

Information System, Technology & Communication

## KATEGORISASI DAN ANOTASI BAHAN PEMBELAJARAN MENGGUNAKAN PENDEKATAN KESAMAAN TEKS

Subhan A. Gani\*

Jurusan Teknik Industri, Fakultas Teknik, Universitas Malikussaleh, Lhokseumawe, Indonesia

\*Corresponding Author: subhan.a.gani@gmail.com

**Abstrak** – Penggunaan kembali bahan pembelajaran dalam sistem e-learning menjadi isu penting karena biaya pembuatan bahan e-learning itu mahal. Memperluas cakupan penggunaan bahan pembelajaran untuk bidang tertentu dapat dicapai dengan membubuhi keterangan (anotasi) di dalam metadata Bahan Pembelajaran (Learning Object, yang selanjutnya disingkat: LO) secara semantik sesuai dengan deskripsi bidang yang diminati oleh pengguna. Penelitian ini bermaksud untuk memperluas penggunaan kembali LO dari sebuah repositori e-learning berbahasa Inggris agar dapat digunakan sebagai materi tambahan pembelajaran mandiri melalui sistem e-learning dalam program pelatihan tenaga kerja. Kemiripan topik sebuah LO dihitung secara kuantitatif menggunakan pendekatan Kesamaan Tingkat Lexikal dengan memanfaatkan database WordNet sebagai basis kesamaan leksikal. Pendekatan ini dimaksudkan untuk mengkategorikan topik LO dengan topik deskripsi pekerjaan seorang peserta pelatihan dengan cara mengukur skor kesamaan antara makna tekstual dari judul, deskripsi, dan kata kunci LO dengan judul, deskripsi, dan kata kunci dalam deskripsi pekerjaan tertentu. Tujuan penelitian ini adalah untuk mengevaluasi tingkat akurasi pendekatan Kesamaan Tingkat Lexikal dalam menyimpulkan kesamaan topik dari LO. Berdasarkan percobaan, pendekatan ini memberikan tingkat akurasi yang rendah ketika menemukan dua topik yang serupa, namun mempunyai tingkat akurasi yang tinggi dalam menemukan dua topik yang berbeda.

**Kata Kunci:** Learning Object, metadata annotation, e-learning repository, text categorization, Natural Language Processing.

### 1 Pendahuluan

Ketersediaan konten pembelajaran yang banyak dalam bentuk media elektronik, yang dikenal sebagai bahan pembelajaran (LO), telah memacu penggunaan platform pembelajaran elektronik untuk memfasilitasi proses pembelajaran. Baik instruktur maupun pembelajar dapat mengakses LO yang disimpan dalam repositori; guru dapat menggunakan LO untuk menyusun kursus sementara peserta didik dapat mengambil LO yang relevan untuk mendukung tujuan pembelajarannya.

LO yang tersimpan pusat dalam repositori kemungkinan dapat digunakan kembali dalam konteks yang berbeda, sehingga LO menjadi sumber daya yang sangat menjanjikan untuk digunakan dalam berbagai lingkungan belajar seperti dalam mendukung program pelatihan masal. Setiap orang yang mempunyai akses akan berkesempatan untuk mendapatkan LO sesuai bidang yang diminati dari repositori tersebut. Namun, masalah bagaimana mencapai tingkat penggunaan

kembali LO ini secara massal masih merupakan tantangan besar sampai hari ini [1].

Untuk mencapai tingkat penggunaan yang tinggi sebuah LO sangatlah penting dibubuhi informasi yang berkaitan dengan topik konten di dalam metadata LO tersebut atau diistilahkan dengan anotasi metadata. Membuat anotasi sumber daya pendidikan dengan metadata terkait kompetensi membantu orang dan organisasi dalam mencari, mengambil, dan berbagi sumber belajar yang sesuai untuk program pengembangan kompetensi [2, 3]. Anotasi untuk tujuan ini terkandung dalam metadata LO dalam bentuk beberapa tag yang menjelaskan tentang artefak pendidikan dari LO terkait seperti judul dokumen, penulis, abstrak, dll. Setelah materi pembelajaran dibuat, maka pencipta harus berurusan dengan penjelasan metadata. Ini melibatkan penambahan beberapa informasi teks yang terkait dengan subjek konten, properti konten, dll. Ke dalam metadata LO. Sejumlah besar formulir

elektronik harus diisi dan agak sulit diisi untuk mendapatkan nilai yang benar [1]. Penandaan metadata manual menimbulkan beberapa masalah seperti mengisi formulir elektronik yang kompleks yang membuat tugas ini tidak menarik [4, 5], dan juga mahal dalam hal waktu dan upaya yang diperlukan. Akibatnya, menjelaskan *LO* dengan metadata secara manual kemudian menjadi tugas yang rumit.

Spesifikasi *IEEE Learning Object Metadata (LOM)* specification [6] adalah standar metadata yang paling banyak diadopsi untuk mengatasi sistem pedagogis. Selain menggambarkan informasi umum, ia juga memiliki informasi klasifikasi mengenai di mana *LO* berada dalam sistem klasifikasi tertentu. Menandai *LO* dengan topik yang sesuai dari sistem klasifikasi dapat meningkatkan kepuasan penelusuran pengguna dalam mencari *LO* terkait yang dibutuhkannya. Namun, penulis *LO* perlu menjelaskan *LO*-nya dengan topik seperti itu dari sistem klasifikasi tertentu yang bukan tugas yang mudah.

Dalam penelitian ini, penulis fokus pada anotasi *LOs* dengan sistem klasifikasi tugas kerja. Menanamkan taksonomi tugas kerja seperti itu dalam elemen kategori klasifikasi dapat membantu karyawan yang ingin menemukan *LO* yang terkait dengan tugas pekerjaannya. Item tugas pekerjaan, yang dijelaskan dalam suatu topik dan uraiannya, dapat dikorelasikan dengan satu atau lebih *LO* sebagaimana dilakukan dengan taksonomi topik seperti dijelaskan di atas. Namun, topik tugas pekerjaan dan uraiannya dibentuk dalam kalimat daripada kombinasi tunggal atau multi kata seperti yang ditemukan dalam kamus hierarki topik (lihat Gelbukh [7]), ontoLOgi domain (lihat Jovanovic et al. [8]) atau taksonomi silabus (lihat Roy et al. [9]). Topik tugas pekerjaan bisa berupa kalimat lengkap atau kombinasi multi-kata, dan uraiannya bisa berupa kalimat lengkap atau bahkan lebih.

Mengaitkan *LO* dengan topik-topik tugas kerja dapat dilakukan dengan mudah oleh manusia, namun, pencarian menyeluruh topik-topik tugas kerja terkait dari taksonomi bisa membosankan. Karena tugas ini lebih seperti proses menghubungkan makna antara dua kelompok kalimat, maka pendekatan *Natural Language Processing (NLP)* lebih cocok untuk digunakan.

Saat ini, NLP menjadi populer dalam aplikasi seperti sistem pencarian informasi. NLP berkaitan dengan interaksi antara komputer dan bahasa manusia (alami), dengan kata lain tujuan utama dari metode ini adalah untuk mendapatkan informasi yang disediakan dalam teks/dokumen untuk diproses oleh mesin. Penggunaan basis pengetahuan leksikal bahasa Inggris yang ada telah dilaporkan memiliki peningkatan signifikan dalam keakuratan pemrosesan otomatis aplikasi NLP. Ini didukung oleh peningkatan pengayaan dalam sumber daya dari proyek *WordNet* [10] dan *WordNetPlus* [11]. Hubungan semantik konsep dalam kosa kata *WordNet*

telah digunakan untuk menggantikan basis pengetahuan bawaan secara manual untuk domain subjek tertentu seperti ontologi kurikulum. Pendekatan kesamaan teks, yang diusulkan oleh Do et al. [12], membandingkan kesamaan dua cuplikan teks atau antara dua kalimat. Ukuran kesamaan yang disebut *WNSim*, yang merupakan kata kesamaan metrik berdasarkan hubungan semantik *WordNet* [10], menjadi dasar untuk menghitung kesamaan kalimat. Ukuran kesamaan akan memberikan hasil kategorisasi teks yang lebih sensitif terhadap konteks daripada makna tekstual.

Dalam tulisan ini, penulis meninjau fitur standar *IEEE LOM* yang dapat digunakan untuk penjelasan topik dalam sistem klasifikasi dan pendekatan ukuran kesamaan kalimat yang digunakan dalam perbandingan kesamaan. Selanjutnya, penulis menggambarkan karya eksperimen, hasilnya, dan analisisnya. Akhirnya, penulis menggambarkan skenario yang menerapkan penelitian ini, kompleksitas algoritma, dan menyimpulkan makalah.

## 2 Landasan Teori

### 2.1 Spesifikasi IEEE LOM (Learning Object Metadata)

Untuk memfasilitasi pembagian dan penggunaan kembali *LOs* di berbagai repositori informasi atau sistem manajemen pembelajaran (LMS), direkomendasikan bahwa *LO* harus dikaitkan dengan beberapa standar metadata umum. Ada beberapa standar metadata umum yang telah muncul untuk tujuan ini seperti inisiatif metadata *Dublin Core* (<http://dublincore.org>), metadata *IMS* (<http://imsglobal.org>), *CanCore* LRM ([www.cancore.ca](http://www.cancore.ca)) dan *IEEE LOM* (<http://ltsc.ieee.org/wg12/index.html>). Metadata *Dublin Core* berisi elemen metadata yang bermanfaat untuk aplikasi tujuan umum tetapi tidak mengandung atribut yang menggambarkan perspektif pedagogis suatu dokumen. Untuk mengatasi masalah pendidikan, standar lain telah dikembangkan.

Metadata Bahan pembelajaran IEEE (*IEEE LOM*) [6] telah secara luas diadopsi sebagai standar metadata untuk mengatasi perspektif pedagogis. Beberapa standar lain yang mendukung perspektif ini seperti *CanCore* dan *IMS* sepenuhnya kompatibel dengan standar ini. Standar ini menetapkan skema data konseptual yang mendefinisikan struktur instance metadata untuk bahan pembelajaran yang menggambarkan karakteristik yang relevan dari bahan pembelajaran yang berlaku. Karakteristik tersebut adalah kelompok dalam kategori Umum, Daur Hidup, Meta-metadata, Pendidikan, Teknis, Hak, Hubungan, Anotasi, dan Klasifikasi [6]. Standar ini dimaksudkan untuk dirujuk oleh orang lain sehingga instance metadata untuk bahan pembelajaran dapat digunakan kembali oleh sistem teknologi pembelajaran

untuk mengelola, menemukan, mengevaluasi atau untuk bertukar bahan pembelajaran.

### 2.1.1 Elemen General category (1, IEEE LOM)

Karakteristik paling dasar dari *LO* dijelaskan dalam kategori Umum. Kategori ini mengelompokkan informasi umum yang menggambarkan bahan pembelajaran ini secara keseluruhan. *IEEE LOM* menentukan bahwa setiap *LO* hanya diperbolehkan untuk memiliki satu elemen Umum (1, *IEEE LOM*) dan satu elemen Judul (1.2, *IEEE LOM*), yang berarti masing-masing elemen ini unik untuk *LO*. Namun, *LO* diizinkan untuk memiliki hingga maksimum sepuluh Identifier (1.1, *IEEE LOM*) elemen untuk memungkinkan identifikasi oleh skema katalog yang berbeda. Kata kunci (1.5, *IEEE LOM*) dan Deskripsi (1.4, *IEEE LOM*) juga diizinkan memiliki hingga sepuluh elemen untuk menjelaskan topik/konten *LO* yang dimaksud.

### 2.1.2 Elemen Classification category (9, IEEE LOM)

Elemen kategori klasifikasi (9, spesifikasi *LOM IEEE*) menjelaskan di mana *LO* ini berada dalam sistem klasifikasi tertentu. Perhatikan bahwa, ini berbeda dari elemen Identifier dalam kategori Umum, karena sistem klasifikasi biasanya mengkategorikan *LO* berdasarkan pada sistem taksonomi tertentu. *LOM IEEE* memungkinkan *LO* ditandai dengan hingga 40 elemen Klasifikasi dari sistem taksonomi yang berbeda. Elemen Tujuan (9.1, spesifikasi *LOM IEEE*) menjelaskan tujuan mengklasifikasikan *LO*. Ini unik untuk setiap sistem taksonomi yang ditandai dalam *LO*. Nilai dapat ditandai dengan label yang ditentukan dari spesifikasi *IEEE LOM* yang telah ditentukan sebelumnya, atau dengan nilai preferensi pengguna / pencipta.

Elemen takson jalur (9.2, spesifikasi *LEE IEEE*) dikhususkan untuk membubuhi keterangan jalur taksonomi node dalam sistem klasifikasi tertentu. Spesifikasi ini memungkinkan pemberian tag pada maksimum 15 elemen jalur Taxon dalam elemen Klasifikasi seandainya ada jalur berbeda dalam klasifikasi yang sama yang mengarah ke topik yang sama. Hal ini juga memungkinkan anotasi *LO* untuk lebih dari satu topik jika *LO* berada dalam banyak topik dalam sistem klasifikasi tertentu.

Elemen sumber (9.2.1, spesifikasi *IEEE LOM*) berisi nama sistem klasifikasi / taksonomi. Elemen ini dapat menggunakan taksonomi "resmi" yang dikenal atau taksonomi apa pun yang ditentukan pengguna. Spesifikasi *IEEE LOM* hanya memungkinkan satu elemen Sumber untuk ditandai untuk setiap lintasan Taxon yang berarti lintasan dan sumbernya harus ditentukan secara unik untuk suatu topik.

Elemen takson (9.2.2, spesifikasi *LOM IEEE*) menyarankan istilah tertentu dalam pohon taksonomi.

*Takson adalah simpul yang memiliki label atau istilah yang ditentukan. Label tekstual dari takson ini dijelaskan dalam elemen Entri (9.2.2.2, spesifikasi IEEE LOM). Takson juga dapat memiliki pengidentifikasi alfanumerik untuk referensi standar yang ditandai dalam elemen Id (9.2.2.1, spesifikasi IEEE LOM). Pengidentifikasi takson dapat berupa angka atau kombinasi huruf yang disediakan oleh sumber taksonomi.*

## 2.2 Ukuran Kesamaan Kalimat

Kesamaan kalimat, juga disebut sebagai kesamaan teks, bertujuan untuk membandingkan makna dari dua cuplikan teks sebagaimana dievaluasi dalam tugas-tugas seperti pengikatan teks, pengenalan parafrase, dan penjawaban pertanyaan [12, 13]. Banyak kasus penggunaan spesifik telah dilaporkan seperti untuk umpan balik relevansi dan klasifikasi teks [14], disambiguasi makna kata [15, 16] dan peringkasan ekstraktif [17], dan metode untuk evaluasi otomatis terjemahan mesin [18] atau peringkasan teks [19]. Ini juga berguna untuk evaluasi koherensi teks [20].

Sebagian besar karya terbaru melaporkan langkah-langkah linguistik mengungguli model ruang vektor dan tindakan kata-tumpang tindih. Langkah-langkah linguistik memanfaatkan pengetahuan linguistik seperti hubungan semantik antara kata-kata dan komposisi sintaksisnya untuk menentukan kesamaan kalimat, sedangkan model ruang vektor menghitung tingkat kesamaan kalimat berdasarkan frekuensi dokumen inversi frekuensi (TF-IDF). Word-overlap ukuran adalah keluarga ukuran kesamaan kombinatorial yang menghitung skor kesamaan berdasarkan sejumlah kata yang dibagi oleh dua kalimat. Achananuparp et al. [13] menyelidiki kinerja dari ketiga langkah kesamaan ini terhadap dataset yang tersedia untuk umum; TREC9 [21], MSRP [22], dan RTE3 [23]. Dia melaporkan bahwa ukuran linguistik adalah penampil terbaik untuk kalimat dengan kompleksitas rendah hingga tinggi. Namun, efektivitas ukuran linguistik ditentukan oleh proporsi kata dalam koleksi teks yang dicakup oleh basis pengetahuannya, dalam hal ini ia menggunakan WordNet [24]. Untuk mengatasi masalah ini, ia menyarankan menggunakan sumber daya pengetahuan lain selain ini seperti Wikipedia atau hasil pencarian web.

Do, et al. [12] merespons kekurangannya dengan memperkenalkan ukuran NESim untuk entitas yang disebutkan. Entitas yang dinamai di WordNet tersedia, tetapi langkah-langkah kesamaan yang ada tidak mengambil keuntungan ini. Dengan memanfaatkan jenis entitas bernama dan menggabungkan dengan ukuran kesamaan kata yang ada, Do, et al. Pendekatan ini mampu meningkatkan keakuratan kesamaan kalimat dalam pengenalan parafrase dan tugas teks.

Perhitungan LLM mengikuti rumus:

$$LLM_{(s_1, s_2)} = \frac{\sum_{v \in s_2} \max_{u \in s_1} sim(u, v)}{|s_2|} \quad (1)$$

di mana,  $sim(u, v)$  adalah ukuran kesamaan yang didefinisikan menggunakan kesamaan Wu dan Palmer [25]:

$$sim(u, v) = \frac{2 * N_3}{N_1 + N_2 + 2 * N_3} \quad (2)$$

di mana  $u$  dan  $v$  menunjukkan simpul kata sinkronisasi dalam hierarki basis data WordNet.

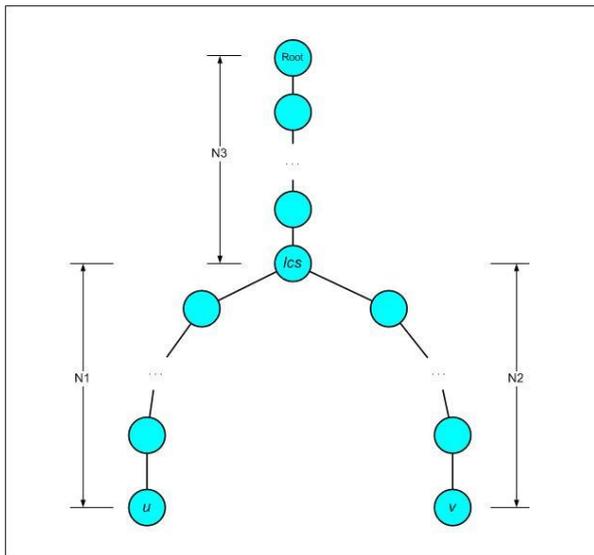


Figure 1. Konsep kesamaan Wu and Palmer

### 3 Metodologi Penelitian

Fokus dari percobaan ini adalah untuk menentukan kombinasi informasi teks dari metadata dan topik tugas pekerjaan yang harus dimasukkan dalam perbandingan kesamaan, kemudian untuk menentukan nilai ambang batas untuk kombinasi terbaik. Nilai ambang diperlukan oleh alat otomatis dalam menilai topik serupa, sehingga mampu menentukan kesamaan topik seperti manusia. Nilai ambang menentukan keputusan kesamaan antara dua topik teks.

#### 3.1 Data Sets

Sistem klasifikasi tugas pekerjaan dalam domain tugas pekerjaan yang terkait dengan komputer telah dibangun. Tugas-tugas pekerjaan ini diperoleh dari dokumen Panduan Teknis untuk Administrator Komputer dari Pegawai Negeri Sipil Indonesia [26]. Judul tugas pekerjaan, deskripsi, dan simpul induknya akan dimasukkan sebagai set fitur topik tugas-pekerjaan.

Sebagai sumber dataset topik  $LO$ , penulis mengumpulkan  $LO$  yang sesuai dengan topik tugas-pekerjaan yang berasal dari repositori gratis seperti MERLOT ([www.merlot.org](http://www.merlot.org)), EdNA ([www.edna.edu.au](http://www.edna.edu.au)), ARIADNE ([www.ariadne-eu.org](http://www.ariadne-eu.org)), dll.

Setiap pasangan topik (topik tugas-pekerjaan dan topik  $LO$ ) dinilai oleh penilai manusia apakah mereka memiliki topik yang sama. Keputusan penilaian ini akan digunakan sebagai referensi untuk mengevaluasi kinerja prediksi alat otomatis. Penulis dapat dengan mudah mengumpulkan fitur metadata  $LO$  karena beberapa repositori telah memberikan judul, kata kunci, dan deskripsi dalam  $LO$  mereka. Namun, bagian yang paling sulit adalah menemukan  $LO$  yang cocok karena setiap pekerjaan-tugas adalah topik yang sangat spesifik.

#### 3.2 Rancangan Eksperimen

Untuk menentukan nilai ambang terbaik, diperlukan untuk mengamati melalui percobaan. Nilai ambang akan dipilih melalui analisis prediksi kesamaan yang memberikan tingkat akurasi terbaik dibandingkan dengan penilaian manusia. Sembilan variasi telah disiapkan untuk percobaan. Setiap variasi menyertakan kombinasi fitur yang berbeda seperti yang ditunjukkan pada Tabel 1. Setiap variasi akan ditentukan untuk nilai ambangnya berdasarkan tingkat akurasi terbaiknya.

Table 1. Varian Desain Eksperimental

	Job Task *)		
	TT	TT+TD	TT+TD+TK
T	Var1	Var2	Var3
<b>LOM features **)</b> T+D	Var4	Var5	Var6
TDK	Var7	Var8	Var9

\*TT = Task Title, TD = Task Description, TK = Task Keywords

\*\*T = Title, T+D = Title + Description, TDK = Title + Description + Keywords

#### 3.3 Kriteria Evaluasi

Telah didefinisikan tiga ukuran evaluasi kinerja berdasarkan gagasan umum penilaian positif dan negatif dalam pencarian informasi dan klasifikasi teks. Ketepatan, ketepatan, daya ingat, dan ukuran F1. Langkah-langkah ini dianggap berdasarkan pada alasan yang dikemukakan oleh Tu [27] bahwa keakuratan saja tidak cukup untuk mengukur evaluasi classifier biner dan beberapa karya juga menyarankan untuk menggunakan ukuran F1 [12, 13].

Akurasi adalah bagian dari semua topik prediksi yang benar dari pasangan topik yang telah dinilai oleh penilai manusia sebagai sama atau tidak, dari semua pasangan topik. Prediksi yang benar dapat berupa *benar serupa (tp)* atau benar tidak-serupa (*tn*). Presisi adalah bagian dari prediksi yang benar untuk pasangan topik serupa (*tp*) dari semua jumlah pasangan topik yang diprediksi sama. Prediksi yang sama dapat berupa *benar serupa (tp)*

ataupun *serupa palsu (fp)*. *Recall* adalah bagian dari prediksi yang benar dari pasangan topik yang sama (*tp*) dari semua jumlah pasangan topik yang sebenarnya sama berdasarkan penilaian manusia. Pasangan aktual sebenarnya yang sebenarnya dapat berupa benar serupa (*tp*) ataupun *serupa palsu (fn)*. Nilai *recall* juga dikenal sebagai tingkat positif sejati atau indeks sensitivitas. F1 adalah rata-rata harmonik presisi dan daya ingat. Ini adalah ukuran evaluasi yang penting karena F1 memperhitungkan angka-angka ketidakseimbangan antara pasangan serupa yang sebenarnya dan pasangan sampel yang sebenarnya berbeda dalam percobaan [28]. Dengan kata lain, ukuran F1 merupakan trade-off antara presisi dan recall.

### 3.4 Pemrosesan awal Data

Untuk setiap dataset, penulis melakukan pra-pemrosesan yang melibatkan 4 langkah; yaitu penandaan sebagian kata, penghentian kata kunci dan tanda baca, pencocokan basis data WordNet, dan stemming, masing-masing seperti yang ditunjukkan pada Gambar 2.

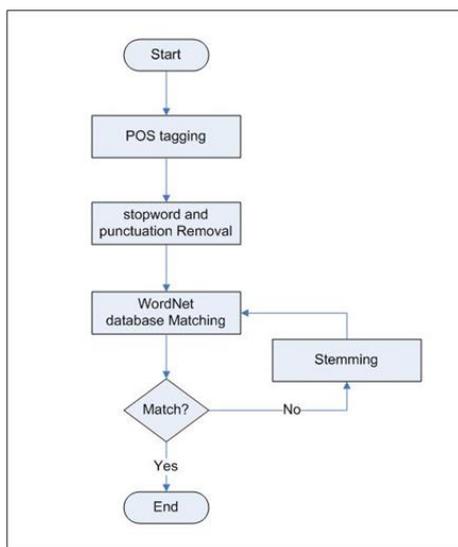


Figure 2. Kerangka kerja pra-pemrosesan kalimat

Tagging POS memberi anotasi pada setiap kata dengan label sintaksis. Langkah ini penting dalam mengurangi ambiguitas makna kata karena urutan kata dalam kalimat menentukan fungsi sintaksis kata dan, oleh karena itu, memengaruhi makna. Tagger POS yang diimplementasikan menggunakan perpustakaan berbasis Java yang disebut Illinois POS tagger [29] yang tokenizes kata-kata dan tanda baca berdasarkan fungsi mereka dalam kalimat.

Stopwords adalah kata-kata yang terlalu umum, seperti kata ganti, kata depan, dan kata sambung dalam bahasa Inggris, dan mereka sering muncul sehingga mereka tidak dapat memberikan informasi yang berguna

tentang konten dan dapat bersifat diskriminatif untuk kelas tertentu [30]. Menghapus stopwords dapat mengurangi efek kata-kata umum yang dapat mengaburkan konteks makna kalimat dan karenanya memengaruhi penghitungan kesamaan. Cara menghapus kata kunci dan tanda baca mengikuti daftar pemetaan yang mengklasifikasikan kata-kata token menjadi tiga kategori, yaitu 1) stopwords, 2) tanda baca, dan 3) kata-kata inti. Kata-kata inti adalah kata-kata yang mengandung makna konteks dalam kalimat mereka. Mereka terdiri dari empat jenis bentuk morfemik, yaitu kata benda, kata kerja, kata sifat dan kata keterangan. Penulis tidak menerapkan perubahan apa pun pada empat jenis kata yang telah dikenali oleh basis data WordNet.

Diperlukan untuk memastikan bahwa setiap kata dikenali dalam basis data WordNet, sehingga proses pra-pencocokan akan mencoba mencocokkan kata yang ditandai dengan basis data WordNet. Tag dari langkah POS-tagging berbeda dari tag WordNet. WordNet hanya mengenali empat bentuk semantik (mis., Kata benda, kata kerja, kata sifat, dan kata keterangan), sedangkan tag mungkin dalam bentuk bentuk morfemiknya. Namun, setelah langkah kedua dari fase pra-pemrosesan (mis., Menghapus kata-berhenti dan tanda baca) tag yang tersisa hanyalah perpanjangan dari empat bentuk sintaksis dasar di atas. Sebagai contoh, tag "gerund / present participle" dapat dikategorikan ke dalam sintaksis "kata kerja". Dengan demikian, bentuk-bentuk sintaksis yang diperluas ini dapat dengan mudah dikembalikan ke bentuk-bentuk dasar mereka yang pasti dikenali di WordNet. Mengambil kata-kata dari database difasilitasi oleh JAWS yang dibuat oleh Brett Spell [31].

Stemming diperlukan jika penulis tidak dapat menemukan kata yang tepat seperti yang dikenali dalam basis data WordNet. Stemming adalah preprocessing untuk mencari tahu morfem akar kata-kata. JAWS juga menyertakan fungsi penelusuran kata root untuk kata-kata morfemik non-root. Dengan menggunakan fungsi ini, melacak kata-kata morfem root menjadi proses yang terintegrasi sejak database kosakata WordNet dilibatkan.

## 4 Hasil dan Pembahasan

### 4.1 Hasil Percobaan

Mengacu pada Bagian 3.2 menggambarkan desain eksperimental, setiap varian hasil eksperimen diplot pada grafik untuk mendapatkan ukuran F1 maksimum. Nilai ambang batas optimal ditentukan berdasarkan ukuran F1 maksimum. Ambang batas optimal ini akan digunakan sebagai pengaturan ambang batas dalam aplikasi alat otomatis. Untuk memperoleh nilai ambang batas

optimal, 418 dataset telah dilatih untuk setiap varian dari 9 varian dan hasilnya ditampilkan pada Gambar 3.

The accuracy values shown in this table indicate the minimum accuracy that expected to be performed either in the further testing or in the tool's application. The accuracy value indicates the correct predictions rate of the setting threshold value. The highest accuracy value is performed by variant 6 and then followed by variant 9 and 8 as the second and the third respectively. This performance summary is also visualized in the Figure 3. As shown the figure, the highest F1 value is performed by four variants, i.e. variant 5, 6, 8 and 9, while the lowest is performed by variant 1 and 7.

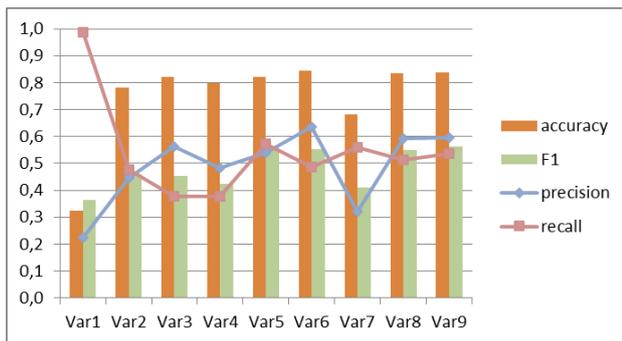


Figure 3. Training Dataset Performance

Pengaturan ambang akan digunakan sebagai referensi dalam alat otomatis. Namun, penting untuk diketahui kinerjanya sebelum menerapkan dalam aplikasi alat otomatis. Ini penting apakah nilai-nilai ini dapat diandalkan. Untuk alasan ini, nilai ambang diuji dengan 128 dataset pengujian. Kinerja pengujian ini secara grafis ditampilkan pada Gambar 4 yang menunjukkan akurasi yang dilakukan diuji dengan 128 dataset. Telah diamati bahwa semua nilai akurasi yang dilakukan lebih tinggi dari nilai akurasi yang diharapkan kecuali varian 4. Sementara itu, nilai akurasi tertinggi dilakukan oleh varian 9 dan kemudian diikuti oleh varian 8 dan dua varian mengikuti di tempat ketiga, yaitu , varian 5 dan 6, masing-masing.

Penampilan bagus yang dicapai oleh empat varian (5, 6, 8 dan 9) juga didukung oleh nilai F1 dan presisi yang lebih tinggi. Ini berbeda dengan semua varian 3, 4 dan 7 di mana nilai akurasi tinggi tidak didukung oleh kedua ukuran. Karena parameter presisi mengukur kemampuan untuk menyimpulkan topik yang sama (relevan dengan definisi nilai presisi oleh Kohavi dan Provost [32]) maka ketiga kandidat ini harus dikecualikan sebagai kandidat untuk digunakan dalam aplikasi.



Figure 4. Testing Dataset Performance

Ukuran lain, yakni *recall*, menunjukkan kemampuan untuk menyimpulkan topik yang berbeda kontras dengan presisi. Nilai penarikan juga disebut indikator sensitivitas. Mengevaluasi empat varian yang baik (5, 6, 8, dan 9), diamati bahwa varian 9 memiliki nilai recall tertinggi diikuti oleh varian 8, 5 dan 6, secara berurutan. Oleh karena itu, sangat disarankan untuk menggunakan varian 9 sebagai kandidat terbaik untuk menyimpulkan topik LO.

## 4.2 Analisa Hasil

### 4.2.1 Membandingkan hanya judul saja

Telah ditunjukkan bahwa varian 1 adalah yang terburuk, baik dalam dataset pelatihan atau dalam dataset pengujian. Varian ini hanya membandingkan dokumen hanya di antara judulnya, yaitu, judul LO dengan judul tugas pekerjaan. Perbandingan semacam itu tidak dapat menyimpulkan dokumen terkait topikal dengan benar. Termasuk lebih banyak informasi lain ke sisi komparator sumber, seperti deskripsi dan kata kunci LO, hanya meningkatkan akurasi sedikit seperti yang ditunjukkan oleh varian 4 dan 7 pertunjukan. Fenomena yang sama berlaku untuk sisi target seperti yang ditunjukkan pada varian 3 dan 2; meskipun varian 2 menunjukkan sedikit peningkatan. Fakta-fakta ini mendorong penggunaan semata-mata judul di satu sisi perbandingan harus dihindari.

### 4.2.2 Lebih banyak informasi akan lebih baik

Juga telah ditunjukkan bahwa semakin banyak informasi yang dimasukkan dalam perbandingan akan memberikan akurasi yang lebih baik. Pencocokan Tingkat Leksikal menggunakan database WordNet menggunakan kesamaan kata leksikal untuk menyimpulkan kesamaan dokumen. Karena semakin banyak informasi yang terlibat dalam perbandingan maka akan dapat membedakan maknanya dengan lebih jelas.

Varian 8 dan 9 sebagai dua tertinggi dalam pengujian, telah menunjukkan pentingnya memasukkan kata kunci dalam menyimpulkan kesamaan. Seperti yang kita ketahui bahwa kata kunci adalah kata atau frasa terpilih

yang digunakan oleh penulis LO dalam metadata. Fitur-fitur ini ternyata bermanfaat untuk menyimpulkan kesamaan antara dokumen.

#### 4.2.3 Hasil uji Recall and Presisi

Nilai presisi menunjukkan tingkat hasil yang diprediksi dengan benar (benar positif) dibandingkan dengan prediksi salah yang dianggap benar (salah positif). Sedangkan nilai recall mewakili tingkat hasil prediksi yang benar (true positive) dibandingkan dengan prediksi yang salah dianggap false (false negative).

Hasil percobaan ini menunjukkan nilai presisi yang adil (58,3 ~ 62,5%) dan nilai penarikan kembali yang rendah (26,9 ~ 38,5%). Menggunakan nilai ambang seperti yang direkomendasikan di atas dapat menghasilkan hasil negatif palsu yang tinggi, dengan kata lain akan ada sejumlah besar topik yang berbeda akan disimpulkan sama dan juga sejumlah topik yang sama akan ditolak. Keakuratan prediksi varian 9 sangat tinggi; ini karena nilai negatif benar tinggi yang meningkatkan nilai akurasi. Sudah disarankan untuk menggunakan varian 9 sebagai pemain terbaik; Namun, penggunaan rekomendasi ini dalam alat otomatis harus dengan hati-hati, karena keandalan dalam menyimpulkan topik yang berbeda tidak berarti dapat diandalkan juga untuk menyimpulkan yang serupa. Ini hanya cara untuk menyimpulkan kalimat yang sama dengan menghilangkan hasil yang berbeda.

## 5 Kesimpulan

Penggunaan kembali obyek pembelajaran ditingkatkan dengan kategorisasi metadata dan penandaan anotasi, sehingga memudahkan pengguna dalam mencari, mengambil, dan berbagi sumber belajar yang sesuai untuk program pengembangan kompetensi. Hasil percobaan menunjukkan bahwa keakuratan dapat ditingkatkan dengan memasukkan lebih banyak informasi dalam tugas perbandingan.

Namun, prediksi ini juga memiliki akurasi rendah yang disebabkan oleh rendahnya pengakuan modul tagger POS. Dengan demikian dalam pekerjaan masa depan dianjurkan untuk fokus pada peningkatan kata penutup tagger POS.

## Daftar Pustaka

- [1] K. Cardinaels, M. Meire, dan E. Duval, *Automating metadata generation: the simple indexing interface*, Proc. 14th International Conference on World Wide Web, 2005, pp. 548-556.
- [2] D. G. Sampson. Competence-related metadata for educational resources that support lifelong competence development programmes. *Educational Technology & Society*, vol. 12, pp. 149-159, 2009.
- [3] B. Chang, Y. Lee, S. Ko, dan J. Cha, *Enhancing ontology-based educational content search service with competency*, Eighth IEEE International Conference on Advanced Learning Technologies, ICALT'08., 2008, pp. 293-294.
- [4] E. Duval dan W. Hodgins, *A LOM research agenda*, Proc. 12th International Conference on World Wide Web, 2003, pp. 1-9.
- [5] J. Greenberg, A. Crystal, W. D. Robertson, dan E. Leadem, *Iterative design of metadata creation tools for resource authors*, International Conference on Dublin Core and Metadata Applications, 2003, pp. pp. 49-58.
- [6] I. L. T. S. Committee, Learning Object Metadata, Final Draft Standard, IEEE 1484.12. 1-2002, Available in: <http://ltsc.ieee.org/wg12/20020612-Final-LOM-Draft.html>, 2002.
- [7] A. Gelbukh, G. Sidorov, dan A. Guzman-Arenas, *Document comparison with a weighted topic hierarchy*, Proc. Tenth International Workshop on Database and Expert Systems Applications, 1999, pp. 566-570.
- [8] J. Jovanović, D. Gašević, dan V. Devedžić. Automating semantic annotation to enable learning content adaptation. *Adaptive Hypermedia and Adaptive Web-Based Systems, 2006*, pp. 151-160.
- [9] D. Roy, S. Sarkar, dan S. Ghose. A Comparative Study of Learning Object Metadata, Learning Material Repositories, Metadata Annotation & an Automatic Metadata Annotation Tool. *Advances in Semantic Computing, vol. 2*, 2010.
- [10] C. Fellbaum dan G. Miller, *Wordnet: An electronic lexical database*, (ed: MIT Press, 1998).
- [11] J. Boyd-Graber, C. Fellbaum, D. Osherson, dan R. Schapire, *Adding dense, weighted connections to wordnet*, Proc. 3rd International WordNet Conference, 2006, pp. 29-36.
- [12] Q. Do, D. Roth, M. Sammons, Y. Tu, dan V. G. V. Vydiswaran. Robust, light-weight approaches to compute lexical similarity. *Computer Science Research and Technical Reports, University of Illinois*. <http://hdl.handle.net/2142/15462>, 2010.
- [13] P. Achananuparp, X. Hu, dan X. Shen. The evaluation of sentence similarity measures. *Data Warehousing and Knowledge Discovery*. pp. 305-316, 2008.
- [14] J. J. Rocchio, Relevance Feedback in Information Retrieval, in G. Salton (Ed.), *The SMART Retrieval System - Experiments in Automatic Document Processing* (Prentice Hall, 1971).
- [15] M. Lesk, *Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone*, Proceedings of the 5th annual international conference on Systems documentation, Toronto, Ontario, Canada, 1986.
- [16] H. Schütze. Automatic word sense discrimination. *Computational linguistics, vol. 24*, pp. 97-123, 1998.
- [17] G. Salton, A. Singhal, M. Mitra, dan C. Buckley. Automatic text structuring and summarization. *Information Processing & Management, vol. 33*, pp. 193-207, 1997.
- [18] K. Papineni, *Why inverse document frequency?*, Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies, Pittsburgh, Pennsylvania, 2001.

- [19] C.-Y. Lin dan E. Hovy, *Automatic evaluation of summaries using n-gram co-occurrence statistics*, Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1, 2003, pp. 71-78.
- [20] M. Lapata dan R. Barzilay, *Automatic evaluation of text coherence: Models and representations*, International Joint Conference On Artificial Intelligence, 2005, p. 1085.
- [21] P. Achananuparp, X. Hu, X. Zhou, dan X. Zhang, *Utilizing sentence similarity and question type similarity to response to similar questions in knowledge-sharing community*, 17th International Conference on World Wide Web, 2008.
- [22] B. Dolan, C. Quirk, dan C. Brockett, *Unsupervised construction of large paraphrase corpora: Exploiting massively parallel news sources*, Proceedings of the 20th international conference on Computational Linguistics, 2004, p. 350.
- [23] I. Dagan, O. Glickman, dan B. Magnini. The pascal Recognising Textual Entailment Challenge. *Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Tectual Entailment*, pp. 177-190, 2006.
- [24] G. A. Miller. WordNet: a lexical database for English. *Communications of the ACM. vol. 38*, pp. 39-41, 1995.
- [25] Z. Wu dan M. Palmer, *Verbs semantics and lexical selection*, Proceedings of the 32nd annual meeting on Association for Computational Linguistics, 1994, pp. 133-138.
- [26] *Technical Guide of Accessing Credit Point of Computer Administrator*, S. C. Body 16/2008, 2008.
- [27] Y. Tu dan D. Roth. Learning English light verb constructions: contextual or statistical. *ACL HLT 2011*, p. 31, 2011.
- [28] H. Hamilton. *Computer Science 831: Knowledge Discovery in Databases*. 2000
- [29] D. Roth dan D. Zelenko, *Part of speech tagging using a network of linear separators*, Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 2, 1998, pp. 1136-1142.
- [30] K. Celik, *A Comprehensive Analysis of Using Wordnet, Part-Of-Speech Tagging, and Word Sense Disambiguation in Text Categorization*, Ph.D. dissertation, Bogaziçi University, 2012.
- [31] B. Spell. (2009, March 20, 2013). *Java API for WordNet Searching (JAWS)*. Available: <http://lyle.smu.edu/~tspell/jaws/index.html>
- [32] R. Kohavi dan F. Provost. Glossary of terms: Editorial for the Special Issue on Applications of Machine Learning and the Knowledge Discovery Process. *Journal of Machine Learning, vol. 30*, 1998.
- [33] V. Adamchik, *Algorithmic Complexity*, ed. <http://www.cs.cmu.edu/~adamchik/15-121/lectures/Algorithmic%20Complexity/complexity.html>, 2009.