

PENDETEKSIAN ISTILAH BERBEDA PADA DOKUMEN SPESIFIKASI KEBUTUHAN PERANGKAT LUNAK (SKPL)

Fitria Vera Sari Sahadi¹, Daniel Oranova Siahaan², Umi Laili Yuhana³

¹²³Teknik Informatika, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh
Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

Email: angel87_ria@yahoo.co.id¹, daniel@if.its.ac.id², yuhana97@gmail.com³

Abstrak. Dalam membuat dan menganalisa suatu dokumen SKPL diperlukan ketelitian dalam penyusunan SKPL. Dokumen SKPL harus jelas, lengkap, dan tidak ambigu. Istilah berbeda merupakan varian dari noise dalam suatu dokumen SKPL. Penelitian ini berfokus mengenai istilah berbeda yang dikenali sebagai sinonim pada pasangan kalimat dalam dokumen SKPL. Sinonim merupakan kata yang memiliki istilah yang berbeda dan bermakna sama. Perancangan metode terdiri dari proses pelatihan dan pengujian, yaitu prapemrosesan, menghitung kemiripan semantik dari pasangan kalimat dan menentukan nilai threshold. Sedangkan nilai Kappa untuk mengetahui perancangan metode dapat diandalkan dan digunakan untuk mendeteksi ketidakkonsisten istilah pada dokumen SKPL. Hasilnya adalah pasangan kalimat yang terdeteksi sebagai istilah berbeda.

Kata Kunci: dokumen SKPL, fakta, istilah berbeda, kemiripan semantic

Banyak permasalahan dalam pengembangan perangkat lunak berakar pada keterbatasan pemahaman pengembang akan kebutuhan pengguna akan perangkat lunak yang dibangun (Siahaan, 2010). Hal ini bisa disebabkan oleh keterbatasan data dan informasi yang didapatkan pada waktu proses pengumpulan, penganalisaan, penspesifikasian, verifikasi, dan validasi kebutuhan dari perangkat lunak yang hendak dibangun. Keterbatasan data dan informasi ini terjadi karena rendahnya kepedulian terhadap proses pelaksanaan rekayasa kebutuhan (Requirement Engineering).

Sebuah spesifikasi kebutuhan perangkat lunak dikatakan konsisten, jika tidak ada kebutuhan yang konflik. Ada tiga jenis konflik, yaitu :

- *Multiple descriptor (Banyak Deskripsi),*

Ketidakkonsistenan ini terjadi ketika dua atau lebih kata-kata yang digunakan untuk referensi item yang sama, yaitu istilah-istilah yang digunakan secara bergantian.

- *Opposing physical requirements (Tidak Sesuai Kebutuhan Fisik),*

Ketidakkonsistenan ini terjadi ketika dalam mendeskripsikan objek dunia nyata secara tidak konsisten, misalnya salah satu indikator dalam peringatan dalam sistem ditunjukkan dengan warna oranye dan indikator yang satunya lagi ditunjukkan dengan warna merah.

- *Opposing functional requirements (Tidak Sesuai dengan Kebutuhan Fungsional),*

Ketidakkonsistenan ini terjadi ketika dalam karakteristik fungsional terjadi konflik misal, tampilkan fungsi X ketika fungsi A dan B dijalankan atau tampilkan fungsi X ketika fungsi A atau B telah dijalankan.

Penelitian ini akan berfokus pada Multiple descriptor (Banyak Deskripsi) dari ketidakkonsistenan spesifikasi kebutuhan.

Rumusan masalah dalam penelitian ini adalah bagaimana membuat kerangka kerja untuk mendeteksi penggunaan istilah secara tidak konsisten dalam dokumen SKPL.

Sedangkan tujuan penelitian ini adalah membangun sebuah kerangka kerja pendeteksian ketidakkonsistenan penggunaan istilah dalam dokumen SKPL. Manfaat yang diusulkan adalah membantu spesifikator dalam mendeteksi ketidakkonsistenan penggunaan istilah dalam dokumen SKPL. Kontribusi dari penelitian ini adalah menyediakan sebuah kerangka kerja untuk mendeteksi ketidakkonsistenan penggunaan istilah dalam dokumen SKPL.

Istilah berbeda merupakan bagian dari perubahan dan pengembangan arti kata-kata. Kosakata terus berubah dan berkembang mengikuti perkembangan zaman. Kosakata dalam suatu bahasa yang hidup dalam sekelompok masyarakat atau kehidupan sehari-

hari tidak ada yang tetap. Sehubungan dengan ruang lingkup kosakata, Dipodjojo (1984, 22) mengatakan bahwa istilah berbeda dapat dikenali menjadi homonim, homofon dan homograph. Namun pada penelitian ini dilakukan penelitian mengenai istilah berbeda yang dikenali sebagai sinonim. Sinonim merupakan kata yang memiliki istilah yang berbeda dan bermakna sama. Alasan menggunakan sinonim sebagai acuan dalam kerangka kerja adalah menggunakan istilah berbeda dan bermakna sama seringkali ditemukan dalam dokumen SKPL. Sehingga diperlukan suatu penelitian untuk mengenali sinonim antar kata dalam dokumen SKPL agar tidak terjadi kesulitan dalam pemahaman istilah yang bermakna sama.

Istilah berbeda dapat dikenali dengan menganalisa semantik dengan mengenali ciri sinonim yaitu kedua istilah yang merupakan sinonim dapat digunakan pada kalimat dalam konteks yang sama. Contoh kalimat yang mengandung sinonim dalam dokumen SKPL :

K1 = The software to be made is called Secure Digital Archiving System (SDA)

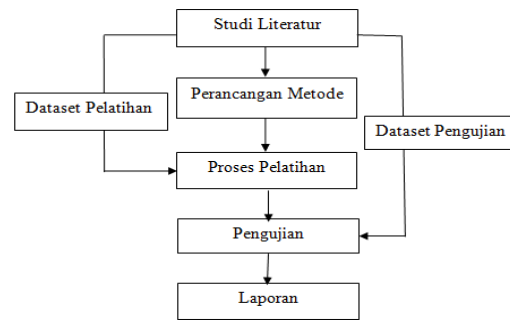
K2 = The main aim of the SDA is to manage the storage of digital documents in such a way

Dari contoh kalimat di atas istilah yang mengandung sinonim terdapat pada K1 adalah Archiving sedangkan pada K2 adalah documents. Hal ini yang dimaksudkan dengan pemakaian istilah yang tidak konsisten.

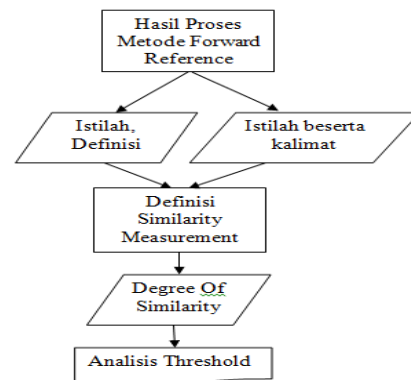
Kerangka kerja yang diusulkan pada penelitian ini adalah mendeteksi istilah berbeda pada pasangan kalimat yang mengandung fakta sama dalam dokumen SKPL. Kerangka kerjanya terdiri dari tahap-tahap ekstraksi fakta, prapemrosesan yaitu tokenisasi, penghapusan stopword, stemming,. Kemudian menghitung nilai similarity pasangan kalimat yang mengandung fakta sama.

I. Metodologi

Data set URD dan SRS diambil dari berbagai sumber dengan pengambilan secara acak. Dapat dilihat pada tabel 1.

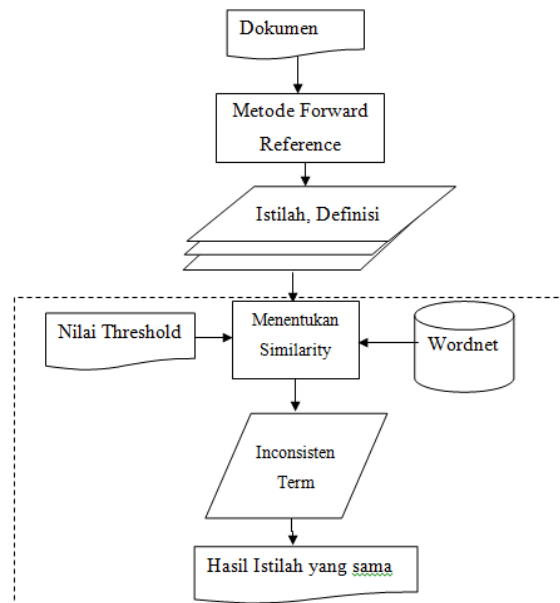


Gambar 1. Metode penelitian



Gambar 2. Proses pelatihan

Perancangan kerangka kerja terdiri dari dua proses yaitu proses pelatihan dan proses pengujian. Tujuan proses pelatihan adalah untuk mendapatkan nilai threshold. Sedangkan proses pengujian dilakukan untuk menentukan data set yang merupakan istilah yang konsisten atau tidak berdasarkan nilai threshold.



Gambar 3. Perancangan kerangka kerja

Proses perancangan kerangka kerja yang akan dibuat terdapat pada bagan yang ditandai yaitu menghitung nilai kemiripan dengan rumus kemiripan. Menentukan nilai threshold untuk menentukan ketidakkonsistenan istilah, Inconsistency term adalah proses dimana menentukan apakah istilah tersebut sudah konsisten atau tidak dan output berupa hasil istilah yang sama. Sedangkan bagan yang tidak ditandai adalah proses dari metode forward reference oleh Umami (2012) dapat dilihat pada gambar 3.

II. Hasil dan Pembahasan
Skenario Pengujian

Berikut adalah langkah-langkah pengujian yang dilakukan.

1. Pengumpulan data uji.

Untuk proses pengujian, data uji diambil dari 30 dokumen SKPL berbahasa Inggris.

2. Penyiapan data uji.

Setiap dokumen SKPL diproses untuk mengekstraksi istilah dan definisi atau kalimat yang ada pada dokumen tersebut. Setiap pasangan istilah-definisi/kalimat dikumpulkan dalam suatu tabel/form yang nanti akan diperiksa oleh para ahli bahasa. Tabel 1. Menunjukkan contoh tabel/form tersebut. Kolom dokumen adalah nama dokumen SKPL atau URD. Kolom kalimat adalah pasangan kalimat K1 dan kalimat K2 yang diambil dari dokumen SKPL dan URD secara acak. Setiap pasangan kalimat telah di periksa oleh ahli bahasa dengan dua kemungkinan jawaban : istilah yang sama atau istilah yang tidak sama. Kolom pengamatan Expert (ahli bahasa) adalah pernyataan dari ahli bahasa yang menyatakan bahwa pasangan kalimat adalah istilah yang

sama atau istilah yang tidak sama. Pada dataset no 1, ahli bahasa menyatakan istilah yang sama.

3. Peniadaan Ahli Bahasa

Ahli bahasa yang dilibatkan sebanyak 5 ahli bahasa. Ahli bahasa yang dilibatkan mempunyai pengalaman di bidang bahasa Inggris tetapi tidak mempunyai latar belakang teknik informatika. Karena setiap ahli memiliki subyektifitas yang independen, maka ada peluang satu atau lebih ahli memiliki pendapat yang menyimpang di banding dengan kebanyakan ahli yang lain. Untuk itu dilakukan peniadaan pencilan (ahli bahasa yang berbeda pendapatnya dari kebanyakan ahli yang lain), dimana ahli bahasa yang dianggap memiliki pendapat yang menyimpang jauh tidak diikutsertakan dalam perhitungan metode DAO. Peniadaan ahli bahasa dilakukan dengan metode MeanDist.

Langkah-langkahnya sebagai berikut :

- Menghitung nilai graf dengan nilai jarak euclidion yaitu 0-1

Rumus: $V1-V2=$

$$\sqrt{(X1 - Y1)^2 + (X2 - Y2)^2 + (X3 - Y3)^2 + \dots + (Xn - Yn)^2}$$

- Menghitung nilai rata-rata

- 1 - 2 = > 2 - 1 =
- 1 - 3 = > 2 - 3 =
- 1 - 4 = > 2 - 4 =
- 1 - 5 = > 2 - 5 =

- 3 - 1 = > 4 - 1 =
- 3 - 2 = > 4 - 2 =
- 3 - 4 = > 4 - 3 =
- 3 - 5 = > 4 - 5 =

Tabel 1 Contoh dataset yang digunakan dalam skenario pengujian

Dokumen	Libra_srs-uji1		
Ahli Bahasa			
No. Urut	Kalimat	Hasil pengamatan	
1.	K ₁ = This software is being developed to for customer Xuan Gu K ₂ = The TA1 Project is a new product that is part of a larger, more complete product for our customer	Ya	
.		.	
.		.	
.		.	
276.	K ₁ = Linux is an open source operating system, with extensive documentation and user support K ₂ = A grid is an infrastructure that couples resources such as computers (PCs, workstations, clusters and so on)		Tidak

- Membuat tabel graf
Untuk membuat tabel 2. graf, nilai diurut sesuai nilai terbesar seperti berikut:

Tabel 2. Graf, nilai yang diurut

	E1	E2	E3	E4	E5
E1	0	0	0	5	3.74
E2	0	0	0	5	3.74
E3	0	0	0	5	3.74
E4	5	5	5	0	6.08
E5	3.7	3.74	3.74	6.08	0

Dari hasil tabel diatas maka ditentukan nilai L1,L2,L3,L4 dan L5. Dengan cara menjumlahkan nilai dari E1-E2, E1-E3, E1-E4, E1-E5 dan dibagi 4. Selanjutnya nilai terbesar pertama dikurangi dengan terbesar kedua, nilai terbesar kedua dikurangi dengan nilai terbesar ketiga, nilai terbesar ketiga dikurangi dengan nilai terbesar keempat dan nilai terbesar keempat dikurangi dengan nilai terbesar kelima.

$$\begin{aligned} L1 &= 2.19 & L4-L5 &= 0.94 \\ L2 &= 2.19 & L5-L1 &= 2.14 \\ L3 &= 2.19 & L1-L2 &= 0 \\ L4 &= 5.27 & L2-L3 &= 0 \\ L5 &= 4.33 & & \end{aligned}$$

Dari hasil tabel 3 di tentukan nilai $t=0,5$ dan ahli bahasa yang tidak diikuti sertakan adalah L5-L1 yaitu ahli bahasa.

Tabel 3. Hasil nilai t dan peniadaan ahli bahasa

t=	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
T=	0.2	0.43	0.64	0.86	1.07	1.28	1.5	1.71	1.93

4. Pengumpulan data hasil dari metode.

Dokumen SKPL yang telah dinilai oleh ahli bahasa kemudian dinilai oleh perangkat lunak yang dibangun berdasarkan metode yang diusulkan. Hasil dari perangkat lunak kemudian dibandingkan dengan hasil dari para ahli bahasa. Perbandingan ini bertujuan mendapat tingkat kesepakatan antara para ahli dan metode yang diusulkan.

5. Analisis hasil

Hasil perhitungan tingkat kesepakatan tersebut kemudian dianalisis untuk menemukan fakta-fakta atau temuan-temuan baru.

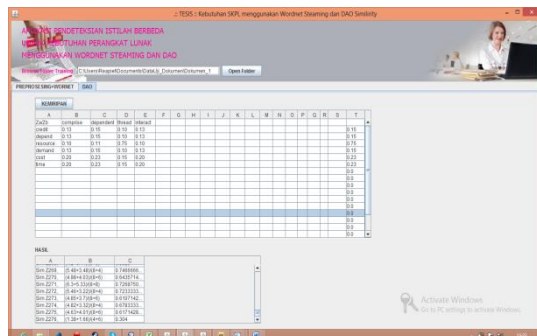
Dokumen yang diuji sebanyak 30 dokumen, setiap dokumen mempunyai banyak pasangan kalimat yang berbeda-beda, ada yang

tidak konsisten dan ada yang konsisten dalam penggunaan istilah berbeda. Jumlah total pasangan kalimat masing-masing juga berbeda. Seperti di tunjukan pada tabel 4.6. Kolom “ya” adalah banyak kesepakatan antara ahli dan tools yang menyatakan 2 istilah tersebut merupakan istilah yang sama(penggunaan istilah yang tidak konsisten), sedangkan kolom “tidak” adalah banyak kesepakatan antara ahli dan tools yang menyatakan 2 istilah tersebut tidak sama(konsisten). Dan pada kolom jumlah total adalah jumlah pasangan kalimat yang diuji.

Tabel 4. Hasil data pengujian dari metode dengan nilai threshold 0.6

Nama Dokumen	Total		Jumlah Total
	Ya	Tidak	
Dokumen 1	235	41	276
Dokumen 2	172	59	231
Dokumen 3	135	36	171
Dokumen 4	109	11	120
Dokumen 5	72	19	91
Dokumen 6	81	39	120
Dokumen 7	82	23	105
Dokumen 8	215	61	276
Dokumen 9	463	98	561
Dokumen 10	187	23	210
Dokumen 11	191	109	300
Dokumen 12	164	46	210
Dokumen 13	249	51	300
Dokumen 14	557	38	595
Dokumen 15	76	29	105
Dokumen 16	401	194	595
Dokumen 17	692	88	780
Dokumen 18	190	41	231
Dokumen 19	142	48	190
Dokumen 20	179	11	190
Dokumen 21	231	45	276
Dokumen 22	157	33	190
Dokumen 23	225	6	231
Dokumen 24	138	15	153
Dokumen 25	18	3	21
Dokumen 26	238	15	253
Dokumen 27	205	48	253
Dokumen 28	140	13	153
Dokumen 29	135	36	171
Dokumen 30	126	45	171

Setelah Data set di nilai oleh ahli bahasa selanjutnya dinilai oleh perangkat lunak yang di bangun. Proses yang akan dilakukan adalah perhitungan pencarian nilai kemiripan menggunakan metode kemiripan semantik berbasis wordnet (Dao et al, 2006) pada gambar 4. Tetapi untuk melakukan perhitungan pencarian penggunaan istilah berbeda, terlebih dahulu melakukan proses pre-posesing. Proses pre-posesing terdiri dari ekstraksi dokumen, tokenizing, penghapusan stopword dan steaming.



Gambar 4. Kakas bantu menghitung nilai kemiripan dengan metode kemiripan semantik berbasis Wordnet

Pada tabel kemiripan berupa hasil menghitung nilai kemiripan secara bergantian menggunakan metode kemiripan semantik berbasis Wordnet (Dao et al, 2006). Kakas

bantu menghitung nilai kemiripan semantik di tunjukan pada gambar 4.

Tabel hasil adalah nilai kemiripan pasangan kalimat yang diperoleh dengan menghitung nilai rata-rata dari setiap pasangan kalimat dan istilahnya masing-masing. Nilai pasangan kalimat ini yang menentukan apakah pasangan kalimat tersebut merupakan istilah yang konsisten atau tidak konsisten.

Perhitungan hasil kesepakatan

Hasil kesepakatan yang baik antara tools dengan ahli bahasa pada 30 dokumen yang diuji terdapat pada ahli bahasa ke 4 dengan nilai kappa rata-rata tertinggi 0.721868229 pada nilai threshold 0.6 dan terendah -0.1544. Sedangkan nilai rata-rata keseluruhan dari kesepakatan antara 4 ahli bahasa dengan tools, nilai kappa tertingginya adalah 0.485304. Di tunjukan pada tabel 6 dan 7.

Analisis Hasil Pengujian

Berdasarkan interpretasi nilai kappa 0.721868229 adalah Banyak (substantial agreement), dengan ini berarti terdapat banyak kesepakatan antara ahli bahasa dan kerangka kerja. Dapat di simpulkan bahwa kerangka kerja yang diajukan dapat digunakan untuk mendeteksi istilah yang sama pada dokumen SKPL.

Tabel 6. Hasil kesepakatan tools dan ahli bahasa pada 30 dokumen

Nama Dokumen	Ahli	Nilai TH	Hasil kesepakatan		Nilai TH
			Tertinggi	Terendah	
Dokumen 1	4	0.6	0.808917197	0.006897	1
Dokumen 2	4	0.6	0.334164	-0.19573	1
Dokumen 3	4	0.6	0.874172185	-0.20362	1
Dokumen 4	4	0.6	0.565807327	-0.07402	1
Dokumen 5	4	0.6	0.85270314	-0.18747	1
Dokumen 6	4	0.6	0.887308827	-0.2047	1
Dokumen 7	4	0.6	0.795967938	-0.187	1
Dokumen 8	4	0.5	0.235955	-0.17195	0.7
Dokumen 9	4	0.6	0.929025352	-0.16713	1
Dokumen 10	4	0.6	0.929025352	-0.16713	1
Dokumen 11	4	0.6	0.922715981	-0.16368	1
Dokumen 12	4	0.6	0.805387372	-0.1793	1
Dokumen 13	4	0.6	0.482066098	-0.1804	1
Dokumen 14	4	0.6	0.528469291	-0.10601	1
Dokumen 15	4	0.7	0.647597254	-0.12	1
Dokumen 16	4	0.6	0.937151428	-0.20237	1
Dokumen 17	4	0.6	0.792727273	-0.13362	1
Dokumen 18	4	0.6	0.834694973	-0.1617	1
Dokumen 19	4	0.6	0.895284752	-0.20563	1
Dokumen 20	4	0.6	0.669303797	-0.11189	1
Dokumen 21	4	0.6	0.719192577	-0.17821	1
Dokumen 22	4	0.6	0.780802953	-0.19302	1
Dokumen 23	4	0.6	0.518893187	-0.03454	1
Dokumen 24	4	0.6	0.658545657	-0.1412	1
Dokumen 25	4	0.6	0.631578947	-0.2069	1
Dokumen 26	4	0.6	0.650515354	-0.09176	1
Dokumen 27	4	0.6	0.742651	-0.16739	1
Dokumen 28	4	0.6	0.56626506	-0.10769	1
Dokumen 29	4	0.6	0.862429606	-0.19145	1
Dokumen 30	4	0.6	0.796728	14 -0.20357	1
Rata-rata			0.721868229	-0.1544	

Walaupun kerangka kerja telah bekerja dengan baik pada nilai threshold 0,6 tetapi ada ketidaksepakatan kerangka kerja dan ahli

bahasa dengan nilai terendah pada dokumen 2 dan 8 yaitu 0.334164 dan 0.235955, ditunjukkan pada tabel 8. Penyebabnya adalah ada pasangan kalimat yang memiliki kesepakatan yang berbeda. Pada dokumen 8 sebanyak 72, dimana 26 pasang kalimat kerangka kerja menyatakan istilah yang sama dan ahli bahasa menyatakan tidak yang disebabkan karena nilai kemiripan pasangan lebih tinggi daripada nilai threshold 0,6. Dan 46 pasangan kalimat dimana kerangka kerja menyatakan istilah yang tidak sama dan ahli bahasa menyatakan istilah yang sama disebabkan karena nilai kemiripan pasangan kalimat yang lebih rendah dari nilai threshold 0,6.

Adanya ketidaksepakatan antara kerangka kerja dan ahli bahasa disebabkan karena nilai kemiripan pasangan kalimat lebih besar dari nilai threshold 0,6. Dimana kerangka kerja menyatakan istilah yang sama dan ahli bahasa menyatakan istilah yang tidak sama dapat dilihat pada Tabel 8.

1. Kolom kalimat adalah K1 dan K2 yang didapat dari dokumen 1 yaitu dokumen 8_HomeSubscriberServer_SRS_Document.
2. Dengan pengambilan kalimat dan istilah yang di hasilkan oleh metode Forward reference.

Tabel 8. Pencarian nilai kemiripan

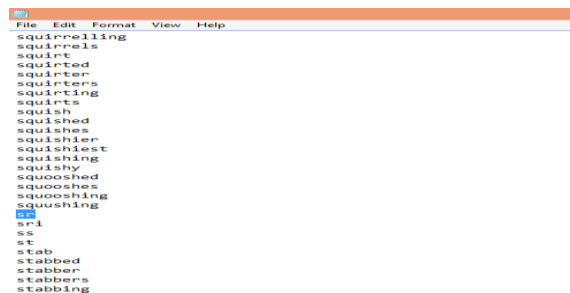
Dokumen	Kalimat	Kemiripan	Nilai kemiripan pasangan kalimat	Variabel kappa threshold
Uji dok 8 HomeSubscriberServer_SRS_Document.No 2	ZA: ip multimedia (im) subsystem cx and dx interfaces signalling flows and message contents ZB: communication integrity check and ciphering	Sim(ZA,ZB)	0.544	C

Nilai kemiripan yang rendah pada kerangka kerja ini disebabkan karena dua hal. Penyebab yang pertama konten Wordnet yang terbatas, sehingga ada hubungan atau makna kata yang tidak dikenali oleh kerangka kerja, atau contoh lebih khusus yang oleh kerangka kerja tetapi dengan makna lain, misalnya pada kalimat ZA,ZB. Hasil penghapusan stopwords dan stemming pada kalimat ZA,ZB adalah :

ZA = sr, describe, requirement, specification, libra, economy, driven, cluster, sheduler

ZB = time, pend, job, memory

Pada hasil diatas dapat dilihat bahwa kata SRS dideteksi sebagai sr. hal ini disebabkan karena kamus wordnet terdapat kata sr dan masih banyak lagi kata-kata yang tidak dapat dibaca oleh wordnet hal ini dapat mengacaukan nilai dan menyebabkan perhitungan kemiripan kalimat bekerja kurang maksimal. Ditunjukkan pada gambar 5 dan hasil yang lain pada gambar 6.



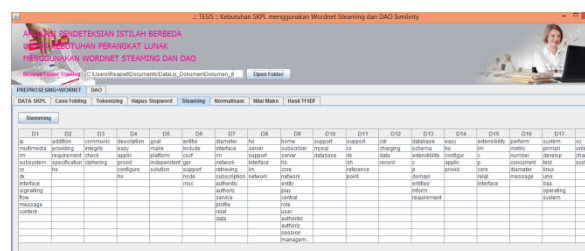
Gambar 5. Contoh kosakata pada wordnet

Tabel 7. Nilai rata-rata Kesepakatan tools dan ahli bahasa pada 30 dokumen

Nama Dokumen	Nilai TH	rata-rata		Nilai TH
		Tertinggi	Terendah	
Dokumen 1	0.6	0.284774	-0.15385	1
Dokumen 2	0.6	0.137637	-0.09374	0.7
Dokumen 3	0.3	0.758603	-0.19642	1
Dokumen 4	0.6	0.386644	-0.1509	1
Dokumen 5	0.6	0.23222	-0.201	0.9
Dokumen 6	0.6	0.3406	-0.203	1
Dokumen 7	0.6	0.553054	-0.183	1
Dokumen 8	0.6	0.138724	-0.12756	1
Dokumen 9	0.6	0.114886	-0.14937	0.9
Dokumen 10	0.6	0.17017	-0.139	1
Dokumen 11	0.6	0.711926	0.089968	0.8
Dokumen 12	0.6	0.746824	-0.19538	1
Dokumen 13	0.6	0.414621	-0.1845	1
Dokumen 14	0.6	0.448784	-0.01472	1
Dokumen 15	0.6	0.667305	-0.01853	1
Dokumen 16	0.6	0.71221	-0.28	1
Dokumen 17	0.6	0.715338	-0.022	1
Dokumen 18	0.6	0.519092	-0.1925	1
Dokumen 19	0.6	0.735699	-0.19485	1
Dokumen 20	0.6	0.314065	-0.10632	1
Dokumen 21	0.6	0.602427	-0.19268	1
Dokumen 22	0.6	0.725548	-0.20528	0.9
Dokumen 23	0.6	0.356436	-0.10179	1
Dokumen 24	0.6	0.45233	-0.19455	1
Dokumen 25	0.6	0.412612	-0.2069	1
Dokumen 26	0.6	0.393767	-0.15694	1
Dokumen 27	0.6	0.550885	-0.18903	1
Dokumen 28	0.6	0.557088	-0.12478	1
Dokumen 29	0.6	0.711341	-0.20058	1
Dokumen 30	0.6	0.693503	-0.14258	1
Rata-rata		0.485304	-0.14773	



Gambar 6. Hasil Preprocessing Dokumen 2



Gambar 7. Hasil preprocessing Dokumen 8

Penyebab yang sama terdapat pada dokumen 2 dan 8, terdapat banyak kata-kata yang tidak dapat dibaca oleh wordnet hal ini menyebabkan perhitungan kemiripan kalimat bekerja kurang maksimal. Di tunjukan pada gambar 6 dan 7.

Penyebab kedua adalah ada kemungkinan terjadi pemakaian istilah yang sama pada setiap pasangan kalimat, tetapi istilah yang sama tersebut terjadi pada sebagian kata dengan nilai kemiripan yang tinggi. Namun pada kata lainnya mempunyai nilai kemiripan yang sangat rendah. Hal ini menyebabkan perhitungan nilai kemiripan kalimat akan menghasilkan nilai yang lebih rendah dari nilai threshold 0,6, ditunjukan pada gambar 8. Tetapi jika hasil perhitungan nilai kemiripan kalimat lebih tinggi daripada nilai threshold, maka hal ini tidak akan mempengaruhi kesepakatan. Walaupun demikian, metode untuk menghitung nilai kemiripan pasangan kalimat dengan mencari rata-rata pasangan kata ini dapat digunakan dalam kerangka kerja yang di ajukan.

KEMIRIPAN	A	B	C	D	E	F	G	H	I	J
berbeda	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
berbeda	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
berbeda	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
berbeda	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
berbeda	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
berbeda	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
berbeda	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
berbeda	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
berbeda	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
berbeda	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Gambar 8. Hasil perhitungan nilai kemiripan yang tidak maksimal

III. Simpulan

Pengujian yang dilakukan pada 30 dokumen spesifikasi kebutuhan perangkat lunak menunjukkan bahwa kerangka kerja bekerja cukup baik pada nilai threshold = 0,6 pada ahli bahasa 4 dengan nilai kappa = 0.721868229 dengan interpretasi nilai Kappa adalah Banyak (substansial agreement), dengan nilai rata-rata 0.485304.

Dari hasil pengujian menunjukkan bahwa kerangka kerja dapat digunakan dan untuk deteksi istilah berbeda dalam dokumen spesifikasi kebutuhan perangkat lunak.

IV. Daftar Pustaka

- [1] Dao, T., & Simpson, T. (2006). Measuring Similarity Between Sentences.

- [2] Dipodjojo, A. S. (1984). Komunikasi Lisan. Yogyakarta.
- [3] Hussain, I. O. (2007). Automatic Quality Assessment of SRS Text by Means of a Decision-Tree-Based Text Classifier”, Proceedings on Seventh International Conference on Quality Software (QSIC 2007). IEEE.
- [4] Umami, Izzatul, 2012. Deteksi Forward Reference Dalam Dokumen Spesifikasi Kebutuhan Perangkat Lunak (SKPL). Fakultas Teknologi Informasi Institut Teknologi Sepuluh Nopember
- [5] K.Pohl. (1997). The Three Dimensions of Requirements Engineering. ESPRIT.
- [6] Landis Richard J, K. G. (1977). An Application of Hierarchical Kappa-type Statistics in the Assessment of Majority Agreement among Multiple Observers. Departement of Biostatistics, University of Michigan.
- [7] Laplante, P. (2007). What Every Engineer Should Know about Software Engineering. Boca Raton: CRC Press.
- [8] Meyer, B. (1985). On Formalism in Specifications. IEEE Software, vol. 3, no. 1, 6- 25.
- [9] Miller. (1995). WordNet: A Lexical Database for English. Volume 38.
- [10] Morton, Thomas. S. (1999). Using Coreference for Question Answering. CorefApp '99 Proceedings of the Workshop on Coreference and its Applications.
- [11] Riggs, F. W. (2005). A semantic/ Onomantic Puzzele on Homonyms. Retrieved 01 26, 2014, from [www2.hawaii.edu: http://www2.hawaii.edu/~fredr/homonymy](http://www2.hawaii.edu/~fredr/homonymy)
- [12] Sarab. (2005). A Semantic Analysis of Homonyms, Heteronyms & allonyms with Some Reference to Word Net. AL-Fateh Journal.
- [13] Siahaan, Daniel. (2010). Rekayasa Kebutuhan. ITS Surabaya.
- [14] Sommerville, I. (2001). Software Engineering. PEARSON.
- [15] SWEBOOK. (2004). Software Engineering Body of Knowledge. IEEE Computer Societ