

EVALUASI KOMBINASI HIPERNIM DAN SINONIM UNTUK KLASIFIKASI KEBUTUHAN NON-FUNCTIONAL BERBASIS ISO/IEC 25010

Lukman Hakim¹, Siti Rochimah², Chastine Fatichah³

^{1,2,3}Jurusan Teknik Informatika, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember, Sursabaya
Email: ¹lukman.if.its@gmail.com, ²siti@if.its.ac.id, ³chastine.fatichah@gmail.com

(Naskah masuk: 12 Desember 2018, diterima untuk diterbitkan: 25 April 2019)

Abstrak

Kebutuhan non-fungsional dianggap mampu mendukung keberhasilan pengembangan perangkat lunak. Namun, kebutuhan non-fungsional sering diabaikan selama proses pengembangan perangkat lunak. Hal ini dikarenakan kebutuhan non-fungsional sering tercampur dengan kebutuhan fungsional. Disamping itu, standar kualitas yang beragam menyebabkan kebingungan dalam menentukan aspek kualitas. Pendekatan yang ada menggunakan ISO/IEC 9126 sebagai referensi untuk mengukur aspek kualitas. ISO/IEC 9126 merupakan standar lama yang dirilis pada tahun 2001. Peneliti sebelumnya mengungkapkan ambiguitas dalam enam sub-atribut pada struktur hirarkis ISO/IEC 9126. Hal ini menimbulkan keraguan serius tentang validitas standar secara keseluruhan. Oleh karena itu, standar kualitas yang digunakan sebagai referensi untuk mengukur aspek kualitas pada penelitian ini adalah ISO/IEC 25010. Selain itu, penelitian ini juga mengusulkan suatu sistem untuk mengidentifikasi aspek kualitas kebutuhan non-fungsional dengan menggunakan 1 level hipernim dan 20 sinonim yang disebut skenario 1. Skenario ini akan dibandingkan dengan 2 level hipernim dan 9 sinonim pada masing-masing sinonim yang disebut skenario 2. Kedua skenario tersebut akan menghasilkan dua data latih berbeda. Kedua data latih tersebut akan dibandingkan menggunakan dua model pengujian yaitu berdasarkan *ground truth* pakar dan sistem dengan menggunakan metode klasifikasi KNN dan SVM. Hasil pengujian menunjukkan skenario 1 terbukti memberikan nilai lebih baik dibandingkan skenario 2 pada kedua model pengujian, dimana nilai precision dari *ground truth* pakar, KNN, dan SVM masing-masing 49.3%, 81.0%, dan 74.6%.

Kata kunci: *kebutuhan non-fungsional, hipernim, sinonim, ISO/IEC 25010*

EVALUATION OF HYPERNYMS AND SYNONYMS COMBINATION FOR CLASSIFICATION OF NON-FUNCTIONAL REQUIREMENT BASED ON ISO/IEC 25010

Abstract

Non-Functional requirements are considered capable of supporting the success of software development. However, non-functional requirements are often ignored during the software development process. This is because the quality aspects of non-functional requirements are often mixed with functional requirements. In addition, the number of diverse quality standards causes confusion in determining quality aspects. The existing approach uses ISO / IEC 9126 as a reference to measure quality aspects. ISO / IEC 9126 is an old standard released in 2001. Previous researchers revealed ambiguity in six sub-attributes on the hierarchical structure of ISO / IEC 9126. This raises serious doubts about the validity of the overall standard. Therefore, the quality standard used as a reference to measure the quality aspects of this study is ISO / IEC 25010. In addition, this study also proposes a system to identify aspects of the quality of non-functional requirements using 1 hypernym level and 20 synonyms called scenario 1. This scenario will be compared with 2 hypernym levels and 9 synonyms in each synonym called scenario 2. Both scenarios will produce two different training data. The two training data will be compared using two testing models ie based on expert ground truth and systems using the KNN and SVM classification methods. The test results showed scenario 1 is proven to provide a better value than scenario 2 in both testing models, where the precision values of expert ground truth, KNN, and SVM respectively 49.3%, 81.0%, and 74.6%.

Keywords: *non-functional requirement, hypernyms, synonym, ISO/IEC 25010*

1. PENDAHULUAN

Kebutuhan non-fungsional adalah kriteria kualitas atau kinerja yang harus dipenuhi oleh sistem perangkat lunak (Xiang et al., 2015). Kebutuhan non-fungsional tersebut dianggap mampu mendukung keberhasilan pengembangan perangkat lunak (Ullah, Iqbal and Khan, 2011; Khatter and Kalia, 2013). Namun, kebutuhan non-fungsional sering diabaikan selama proses pengembangan perangkat lunak. Ini karena aspek kualitas kebutuhan non-fungsional sering dicampur dengan persyaratan fungsional (Casamayor, Godoy and Campo, 2010; Rahimi, Mirakhorli and Cleland-Huang, 2014). Oleh karena itu diperlukan metode untuk dapat mengidentifikasi aspek kualitas kebutuhan non-fungsional, salah satunya adalah dengan mengklasifikasikan kalimat - kalimat kebutuhan yang ditulis dalam dokumen kebutuhan itu (Ramadhani, Rochimah and Yuhana, 2015).

Berbagai metode telah digunakan untuk memecahkan masalah identifikasi aspek kualitas. Diantaranya adalah algoritma Naive Bayes yang dikombinasikan dengan metode Expectation Maximization (EM) yang bertujuan untuk mengurangi data latih. Metode ini termasuk dalam pendekatan pembelajaran semi-diawasi (Casamayor, Godoy and Campo, 2010). Metode lain menggunakan pendekatan pembelajaran yang diawasi adalah mengidentifikasi aspek kualitas kebutuhan non-fungsional berbasis algoritma klasifikasi model probabilistik (Cleland-Huang et al., 2006). Kedua penelitian ini memiliki tingkat subjektivitas yang tinggi karena proses pelabelan data latih dilakukan secara manual. Disamping itu, banyaknya aspek kualitas yang ditawarkan oleh model-model kualitas seperti model *McCall*, *Boehm*, *Dromey*, *FURPS*, *BNN*, *Star*, dan *ISO* menyebabkan kebingungan dalam menentukan aspek kualitas (Behkamal, Kahani and Akbari, 2009).

Penelitian selanjutnya yaitu mengidentifikasi aspek kualitas kebutuhan non-fungsional menggunakan standar kualitas ISO/IEC 9126 dilakukan oleh Suharso (Suharso and Rochimah, 2012) dan Ramadhani (Ramadhani, Rochimah and Yuhana, 2015). Kedua penelitian ini masih menggunakan ISO/IEC 9126 sebagai referensi untuk mengukur aspek kualitas. ISO/IEC 9126 merupakan standar lama yang dirilis pada tahun 2001. Evaluasi kritis dari penelitian yang dilakukan oleh Al-Kilidar (Al-Kilidar, Cox and Kitchenham, 2005) dan Al-Qutaish (Al-Qutaish, 2010) mengungkapkan ambiguitas dalam enam sub-atribut dalam struktur hirarkis ISO / IEC 9126. Sejalan dengan evaluasi ini, Behkamal (Behkamal, Kahani and Akbari, 2009) hanya menggunakan 21 sub-atribut dari total 27 sub-atribut yang dimiliki oleh ISO/IEC 9126. Oleh karena itu, standar kualitas yang digunakan dalam penelitian ini adalah ISO/IEC 25010.

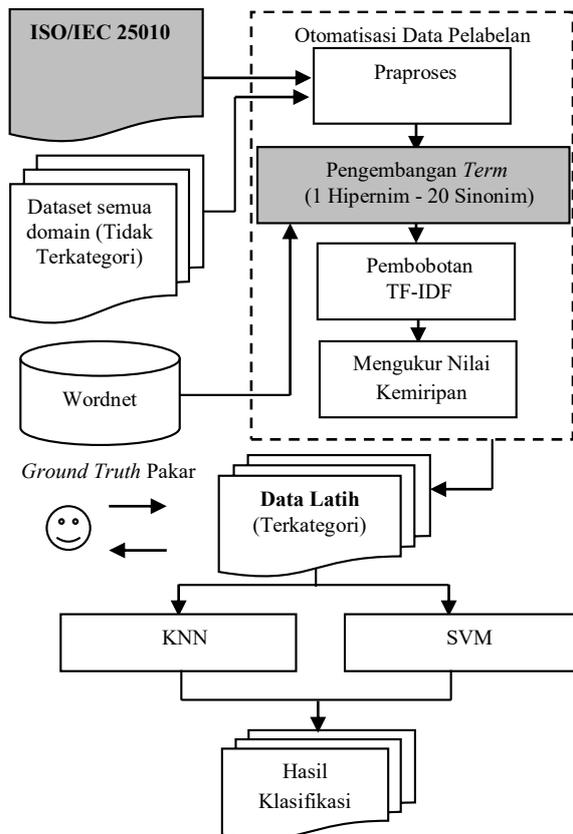
Penelitian ini juga membuat pengamatan tentang efek penggunaan kombinasi hipernim dan sinonim dalam pembobotan TF-IDF. Ramadhani (Ramadhani, Rochimah and Yuhana, 2015) melakukan perbandingan pengembangan term dengan sinonim dan term gabungan antara 2 level hipernim dan 9 sinonim. Hasilnya menggunakan term gabungan hipernim-sinonim terbukti meningkatkan kinerja klasifikasi. Akan tetapi, terlalu banyak penggunaan hipernim dapat menghilangkan makna asli dari *term* yang dikembangkan. Menurut Rozi (Rozi, Faticah and Purwitasari, 2015) penggunaan hipernim mampu meningkatkan akurasi. Namun, penggunaan hipernim dalam beberapa dataset hanya akan menambah jumlah *term* yang akan dianggap sebagai noise sehingga mengurangi akurasi. Oleh karena itu, dua skenario diusulkan untuk membuktikan hipotesa tersebut.

Penelitian ini mengusulkan suatu sistem untuk mengidentifikasi aspek kualitas kebutuhan non-fungsional dengan menggunakan 1 level hipernim dan 20 sinonim yang disebut skenario 1. Skenario ini akan dibandingkan dengan 2 level hipernim dan 9 sinonim pada masing-masing hipernim seperti yang dilakukan penelitian sebelumnya yang disebut skenario 2 (Ramadhani, Rochimah and Yuhana, 2015). Metode pembobotan yang diusulkan akan mengintegrasikan kekuatan frekuensi. Jumlah frekuensi diharapkan dapat meningkatkan akurasi.

Kedua skenario tersebut akan evaluasi pada dua tahap pengujian. Tahap pertama yaitu data latih yang dihasilkan akan dibandingkan dengan *ground truth* dari pakar. *Ground truth* adalah hasil kesepakatan antara ketiga pakar terhadap pelabelan manual yang dilakukan terhadap kalimat kebutuhan tertentu masuk ke dalam aspek kualitas tertentu. *Gwet's AC1* digunakan untuk mengukur kesepakatan yang dihasilkan. Tujuan evaluasi ini adalah untuk mengukur seberapa baik data latih yang dihasilkan sebelum dilanjutkan pada proses klasifikasi. Tahap kedua yaitu data latih yang sudah diketahui reliabilitasnya berdasarkan *ground truth* akan diuji menggunakan dua metode klasifikasi. Metode klasifikasi yang digunakan yaitu KNN dan SVM. Kedua metode klasifikasi tersebut bertujuan untuk mengukur performa dari data latih yang dihasilkan dari kedua skenario.

2. METODE PENELITIAN

Desain sistem ditunjukkan pada Gambar 1. Pada Gambar 1 terdiri dari dua proses utama yaitu otomatisasi data pelabelan, dan klasifikasi. Kontribusi utama pada penelitian ini terdapat pada proses otomatisasi data pelabelan yaitu ISO/IEC 25010, dan pengembangan term menggunakan 1 level hipernim dan 20 sinonim. Sedangkan pada fase klasifikasi pada penelitian ini menggunakan metode KNN dan SVM.



Gambar 1. Bagan Metodologi

Tabel 1. Sub-atribut/ Model Kualitas ISO/IEC 25010.

ID	Model Kualitas	ID	Model Kualitas
1	Functional Completeness	17	Fault Tolerance
2	Functional Correctness	18	Recoverability
3	Functional Appropriateness	19	Confidentiality
4	Time Behaviour	20	Integrity
5	Resource Utilization	21	Non-repudiation
6	Capacity	22	Authenticity
7	Co-existence	23	Accountability
8	Interoperability	24	Modularity
9	Appropriateness Recognizability	25	Reusability
10	Learnability	26	Analysability
11	Operability	27	Modifiability
12	User Error Protection	28	Testability
13	User Interface Aesthetics	29	Adaptability
14	Accessibility	30	Installability
15	Maturity	31	Replaceability
16	Availability		

2.1. ISO/IEC 25010

Mengukur kualitas sistem bisa dilakukan dengan cara seberapa baik suatu sistem dapat memenuhi berbagai kebutuhan pemangku kepentingan. Oleh karena itu, perlu untuk mempertimbangkan kualitas dari perspektif pemangku kepentingan yang berbeda. Model-model kualitas menyediakan kerangka untuk mengumpulkan kebutuhan pemangku kepentingan. Dokumen dalam standar ISO / IEC 25010 mendefinisikan model kualitas produk yang terdiri dari delapan atribut yang selanjutnya dibagi lagi menjadi sub-atribut. Model ini dipahami sebagai

model struktural yang didefinisikan oleh SquaRE sebagai "Model Kualitas: set karakteristik yang ditetapkan, dan hubungan di antara mereka. Standar ini juga memberikan definisi dari masing-masing atribut kualitas dan sub-atribut (Febrero, Calero and Moraga, 2016). Detail model kualitas dapat dilihat pada Tabel 1.

2.2. Data

Data yang digunakan pada penelitian ini terdiri dari 21 spesifikasi kebutuhan perangkat lunak komunikasi VoIP sejumlah 1300 kalimat kebutuhan dan 2 spesifikasi kebutuhan perangkat lunak geolocation sejumlah 66 kalimat kebutuhan. Jadi total kalimat kebutuhan yang digunakan pada penelitian ini sebanyak 1366. Kedua data tersebut terdiri dari 1141 kebutuhan fungsional dan 225 kebutuhan non-fungsional (Cleland-Huang, 2006).

2.3. Praproses

Kalimat kebutuhan pada keseluruhan dataset akan melalui tahapan praproses. Pada tahap ini yang pertama dilakukan adalah tokenisasi dan case folding. Tokenisasi berfungsi mengubah kalimat menjadi potongan kata. Sedangkan case folding merubah huruf menjadi huruf kecil. Contoh dari tokenisasi dan case folding misalnya salah satu kalimat kebutuhan yaitu "The system must be maintainable" akan ditokenisasi menjadi "the", "system", "must", "be", dan "maintainable".

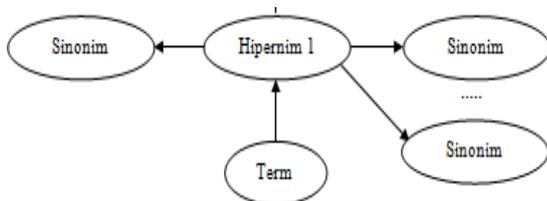
Langkah selanjutnya dalam tahap ini adalah stopword. Stopword berfungsi menghilangkan kata, tanda baca, simbol dan bilangan yang tidak berpotensi menjadi indikator topik pada kalimat kebutuhan. Contoh dari stopwords removal yaitu "a", "an", "and", "as", "be", tanda baca koma dan sebagainya.

Setelah stopword selesai kemudian dilakukan stemming untuk menjadikan teks menjadi kata dasar. Misal dari contoh kalimat kebutuhan pada tahap sebelumnya maka kata "maintainable" akan diubah menjadi "maintan". Tahapan praproses ini akan diterapkan untuk kalimat kebutuhan pada keseluruhan dataset maupun standard acuan model kualitas ISO/IEC 25010.

2.4. Kombinasi Hipernim dan Sinonim

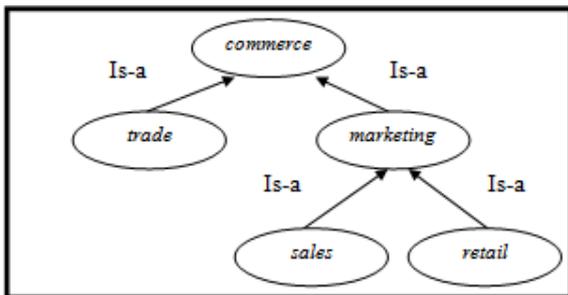
Pengembangan term menggunakan hipernim dan sinonim hanya dilakukan pada term yang berkaitan langsung dengan ISO/IEC 25010. Hal ini dikarenakan term-term tersebut dijadikan acuan dasar untuk melakukan proses klasifikasi kalimat kebutuhan. Penggunaan hipernim dari suatu term harus dibatasi. Hal ini dikarenakan setiap kenaikan satu level hipernim menandakan term tersebut diperluas ke dalam makna yang lebih umum. Penggunaan hipernim terlalu banyak dikhawatirkan akan menghilangkan makna asli dari term yang

dikembangkan. Pola pengembangan kombinasi hipernim dan sinonim ditunjukkan pada Gambar 2.



Gambar 2. Pola Pengembangan kombinasi Hipernim dan Sinonim

Secara umum hipernim merupakan suatu kata yang memiliki ruang lingkup makna yang mencakup hal - hal umum atau mencakup aspek yang luas. Relasi Hipernim dalam Wordnet menurut Miller (Miller, 1995) memiliki kata lain *super-name*, yaitu hubungan transitif antar *synsets*, hubungan semantik ini mengatur makna kata dalam struktur hirarki. Struktur hirarki yang dimaksud adalah struktur hirarki pohon *IS-A*, dimana kata Y merupakan hipernim dari suatu kata X, jika setiap X adalah (*IS-A*) Y (Koirala & Rasheed, 2008). Berikut ilustrasi dari relasi hipernim berdasarkan leksikal Wordnet pada Gambar 3.



Gambar 3. Representasi Relasi Hipernim dalam Wordnet (Chen, Tseng and Liang, 2010)

Pada Gambar 3 diatas dapat diketahui bahwa berdasarkan hubungan *IS-A* pada struktur hirarki pohon diatas, kata *marketing* merupakan hipernim untuk kata *sales* dan *retail* pada level 1, sedangkan kata *commerce* yang berada pada level 2 merupakan hipernim untuk kata *sales*, *retail*, *marketing*, dan *trade*.

2.5. Wordnet

Wordnet adalah sistem referensi leksikal bahasa inggris yang dikembangkan oleh para ilmuwan di Universitas Princeton pada tahun 1990 (Christiane, 1998; Miller, 1995). Wordnet dikembangkan untuk menyediakan kamus online yang dibangun tidak hanya dalam urutan abjad tetapi dalam cara yang lebih konseptual yang menunjukkan hubungan semantik.

Wordnet telah meningkatkan jumlah kata dan menambahkan hubungan kata dalam 17 jenis berbeda diantaranya sinonim, antonim, hyponyms, dan mironim. Sekarang ini, wordnet menampung lebih

dari 150.000 *term*. *Term* tersebut diorganisasikan didalam 115.000 arti kata atau set sinonim yang disebut *synsets*, dan setiap *synsets* dapat terhubung dengan *synsets* yang lainnya. Menggunakan wordnet dapat untuk mengukur keterkaitan *term* dari kata - kata, sehingga dapat dihitung nilai similaritas dan keterkaitan semantik antar *term* yang ditemukan dalam wordnet. Wordnet memiliki 41 kategori leksikal sebagai *nouns*, *verbs*, *adjectives*, dan *adverb*. Wordnet juga dapat digunakan untuk mendapatkan hubungan sinonim, hipernim, hiponim, meronim dan informasi topik yang didapatkan dari suatu kata (Chua, 2014; Çelik, 2013; Patil, 2013).

Dalam penelitian ini, wordnet digunakan sebagai referensi leksikal untuk melakukan pengembangan *term* atau memperkaya *term* dari data latihan dengan menggunakan relasi hipernim yang didapatkan dari wordnet tersebut.

2.6. Pembobotan TF-IDF

Term frekuensi (tf) adalah menghitung jumlah kemunculan (*f*) *term* t_i dalam setiap kalimat kebutuhan d_j . Untuk proses perhitungan *tf* bisa dilihat pada persamaan (1).

$$TF = (t_i, D) \tag{1}$$

Inverse document frequency (idf) adalah perhitungan bobot *term* dari suatu kalimat kebutuhan terhadap kalimat kebutuhan lain. *Term* yang jarang sekali muncul dalam suatu dokumen merupakan *term* yang sangat bernilai. Log dari jumlah kalimat kebutuhan secara keseluruhan D dibagi dengan jumlah kalimat kebutuhan d yang mengandung *term* t_i menghasilkan sebuah bobot *term* terhadap kalimat kebutuhan yang ada. Untuk proses perhitungan *idf* bisa dilihat pada persamaan 2.

$$IDF = \log \frac{N}{df(t_i, D)} + 1 \tag{2}$$

Proses selanjutnya adalah mengalikan *tf* dan *idf*. Proses ini perlu dilakukan normalisasi agar nilai bobot yang didapatkan berada dalam rentang nilai 0 hingga 1. Untuk proses perhitungan $tf * idf$ bisa dilihat pada persamaan 3.

$$\omega_K(t_i) = \frac{tf(t_i, K) \times \log \frac{N}{df(t_i, K)} + 1}{\sqrt{\sum_{t_i} (tf(t_i, K) \times \log \frac{N}{df(t_i, K)} + 1)^2}} \tag{3}$$

2.7. Cosine Similarity

Cosine Similarity digunakan untuk mengukur kemiripan antara kalimat kebutuhan dengan definisi aspek kualitas menggunakan ISO/IEC 25010. Hasil nilai kosinus berkisar antara 0 hingga 1, dimana nilai 0 menandakan bahwa query dan dokumen tidak mirip sama sekali, dan 1 menandakan bahwa query dan dokumen benar-benar identik. *Cosine Similarity* bisa dilihat pada persamaan 4.

$$Cos(Q, K) = \sum_{r=1}^M \omega_Q(t_i) \times \omega_K(t_i) \quad (4)$$

Dimana $\omega_Q(t_i)$ merupakan bobot dari term t_i yang berasal atribut aspek kualitas, sedangkan $\omega_K(t_i)$ merupakan bobot yang didapatkan dari term t_i yang berasal suatu kalimat kebutuhan K . Dengan menggunakan penentuan nilai *threshold* tertentu maka akan dapat dihasilkan kalimat - kalimat kebutuhan K yang terlabeli ke dalam satu atau lebih atribut aspek kualitas. Proses selanjukan akan dilakukan klasifikasi menggunakan metode KNN dan SVM untuk mengukur seberapa baik pengaruh penggunaan 1 level hipernim dan 20 sinonim yang diusulkan.

2.8. K-Nearest Neighbor(KNN)

K-NN adalah metode klasifikasi teks dengan pendekatan pembelajaran diawasi berdasarkan jarak terdekat (Putri, Herlambang and Wihandika, 2017). Untuk tujuan klasifikasi, metode mengevaluasi kelas-kelas dari data-data yang dipilih(Gunawan, 2018).

2.9. Support Vector Machine (SVM)

(SVM dianggap sebagai salah satu metode klasifikasi dengan kinerja yang baik. SVM memiliki tiga keunggulan dibandingkan metode klasifikasi lainnya. Pertama, SVM berhasil diterapkan pada klasifikasi teks dengan pendekatan pembelajaran diawasi (Chau and Chen, 2008). Kedua, Proses klasifikasi menggunakan SVM yang mengungguli pendekatan yang ada dalam klasifikasi (Rashwan, Ormandjieva and Witte, 2013). Ketiga, SVM adalah klasifikasi biner(Hasanah, 2016), Strategi ini dapat meningkatkan ketepatan proses pengelompokkan SVM untuk mengatasi ketidakseimbangan data. Karena jika menggunakan klasifikasi multi-label maka ketidakseimbangan data latih tidak dapat dihindari (Sun, Lim and Liu, 2009).

2.10. Evaluasi

Ada banyak jenis metode evaluasi yang dapat digunakan untuk mengukur keandalan metode yang diusulkan. Penelitian ini mengusulkan metode pengukuran menggunakan akurasi, presisi, dan recall. Proses perhitungan *accuracy*, *precision*, dan *recall* ditentukan dari prediksi informasi pada nilai aktual yang diwakili oleh True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Persamaan ketiga dari metode evaluasi ditunjukkan dalam persamaan 5 untuk *accuracy*, persamaan 6 untuk *precision*, dan persamaan 7 untuk *recall*.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

2.11. Ground Truth

Ground truth adalah hasil kesepakatan antara satu pakar dengan pakar yang lain. Pada penelitian ini, pakar dilibatkan untuk melakukan pelabelan secara manual antara kalimat kebutuhan dengan aspek kualitas. Misalnya pakar pertama melabeli kalimat kebutuhan ke 1 ke dalam aspek kualitas *capacity*, dan *operability*. Sedangkan pakar kedua melabeli kalimat kebutuhan ke 1 ke dalam aspek kualitas *capacity* dan *learnability*. Hasilnya ditunjukkan pada Tabel 2, Tabel 3 dan Tabel 4.

Tabel 2. Hasil kesepakatan kedua pakar terhadap kalimat kebutuhan ke 1 pada aspek kualitas *capacity*.

Pakar 1	Pakar 2	
	Ya	Tidak
Ya	2	0
Tidak	0	0

Tabel 3. Hasil kesepakatan kedua pakar terhadap kalimat kebutuhan ke 1 pada aspek kualitas *operability*.

Pakar 1	Pakar 2	
	Ya	Tidak
Ya	1	1
Tidak	0	0

Tabel 4. Hasil kesepakatan kedua pakar terhadap kalimat kebutuhan ke 1 pada aspek kualitas *learnability*.

Pakar 1	Pakar 2	
	Ya	Tidak
Ya	1	0
Tidak	1	0

Berdasarkan Tabel 2, Tabel 3, dan Tabel 4 diketahui hasil pelabelan antara kalimat kebutuhan dan aspek kualitas. Selanjutnya, untuk mencari *Ground truth* dengan menggunakan data pada ketiga tabel tersebut diukur menggunakan Gwet’s AC1

2.12. Gwet’s AC1

Gwet’s AC1 adalah sebuah metode yang berfungsi untuk mengukur indeks kesepakatan antara dua pakar(Gwet, 2013). Penghitungan nilai AC1, hasil dari observasi pengamat dimasukkan dalam matrik 2x2 seperti yang ditunjukkan pada Tabel 5. Penulisan Hasil Pengamatan di bawah ini:

Tabel 5. Tabel Penulisan Hasil Pengamatan oleh kedua pakar.

Pakar 1	Pakar 2		Total
	Ya	Tidak	
Ya	w	x	B1=w+x
Tidak	y	z	B2=y+z
Total	A1=w+y	A2=x+z	N

Pada Tabel 5 bisa dilihat ada dua pakar yang terlibat. Masing-masing pakar bisa salah satu dari dua kategori. Dua kategori tersebut dilabelkan sebagai “ya” dan “tidak”. Variabel w diklasifikasikan oleh dua pakar sebagai ya, Variabel x diklasifikasikan oleh pakar 2 sebagai tidak, Variabel z diklasifikasikan oleh dua pakar sebagai tidak, dan Variabel y diklasifikasikan oleh pakar 1 sebagai tidak. Variabel

B1 menunjukkan jumlah dari variabel w dan variabel x, Variabel B2 menunjukkan jumlah dari variabel y dan variabel z, Variabel A1 menunjukkan jumlah dari variabel w dan variabel y, dan variabel A2 menunjukkan jumlah dari variabel x dan variabel z.

Perhitungan AC1-statistic :

$$P_o = \frac{w+z}{N}$$

(2.28)

Dimana variabel w adalah banyaknya data yang dikelompokkan ke dalam kategori “Ya” oleh kedua pakar sedangkan variabel z adalah banyaknya data yang dikelompokkan ke dalam kategori “tidak” oleh kedua pakar.

Sedangkan untuk menghitung probabilitas acak antara kedua pakar yaitu :

- Pakar 1 mengatakan Ya sebanyak 1 dan Tidak sebanyak 0.
- Pakar 2 mengatakan Ya sebanyak 1 dan Tidak sebanyak 0.

Diketahui bahwa kedua pengamat mengatakan Ya sebanyak 100 %. Jadi, probabilitas yang diharapkan dari kedua pengamat mengatakan Ya secara acak ditunjukkan pada persamaan 8, 9 dan untuk probabilitas keseluruhan ditunjukkan pada persamaan 10. Sedangkan untuk menghitung nilai indeks Kappa menggunakan Cohen Kappa statistik dengan persamaan 11.

$$P_{ya} = \frac{w+x}{w+x+y+z} * \frac{w+y}{w+x+y+z} \tag{8}$$

$$P_{tidak} = \frac{y+z}{w+x+y+z} * \frac{x+z}{w+x+y+z} \tag{9}$$

$$P_e = P_{ya} + P_{tidak} \tag{10}$$

$$k = \frac{P_o + P_e}{1 - P_e} \tag{11}$$

Probabilitas *chance-agreement*,
 $e(y) = 2P_1(1 - P_1) \tag{5}$

$$AC1\text{-statistic, } AC1 = \frac{P - e(y)}{1 - e(y)} \tag{6}$$

(2.34)

Berikut adalah contoh perhitungan *AC1-static*:

Terdapat dua orang pakar yang dimintai pendapatnya terhadap satu kalimat kebutuhan berdasarkan satu aspek kualitas ISO/IEC 25010:2011, apakah kalimat kebutuhan tersebut masuk ke dalam aspek kualitas *Functional Completeness* atau tidak. Hasil pengamatan terlihat seperti Tabel 7 di bawah ini:

Tabel 7. Data Hasil Pengamatan kedua pakar terhadap aspek kualitas functional completeness

Pakar 1	Pakar 2		Total
	Ya	Tidak	
Ya	2	0	2
Tidak	0	0	0
Total	2	0	2

$$P_o = \frac{a+d}{N} = \frac{2+0}{2} = 1$$

Jika dihitung menggunakan indeks Cohen Kappa statistik maka didapat nilai indeks Kappa sesuai persamaan 8, 9, 10, dan 11 sebagai berikut :

$$P_{ya} = \left(\frac{2}{2}\right) \times \left(\frac{2}{2}\right) = 1$$

$$P_{tidak} = \left(\frac{0}{2}\right) \times \left(\frac{0}{2}\right) = 0$$

$$P_e = 1 + 0 = 1$$

$$Kappa = (1 - 1) / (1 - 1) = 0$$

Nilai indeks Cohen Kappa yang dihasilkan pada data pengamatan Tabel 7 sangatlah rendah yaitu 0, hal ini dikarenakan pengamatan kedua pakar hanya berdasarkan satu aspek kualitas saja. Hasilnya akan tetap bernilai 0 meskipun kedua pengamat sudah menyatakan persetujuannya dengan memilih kategori Ya. Oleh karena itu, penelitian ini menggunakan Gwet's AC1. Gwet's AC1 terbukti mampu menghasilkan nilai yang benar sesuai dengan kesepakatan kedua pengamat. Berikut perhitungan dari Gwet's AC1 sesuai persamaan 5, dan 6 :

$$e(y) = 2\left(\frac{2+2}{2 \times 2}\right) \left(1 - \frac{2+2}{2 \times 2}\right) = 2 * 0 = 0$$

$$AC1 = (1 - 0) / (1 - 0) = 1$$

Berdasarkan nilai AC1, maka dapat disimpulkan bahwa kalimat kebutuhan masuk ke dalam aspek kualitas *Functional Completeness*.

Perhitungan AC1 ini diujikan terhadap hasil pelabelan manual yang dilakukan oleh tiga orang pakar dibidang requirement. Masing-masing pakar melabeli 225 kalimat kebutuhan non-fungsional ke dalam lebih dari satu aspek kualitas (*multi-label*). Contoh hasil pelabelan beberapa kalimat kebutuhan non-fungsional ditunjukkan pada Tabel 8.

Tabel 8. Contoh Hasil Pelabelan

KK	Aspek Kualitas		
	Capacity	Operability	Learnability
1	√	√	√
2	√		√
3		√	
4	√		
5			√
...			
225			√

3. HASIL DAN PEMBAHASAN

Pengujian pada penelitian ini dibagi menjadi dua tahap yaitu skenario uji dan skenario hasil.

3.1. Skenario Uji

Di bagian ini, jelaskan hasil tes dari setiap skenario yang telah dilakukan. Dataset terdiri dari 1366 kalimat yang akan dilabeli menjadi 31 aspek kualitas berdasarkan ISO 25010. Setiap data latih yang dihasilkan oleh sistem pada fase otomatisasi data pelabelan akan dicocokkan dengan *ground truth*. *Ground truth* ini berisi daftar kebutuhan non-fungsional untuk setiap model kualitas yang telah diklasifikasikan secara manual oleh pakar sebagai jawaban yang benar. Klasifikasi manual ini didasarkan pada kesesuaian makna dan *term* potensial dari kalimat dalam data uji. Sehingga total jumlah kalimat yang diambil, dan jumlah kalimat yang diambil dan relevan berdasarkan *ground truth* yang relevan akan membentuk kinerja rata-rata *accuracy*, *precision*, dan *recall*. Evaluasi ini bertujuan untuk mengukur seberapa baik data latih yang dihasilkan oleh sistem didasarkan pada pelabelan yang dilakukan oleh para ahli.

Percobaan berikutnya akan diuji data latih yang dihasilkan dari kedua skenario menggunakan metode klasifikasi KNN dan SVM. Tes ini bertujuan untuk mengetahui seberapa baik metode KNN dan SVM dapat mengklasifikasikan kalimat kebutuhan berdasarkan nilai *accuracy*, *precision*, dan *recall* dari kedua data latih yang dihasilkan pada proses sebelumnya.

Setiap metode dalam skenario ujicoba akan diuji menggunakan jumlah parameter tetangga terdekat (k) = 1 untuk metode KNN dan menggunakan *threshold* 0,5 hingga 0,7 untuk kedua metode klasifikasi yaitu KNN dan SVM. Percobaan menggunakan metode KNN dengan menggunakan data latih dari skenario 1, dan skenario 2 dapat dilihat pada Tabel 10. Sedangkan untuk metode SVM dapat dilihat pada Table 11.

3.2. Skenario Hasil

Pada Tabel 9, *accuracy*, *precision*, dan *recall*, pada Skenario 1 lebih besar dari Skenario 2. Skenario 1 menghasilkan *accuracy* tertinggi 97.8% pada *threshold* 0.6, *precision* tertinggi 49.3% pada *threshold* 0.7, dan *recall* tertinggi 86.4% pada *threshold* 0.1. Skenario 1 terbukti menghasilkan data latih yang lebih baik menurut penilaian pakar.

Pada Tabel 10, *accuracy*, *precision*, dan *recall*, pada Skenario 1 lebih besar dari Skenario 2. Skenario 1 menghasilkan *accuracy* tertinggi 82.2% pada *threshold* 0.7, *precision* tertinggi 81.0% pada *threshold* 0.7, dan *recall* tertinggi 82.6% pada *threshold* 0.7. Pada Tabel 11, *accuracy*, *precision*, dan *recall*, pada Skenario 1 lebih besar dari Skenario 2. Skenario 1 menghasilkan *accuracy* tertinggi 85.7% pada *threshold* 0.7, *precision* tertinggi 74.6% pada *threshold* 0.7, dan *recall* tertinggi 85.7% pada *threshold* 0.7. Sedangkan pada Table 12, *accuracy*, *precision*, dan *recall*, metode SVM lebih besar dari pada metode KNN.

Tabel 9. Tabulasi rata-rata kinerja *accuracy*, *precision*, dan *recall* antara skenario 1 dan skenario 2

Threshold	Skenario 1 1 Level Hipernim @20 Sinonim			Skenario 2 2 Level Hipernim @9 Sinonim		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
0.1	71.8%	4.2%	86.4%	75.5%	5.5%	75.8%
0.2	71.8%	4.2%	86.4%	75.7%	5.5%	75.8%
0.3	86.2%	10.1%	65.9%	89.1%	10.6%	54.1%
0.4	93.0%	15.4%	53.4%	95.6%	15.0%	37.0%
0.5	94.0%	35.1%	34.0%	97.4%	38.1%	18.7%
0.6	97.8%	37.1%	22.5%	97.7%	34.5%	11.9%
0.7	97.8%	49.3%	12.2%	97.6%	23.5%	3.9%

Tabel 10. Hasil metode KNN menggunakan data latih dari kedua skenario

Threshold	Skenario 1 1 Level Hipernim @20 Sinonim			Skenario 2 2 Level Hipernim @9 Sinonim		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
0.5	71.1	68.3	71.1	72.2	70.2	72.2
0.6	77.3	74.5	77.3	78.0	76.3	78.0
0.7	82.2	81.0	82.6	81.2	80.2	81.0

Tabel 11. Hasil metode SVM menggunakan data latih dari kedua skenario

Threshold	Skenario 1 1 Level Hipernim @20 Sinonim			Skenario 2 2 Level Hipernim @9 Sinonim		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
0.5	75.8	63.9	75.8	76.0	61.1	76.0
0.6	83.3	71.9	83.3	81.1	69.1	81.1
0.7	85.7	74.6	85.7	84.1	72.2	84.1

Tabel 12. Hasil perbandingan KNN dan SVM menggunakan data latih dari skenario 1

Threshold	KNN			SVM		
	1 Level Hipernim @20 Sinonim			1 Level Hipernim @20 Sinonim		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
0.5	71.1	68.3	71.1	75.8	63.9	75.8
0.6	77.3	74.5	77.3	83.3	71.9	83.3
0.7	82.2	81.0	82.6	85.7	74.6	85.7

4. KESIMPULAN

Berdasarkan hasil pengujian dalam penelitian ini, dapat disimpulkan bahwa penggunaan 1 level hipernim dan 20 sinonim pada fase otomatisasi data pelabelan terbukti meningkatkan akurasi baik berdasarkan *ground truth* pakar maupun menggunakan metode klasifikasi KNN dan SVM.

Pekerjaan pada penelitian saat ini sedang berlangsung dan akan dilakukan tiga pengembangan. Pertama adalah kita akan menggunakan metode A-SUWO untuk mengatasi ketidakseimbangan data sebelum proses klasifikasi. Kedua, mempertimbangkan hubungan semantik antara term relevan pada setiap aspek kualitas dengan term-term pada seluruh dataset. Ketiga, menggunakan metode FSKNN untuk melakukan klasifikasi kalimat kebutuhan. Metode FSKNN merupakan pengembangan dari metode KNN. Metode FSKNN belum bisa mengatasi permasalahan data tidak seimbang. Oleh karena itu, pada pengembangan kedua, sebelum klasifikasi dibantu oleh metode A-SUWO.

DAFTAR PUSTAKA

- AL-KILIDAR, H., COX, K. and KITCHENHAM, B., 2005. The use and usefulness of the ISO/IEC 9126 quality standard. *2005 International Symposium on Empirical Software Engineering, ISESE 2005*, pp.126–132.
- AL-QUTAISH, R.E., 2010. Quality models in software engineering literature: an analytical and comparative study. *Journal of American Science*, 6(3), pp.166–175.
- BEHKAMAL, B., KAHANI, M. and AKBARI, M.K., 2009. Customizing ISO 9126 quality model for evaluation of B2B applications. *Information and Software Technology*, [online] 51(3), pp.599–609..
- CASAMAYOR, A., GODOY, D. and CAMPO, M., 2010. Identification of non-functional requirements in textual specifications: A semi-supervised learning approach. *Information and Software Technology*, [online] 52(4), pp.436–445.
- ÇELIK, K., 2013. A Comprehensive Analysis of using Semantic Information in Text Categorization.
- CHAU, M. and CHEN, H., 2008. A machine learning approach to web page filtering using content and structure analysis. *Decision Support Systems*, 44(2), pp.482–494.
- CHEN, C., TSENG, F.S.C. and LIANG, T., 2010. Data & Knowledge Engineering An integration of WordNet and fuzzy association rule mining for multi-label document clustering. *DATAK*, [online] 69(11), pp.1208–1226..
- CHRISTIANE, F., 1998. Applications of Wordnet. In: *WordNet: An Electronic Lexical Database*. [online] MITP.
- CHUA, S., 2014. Semantic Feature Selection Using WordNet Narayanan Kulathuramaiyer. pp.0–6.
- CLELAND-HUANG, J., SETTIMI, R., ZOU, X. and SOLE, P., 2006. The detection and classification of non-functional requirements with application to early aspects. *Proceedings of the IEEE International Conference on Requirements Engineering*, pp.36–45.
- FEBRERO, F., CALERO, C. and MORAGA, M.Á., 2016. Software reliability modeling based on ISO/IEC SQuARE. *Information and Software Technology*, 70, pp.18–29.
- GUNAWAN, F.E., ASHANTI, L., SEKISHITA, N., KEBON, J. and NO, J.R., 2018. A Simple Classifier for Detecting Online Child Grooming Conversation. *Telkonnika*, 16(3), pp.1693–6930.
- GWET, K., 2013. Self-care behaviors of Italian patients with heart failure. *European Journal of Cardiovascular Nursing*, [online] 12(1), p.S11.
- HASANAH, U., RESITA, L., PRATAMA, A. and CHOLISSODIN, I., 2016. Perbandingan Metode Svm, Fuzzy-Knn, Dan Bdt-Svm Untuk Klasifikasi Detak Jantung Hasil Elektrokardiografi. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 3(3), pp.201–207.
- KHATTER, K. and KALIA, A., 2013. Impact of non-functional requirements on requirements evolution. *International Conference on Emerging Trends in Engineering and Technology, ICETET*, pp.61–68.
- MILLER, G.A., 1995. WordNet : A Lexical Database for English. 38(11), pp.39–41.
- PATIL, L.H., 2013. A Novel Feature Selection based on information gain using WordNet. pp.625–629.
- PUTRI, R.R.M., HERLAMBANG, R.Y. and WIHANDIKA, R.C., 2017. Implementasi Metode K-Nearest Neighbour Dengan Pembobotan TF.IDF.ICF Untuk Kategorisasi Ide Kreatif Pada Perusahaan. *Jurnal Teknologi*

- Informasi dan Ilmu Komputer*, [online] 4(2), p.97.
- RAHIMI, M., MIRAKHORLI, M. and CLELAND-HUANG, J., 2014. Automated extraction and visualization of quality concerns from requirements specifications. *2014 IEEE 22nd International Requirements Engineering Conference, RE 2014 - Proceedings*, pp.253–262.
- RAMADHANI, D.A., ROCHIMAH, S. and YUHANA, U.L., 2015. Classification of Non-Functional Requirements Using Semantic-FSKNN Based ISO/IEC 9126. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, [online] 13(4), p.1456.
- RASHWAN, A., ORMANDJIEVA, O. and WITTE, R., 2013. Ontology-based classification of non-functional requirements in software specifications: A new corpus and SVM-based classifier. *Proceedings - International Computer Software and Applications Conference*, (ii), pp.381–386.
- ROZI, F., FATICHAH, C. and PURWITASARI, D., 2015. Ekstraksi Kata Kunci Berdasarkan Hipernim Menggunakan Fuzzy Association Rule Mining untuk Pengelompokan Dokumen. *Jurnal Ilmiah Teknologi Informasi (JUTI)*, 13(2), pp.190–197.
- SUHARSO, W. and ROCHIMAH, S., 2012. *Sistem Penambang*.
- SUN, A., LIM, E.P. and LIU, Y., 2009. On strategies for imbalanced text classification using SVM: A comparative study. *Decision Support Systems*, [online] 48(1), pp.191–201..
- ULLAH, S., IQBAL, M. and KHAN, A.M., 2011. A survey on issues in non-functional requirements elicitation. *Proceedings - International Conference on Computer Networks and Information Technology*, pp.333–340.
- XIANG, H., MA, Q., FENG, Y., Tan, Y., Hu, H., FU, C. AND ZHANG, T., 2015. Semantic modelling and automated reasoning of non-functional requirement conflicts in the context of softgoal interdependencies. *IET Software*, [online] 9(6), pp.145–156.

Halaman ini sengaja dikosongkan