2 sampai Tabel 8 berikut ini.

Tabel 2. Hasil Similarity Antar Bagian Query Q dan Dokumen D (PKMP 2015)

Bagian Dokumen Q	Nilai Cosine Measure Dokumen D			
	PKMP1	PKMP2	PKMP3*	PKMP4*
Daftar Isi	0,540	0,601	0,714	0,539
Ringkasan	0,323	0,409	0,374	0,428
Bab 1	0,498	0,479	0,524	0,428
Bab 2	0,425	0,337	0,444	0,371
Bab 3	0,444	0,470	0,417	0,308
Bab 4	0,000	0,627	0,612	0,558
Rata-Rata	0,3464	0,48717	0,51417	0,43867

Keterangan : * Data Pelatihan

Tabel 3. Hasil Similarity Antar Bagian Query Q dan Dokumen D (PKMP 2015)

Bagian Dokumen Q	Nilai Cosine Measure Dokumen D			
	PKMP5	PKMP6	PKMP7	PKMP8
Daftar Isi	0,574	0,415	0,814	0,538
Ringkasan	0,401	0,409	0,425	0,471
Bab 1	0,507	0,454	0,47	0,528
Bab 2	0,405	0,371	0,322	0,401
Bab 3	0,418	0,482	0,447	0,380
Bab 4	0,584	0,467	0,486	0,377
Rata-Rata	0,4815	0,433	0,494	0,44917

Tabel 4. Hasil Similarity Antar Bagian Q dan D (PKMK 2015)

Bagian Dokumen Q	Nilai Cosine Measure Dokumen D			
	PKMK1*	PKMK2	PKMK3	PKMK4
Daftar Isi	0,770	0,621	0,639	0,796
Ringkasan	0,236	0,236	0,425	0,425
Bab 1	0,538	0,462	0,535	0,626
Bab 2	0,596	0,298	0,428	0,506
Bab 3	0,464	0,366	0,412	0,492
Bab 4	0,541	0,545	0,416	0,573
Rata-Rata	0,52417	0,42133	0,47583	0,56967

Tabel 5. Hasil Similarity Antar Bagian Q dan D (PKMM 2015)

Bagian Dokumen Q	Nilai Cosine Measure Dokumen D			
	PKMM1	PKMM2	PKMM3	PKMM4*
Daftar Isi	0,710	0,720	0,676	0,749
Ringkasan	0,547	0,216	0,311	0,311
Bab 1	0,442	0,505	0,358	0,462
Bab 2	0,718	0,717	0,749	0,767
Bab 3	0,499	0,424	0,568	0,575
Bab 4	0,533	0,511	0,528	0,557
Rata-Rata	0,57483	0,5155	0,53167	0,57017

Tabel 6. Hasil Similarity Antar Bagian Q dan D (PKMT 2015)

Bagian Dokumen Q	Nilai Cosine Measure Dokumen D			
	PKMT1*	PKMT2	PKMT3	PKMT4
Daftar Isi	0,657	0,503	0,527	0,543
Ringkasan	0,311	0,216	0,288	0,311
Bab 1	0,337	0,275	0,340	0,443
Bab 2	0,257	0,232	0,343	0,423
Bab 3	0,478	0,550	0,550	0,378
Bab 4	0,580	0,559	0,532	0,453
Rata-Rata	0,436667	0,38917	0,430	0,42517

Tabel 7. Hasil Similarity Antar Bagian Q dan D (PKMP 2016)

Bagian Dokumen Q	Nilai Cosine Measure Dokumen D			
	PKMP1	PKMP2	PKMP3	PKMP4*
Daftar Isi	0,762	0,558	0,541	0,729
Bab 1	0,416	0,485	0,504	0,443
Bab 2	0,425	0,388	0,292	0,332
Bab 3	0,323	0,406	0,417	0,437
Bab 4	0,575	0,601	0,601	0,432
Rata-Rata	0,5002	0,4876	0,471	0,4746

Tabel 8. Hasil Similarity Antar Bagian Q dan D (PKMK 2016)

Bagian Dokumen Q	Nilai Cosine Measure Dokumen D			
	PKMK1	PKMK2	PKMK3*	PKMK4
Daftar Isi	0,753	0,821	0,751	0,804
Bab 1	0,589	0,468	0,568	0,536
Bab 2	0,597	0,413	0,394	0,542
Bab 3	0,394	0,394	0,365	0,486
Bab 4	0,558	0,525	0,637	0,319
Rata-Rata	0,5782	0,5242	0,543	0,5374

Tabel 9. Hasil Similarity Antar Bagian Q dan D (PKMM 2016)

Bagian Dokumen Q	Nilai Cosine Measure Dokumen D			
	PKMM1	PKMM2*	PKMM3	PKMM4
Daftar Isi	0,688	0,846	0,680	0,770
Bab 1	0,428	0,471	0,384	0,468
Bab 2	0,645	0,667	0,635	0,631
Bab 3	0,519	0,499	0,451	0,505
Bab 4	0,560	0,570	0,519	0,522
Rata-Rata	0,568	0,6106	0,5338	0,5792

Dari hasil pengujian kemiripan antar bagian, maka dihasilkan nilai scoring dari total rata-rata setiap tree dokumen proposal ditunjukkan dalam Tabel 10 untuk usulan tahun 2015, dan Tabel 11 untuk usulan tahun 2016.

Tabel 10. Nilai Scoring Sort By Descending (PKM 2015)

Skim	Proposal PKM Tahun 2015, Tertinggi							
	1	2	3	4	5	6	7	8
PKM P	0,514	0,487	0,481	0,449	0,433	0,438	0,494	0,3464
PKMK	0,569	0,524	0,475	0,421				
PKMM	0,574	0,570	0,531	0,515				
PKMT	0,436	0,430	0,425	0,389				

Tabel 11. Nilai Scoring Sort By Descending (PKM 2016)

SKIM	Proposal PKM Tahun 2016, Tertinggi			
	1	2	3	4
PKMP	0,5002	0,4876	0,4746	0,471
PKMK	0,5782	0,543	0,5374	0,5242
PKMM	0,6106	0,5792	0,568	0,5338

Berdasarkan Tabel 10 dan Tabel 11 disimpulkan bahwa prototipe perangkat lunak telah menghasilkan nilai scoring akhir antara dokumen proposal mahasiswa dan pedoman PKM. Data pelatihan dari proposal yang lolos pendanaan (tercetak tebal) tidak selalu memiliki nilai tertinggi dibandingkan dengan data pengujian, sehingga kepatuhan terhadap pedoman harus diikuti oleh kualitas dari proposal. Sebagian besar data pelatihan memiliki nilai tertinggi pertama dan kedua, sehingga kepatuhan terhadap pedoman tetap menjadi pertimbangan dalam menentukan proposal yang didanai

5. KESIMPULAN DAN SARAN

Kesimpulan dari penelitian dengan judul Pengembangan Sistem Deteksi Kesesuaian Dokumen Proposal Program Kreativitas Mahasiswa Dengan Metode *Extended Weighted Tree Similarity* yaitu : (1) Prototipe perangkat lunak sistem pendeteksian kesesuaian dokumen telah dapat menghasilkan nilai *auto matching* kemiripan antar bagian (*sub tree*) yang sama dari dokumen, dan menghasilkan nilai *scoring* kesesuaian total rata-rata tree dokumen proposal dan dokumen pedoman PKM, (2) Hasil akhir dari scoring

menunjukkan bahwa data pelatihan dari proposal yang lolos pendanaan tidak selalu memiliki nilai scoring tertinggi dibandingkan dengan data pengujian dari dokumen proposal yang tidak lolos, sehingga kepatuhan terhadap pedoman harus diikuti oleh kualitas dari proposal PKM yang diusulkan, (3) Sebagian besar data pelatihan memiliki nilai tertinggi pertama dan kedua, sehingga kepatuhan terhadap pedoman tetap menjadi pertimbangan dalam menentukan proposal didanai.

Penelitian ini dapat dikembangkan dengan mempertimbangkan kontekstual dari substansi isi proposal yang diusulkan. Bagian (*sub tree*) dokumen dapat diperluas dengan bagian yang lebih kecil sehingga dapat menghasilkan informasi lebih detail tentang ketidaksesuaian atau kekurangan yang harus dilengkapi. Disamping itu prototipe perangkat lunak perlu dikembangkan dalam antar muka yang *user friendly* agar dapat memudahkan pengguna dan hasilnya dapat divisualisasikan dalam bentuk yang lebih mudah dipahami

DAFTAR PUSTAKA

- Sarno, Riyanarto; Rahutomo, Faisal. (2008), Penerapan Algoritma Weighted Tree Similarity Untuk Pencarian Semantik, JUTI. Vol 7 No. 1 : 35-42.
- Sulistyo, Wiwin; Sarno, Riyanarto. (2008), Pemodelan Kesesuaian Dokumen Sekuriti Manajemen Aset Teknologi Informasi Menggunakan Algoritma Extended Weighted Tree Similarity, Prosiding Seminar Nasional Sistem & Teknologi Informasi (SNASTI), Surabaya : 267-375.
- Sulistyo, Wiwin; Sarno, Riyanarto. (2008), Auto Matching Antar Dokumen Dengan Metode Cosine Measure, Seminar Nasional Teknologi

Informasi dan Komunikasi”,
Indonesia.

Suharso, Wiwik. (2015), Identifikasi
Kebutuhan Non-Fungsional Dalam
Spesifikasi Tekstual Berdasarkan
Atribut ISO/IEC 9126, Prosiding
Seminar Nasional Manajemen
Teknologi XXII, MMT ITS Surabaya.

Pedoman Program Kreativitas Mahasiswa
2015, Direktorat Jenderal
Pembelajaran dan Kemahasiswaan
(Belmawa), Kementerian Riset,
Teknologi dan Pendidikan Tinggi.

Pedoman Program Kreativitas Mahasiswa
2016, Direktorat Jenderal
Pembelajaran dan Kemahasiswaan
(Belmawa), Kementerian Riset,
Teknologi dan Pendidikan Tinggi.

Laporan Tim PKM 2016, Bagian
Kemahasiswaan Universitas
Muhammadiyah Jember.

Surat Penugasan PKM 5 Bidang Tahun
2015, 2016.

SK Rektor Nomor :
0203/KEP/II.3.AU/F/2016.

<http://simlitabmas.dikti.go.id/> (diakses 4
Mei 2016)

<http://simbelmawa.ristekdikti.go.id/> (1 Juli
2017)