

Implementasi dan Analisis Perbandingan Kesamaan *Controlled Vocabulary* pada *Library of Congress Subject Headings* dan *Tagging* pada *LibraryThing* dengan Metode *Wikipedia Explicit Semantic Analysis*

Implementation and Analysis Comparative Similarities of Controlled Vocabulary on Library of Congress Subject Headings and Tagging on LibraryThing with Wikipedia Explicit Semantic Analysis Method

Ilham Akbar¹, Dana S, Kusumo, Ph.D², Ibnu Asror, S.T., M.T³

^{1,2,3}Prodi S1 Teknik Informatika, Fakultas Teknik, Universitas Telkom

^{1s}ilhamakbar@student.telkomuniversity.ac.id, ²danakusumo@telkomuniversity.ac.id,

³iasror@telkomuniversity.ac.id

Abstrak

Information Architecture (IA) atau arsitektur informasi adalah suatu bentuk pengorganisasian penataan dan pelabelan terhadap suatu informasi yang diakses oleh pengguna. Seperti halnya dalam mengatur buku membutuhkan sebuah katalog yang dibuat secara tepat untuk memberi kemudahan bagi para pengguna dalam menemukan buku yang dicari. Pada awalnya sistem katalog hanya dituangkan dalam tulisan tangan manusia. Seiring berjalannya waktu, era digital semakin berkembang sehingga untuk katalogisasi buku dapat disebarluaskan dengan mudah menggunakan *tagging* melalui internet seperti pada *LibraryThing.com*. Sistem *tagging* dimana pengguna dapat memberi *tag* untuk menyimpan dan mengingat bahan pustaka yang pernah dibaca atau dikunjungi. Dalam hal ini memberikan kesempatan bagi penulis untuk membandingkan *tag* pada sistem informasi katalog buku untuk dicari kesamaannya terhadap tajuk subjek pada *Library of Congress Subject Headings (LCSH)*. Pada tugas akhir ini akan dicari kesamaan *tag* yang dibuat oleh pengguna *LibraryThing* terhadap tajuk subjek sebagai *controlled vocabulary* pada *Library of Congress Subject Headings (LCSH)* yang telah dikembangkan oleh para ahli dengan menggunakan metode *Explicit Semantic Analysis (ESA)* terhadap artikel *wikipedia* berbahasa Inggris. Berdasarkan data yang digunakan pada percobaan menemukan bahwa kata kunci yang dibuat oleh pengguna sebagai *tag* dan para ahli sebagai *LCSH* memiliki kesamaan sebesar 50%. Dengan melihat kesamaan ini, maka kata kunci akan dievaluasi oleh ahli pada saat menentukan tajuk subjek yang tepat dalam menganotasi buku, yaitu dengan lebih mempertimbangkan persepsi pengguna daripada ahli yang bertujuan untuk memudahkan pengguna dalam menemukan buku yang diinginkan.

Kata kunci : *tag, controlled vocabulary, Wikipedia, Explicit Semantic Analysis (ESA), tajuk subjek, LibraryThing, LCSH, arsitektur informasi.*

Abstract

Information Architecture (IA) is a form of organizing the arrangement and labeling of an information accessed by the user. Just as in organizing a book requires a precise catalog to make it easy for users to find the book they are looking for. Initially the catalog system was only poured in human handwriting. As time goes by, the digital age is expanding so that book cataloging can be easily disseminated using tagging via the internet like *LibraryThing.com*. Tagging system where users can tag to store and recall library materials that have been read or visited. In this case it provides an opportunity for the author to compare the tags on the catalog information system of books to look for similarity to the subject heading in the *Library of Congress Subject Headings (LCSH)*. In this final project will look for the similarity of tags created by *LibraryThing* users to subject headings as *controlled vocabulary* in the *Library of Congress Subject Headings (LCSH)* developed by experts using the *Explicit Semantic Analysis (ESA)* method of English *wikipedia* articles. Based on the data used in this experiment found that keywords created by users as tags and experts as *LCSH* have in common by 50%. By looking at these similarities, the keywords will be evaluated by an expert when determining the appropriate subject headline in the book, by more consideration of the user's perception than the expert intended to facilitate the user in finding the desired book.

Keywords: *tag, controlled vocabulary, Wikipedia, Explicit Semantic Analysis (ESA), subject headings, LibraryThing, LCSH, Information Architecture.*

1. Pendahuluan

Dewasa ini, pustakawan memiliki cara tersendiri untuk mengatur bahan koleksi buku-buku untuk pengguna perpustakaan agar dapat mencari buku sesuai yang dibutuhkan yaitu dengan cara membuat katalog. Katalog buku merupakan daftar dari koleksi buku dari perpustakaan yang disusun secara sistematis, sehingga memungkinkan pengguna perpustakaan dapat mengetahui dengan mudah koleksi yang dimiliki oleh perpustakaan dan dimana koleksi tersebut dapat ditemukan [1]. Namun dalam pemanfaatan sistem katalog buku memiliki beberapa masalah seperti perbedaan bahasa atau istilah yang digunakan dalam pemberian tajuk subjek. Tajuk subjek adalah kumpulan kata yang menentukan subjek buku serta menyatukan materi perpustakaan dibawah subjek yang sama [2]. Dalam penentuan tajuk subjek ini memiliki panduan yang dihasilkan oleh kesepakatan para ahli.

Kini katalog buku dapat diatur menggunakan sistem secara *online* dengan menggunakan penanda/label yang sering kali disebut sebagai *tag*, yaitu kata kunci yang berperan untuk mengelola dan mengatur serta menyimpan sebuah informasi[3]. Sistem *tagging* telah menjadi cara untuk mengatur katalogisasi buku seperti *LibraryThing.com* yaitu sebuah *website* yang menyediakan buku bacaan yang dapat disimpan dan berbagi katalog buku [4] dimana pengguna dapat memberikan *tag* terhadap buku yang tersedia. Penentuan tajuk subjek merujuk pada *Library of Congress Subject Headings (LCSH)* [5] yakni daftar tajuk subjek buku berbahasa Inggris dibawah naungan *Library of Congress*. Tajuk subjek dapat menjadi suatu cara mencari saat pengguna tidak mengetahui judul ataupun pengarang buku. Tajuk subjek yang dibuat oleh para ahli mempertimbangkan struktur kata dengan menggunakan *controlled vocabulary*, yaitu daftar kata yang dibatasi penggunaannya untuk keperluan tertentu yang mempunyai skema klasifikasi masing-masing[6]. Namun dalam hal ini terdapat perbedaan persepsi antara *tag* oleh pengguna dan tajuk subjek yang diberikan para ahli sehingga membuat pengguna kesulitan dalam menemukan buku yang dibutuhkan berdasarkan subjek yang dibuat oleh ahli[7]. Situasi ini membuat kesempatan untuk melakukan pencarian kesamaan tajuk subjek yang dibuat oleh para ahli terhadap *tag* yang diberikan oleh pengguna untuk mengetahui sejauh mana peran kata kunci tajuk subjek sebagai *controlled vocabulary* terhadap kemudahan pengguna dalam menemukan koleksi yang dicari.

Pada tugas akhir ini, Penulis akan menganalisis kata kunci *tag* yang diberikan oleh pengguna *LibraryThing* terhadap kata kunci tajuk subjek yang dibuat oleh para ahli pada *LCSH* untuk melihat kesamaan pasangan kata secara semantik yang bisa jadi memiliki kesamaan yang tinggi meskipun artinya berbeda. Data yang diberikan sebagai inputan adalah dua buah kata yang akan dinilai seberapa terkait kata-kata tersebut berdasarkan artikel *wikipedia* berbahasa inggris. Nilai kesamaan tersebut diberikan dengan rentang 0-1 berdasarkan tingkat kesamaannya. Metode yang digunakan dalam perhitungan ini yaitu dengan menggunakan metode *Explicit Semantic Analysis (ESA)*. Pendekatan ini dipilih karena merupakan salah satu metode yang memiliki kecocokan dengan struktur artikel *wikipedia* dalam perhitungan kesamaan kata berbasis vektor[8]. Artikel *Wikipedia* dapat didistribusikan atau diakses langsung kontennya oleh pengguna tanpa perlu dilakukan perhitungan sistematikal, oleh karena itu *wikipedia* cocok diimplementasikan dengan menggunakan metode *ESA* yang berarti eksplisit/langsung.

2. Dasar Teori

2.1 Controlled Vocabulary

Controlled Vocabulary atau kata kunci terkendali adalah daftar kata yang dibatasi penggunaannya untuk keperluan tertentu, yang tersedia dalam berbagai bentuk dan ukuran, subset dari kata kunci didefinisikan dari bahasa alami [6]. Sederhananya kata kunci tersebut adalah daftar istilah yang setara dalam bentuk sinonim, atau daftar pilihan istilah dalam bentuk file yang telah di sepakati oleh para ahli. Mendefinisikan hubungan hirarkis antara istilah yang mempunyai skema klasifikasi. Kata kunci terkendali menyediakan cara untuk mengatur pengetahuan untuk pengambilan berikutnya [3]. Pada tugas akhir ini *controlled vocabulary* adalah kata kunci tajuk subjek yang terdapat pada *The Library of Congress Subject Headings (LCSH)*. Contoh penggunaan, misalnya “ada pasien dalam kondisi demam”, maka kosa kata terkendali cukup mencantumkan salah satu istilah saja: demam atau pireksia, agar pengguna dapat memilih satu saja dari daftar kosa kata terkendali yang muncul. Dengan demikian, maka menjadi jelas bahwa ketika pengguna memilih “demam” dari kosa kata terkendali, maka yang dimaksudkan oleh pengguna adalah kondisi demam yang sedang dialami pasien[9].

2.2 Tag

Tag atau label adalah kata kunci yang berperan untuk mengelola dan mengatur sebuah informasi. *Tag* adalah tipe *metadata* yang menangkap atau menyimpan pengetahuan tentang sumber informasi [3]. Dalam hal ini, *tag* berupa kata kunci diberikan oleh pengguna untuk kemudahan dalam mencari buku yang sebelumnya sudah pernah di kunjungi. Contoh penggunaan, misal untuk item yang terkait pada buku “*Harry Potter*” user 1 memberikan *tag* “*harry*”, user 2 memberi *tag* “*magic*” user 3 memberi *tag* “*fantasy*” maka *tag* yang terdapat pada buku tersebut adalah “*harry, magic, fantasy*”. Dengan demikian menggunakan *tag* yang ada akan mempermudah koneksi ke item yang terkait melalui fitur *searching system*[10].

2.3 The Library of Congress Subject Headings (LCSH)

The Library of Congress Subject Headings (LCSH) adalah daftar kata kunci tajuk subjek yang paling banyak diadopsi bahasa pengindeksan subjek di dunia, telah diterjemahkan ke dalam banyak bahasa, dan digunakan di seluruh dunia oleh perpustakaan yang dikelola oleh *Library of Congress*[5]. Proposal untuk penambahan dan perubahan kata kunci direview secara berkala pada rapat staf dalam *Policy and Standard Division (PSD)* dan daftar tajuk subjek yang telah disetujui akan diterbitkan [5].

2.4 LibraryThing

LibraryThing adalah aplikasi katalog buku berbasis web untuk menyimpan dan berbagi katalog buku dan berbagai jenis metadata buku. Aplikasi *web* ini digunakan oleh penulis, perpustakaan, penerbit, maupun individu. Pengguna dapat mengambil informasi dari banyak perpustakaan yang ada di dunia seperti *Library of Congress*, *National Library of Australia*, dan lain-lain[4]. Pada tugas akhir ini penulis mengambil *tag* dari *LibraryThing*[4]. *Tag* pada web ini diberikan oleh pengguna dan dapat dilihat oleh pengguna lain bertujuan untuk kemudahan dalam mengakses buku melalui pencarian menggunakan *tag* tanpa harus memasukkan kata kunci bahasa yang terkendali (*controlled vocabulary*).

2.5 Explicit Semantic Analysis (ESA)

Explicit Semantic Analysis merupakan salah satu metode berbasis vektor yang mengupayakan pengklasifikasian sekumpulan dokumen secara eksplisit (tidak berbeli-belit) dan kategorinya diberikan oleh pihak eksternal[8]. Metode *ESA* ini cocok digunakan dengan artikel *wikipedia* sebagai pihak eksternal karena dalam metode ini, tiap kata yang menjadi inputan digambarkan sebagai dimensi vektor besar dan dibandingkan dengan kata-kata yang ada dalam artikel *wikipedia*. Dalam hal ini, eksplisit berarti teks inputan dicek secara langsung dari setiap kategori yang diberikan oleh *wikipedia* yang berupa judul artikel. Melihat bahwa struktur kata dari tiap judul artikel dalam *wikipedia* dapat digambarkan sebagai vektor-vektor besar yang membuat mereka cocok satu sama lain. Tiap komponen vektor merepresentasikan bobot *TF-IDF* antara *term* atau teks satu dengan dokumen kamus (*wikipedia*)[11].

Pada dasarnya, metode *ESA* dapat dibagi kedalam tiga proses utama, yaitu proses penerjemahan semantik, pembobotan nilai vektor dan perbandingan nilai vektor[8]. Penerjemahan semantik proses mulai dibuatnya vektor untuk masing-masing inputan dengan mencari frekuensi kemunculan dalam setiap dokumen (artikel). Pembobotan nilai vektor yaitu dengan menghitung bobot hasil frekuensi kemunculan *term* untuk menemukan nilai *tf.idf* dari setiap dokumennya. Sedangkan perbandingan vektor menghitung hasil pembobotan vektor dengan persamaan *cosine similarity* agar nilai kesamaan diketahui[8].

2.6 TF.IDF

Tf.idf merupakan singkatan dari *term frequency - inverse document frequency*, sedangkan bobot *tf.idf* yaitu bobot atau nilai yang sering digunakan dalam perhitungan *information retrieval* dan juga *text mining*[12]. Bobot ini merupakan perhitungan statistik yang sering digunakan untuk menentukan seberapa penting sebuah *term/kata* dari sebuah dokumen dalam kumpulan dokumen (artikel *wikipedia*) yang dapat digambarkan sebagai vektor. Nilai kepentingan tersebut meningkat sesuai dengan seberapa seringnya kata tersebut muncul dalam artikel. Untuk itu, menggunakan perhitungan *tf.idf* inilah tahap pembobotan vektor pada metode *ESA* dilakukan.

Biasanya, bobot *tf.idf* tersusun dari dua *term*[12]. Pertama menghitung normalisasi *term frequency (tf)*, yaitu jumlah seberapa sering kata muncul dari sebuah dokumen. Kedua adalah *invers document frequency (idf)*, yaitu nilai logaritmik yang merupakan nilai *invers/kebalikan* dari *document frequency (df)*. Semakin sering kata muncul dalam dokumen, maka bobot *idf* akan semakin kecil. *Document frequency (df)* merupakan jumlah dokumen yang mengandung kata tersebut, semakin sering kata muncul, maka nilai *df* akan semakin besar. Pembobotan dapat dilihat pada persamaan (2) berikut[12]:

$$\text{bobot} (tf) = 1 + \log_{10}(tf) \quad (1)$$

tf merupakan representasi jumlah kemunculan kata dari setiap dokumen/artikel. Kemudian untuk perhitungan *idf* dapat dilihat pada persamaan (3) berikut[12]:

$$\text{idf} = \log_{10}\left(1 + \frac{n}{df}\right) \quad (2)$$

df merupakan jumlah dokumen yang mengandung kata tersebut, sedangkan *n* merupakan jumlah total dokumen/artikel yang digunakan. Pada perhitungan ini dilakukan *smoothing* agar meminimalisir nilai nol pada perhitungannya, dengan menambahkan nilai 1 didalam persamaan logaritmik [4]. Jika bobot awal (*tf*) dan *idf* dari perhitungan ini sudah didapatkan, kemudian kalikan masing-masing bobot sesuai dengan dokumennya untuk mendapatkan bobot *tf.idf*-nya. Persamaan ini dapat dilihat pada persamaan (4) berikut[12]:

$$tf.idf = \text{bobot}(tf) \times idf \quad (3)$$

2.7 Cosine Similarity

Cosine Similarity merupakan perhitungan yang digunakan dalam tahap perbandingan vektor pada metode *ESA*[8]. Perhitungan ini menghitung nilai keterkaitan antara dua *non-zero* vektor dari *inner product* yang merepresentasikan derajat *cosine* dari kedua vektor tersebut[13]. Vektor yang dimaksud adalah vektor dari bobot *tf.idf* kata 1 dan kata 2, dan *non-zero* vektor berarti masing-masing vektor tersebut bukan merupakan vektor kosong. Jika kedua vektor tersebut memiliki derajat orientasi yang sama (sudut = 0), maka kedua vektor tersebut memiliki nilai *cosine similarity* sebesar 1, yang berarti sangat mirip/terkait. Derajat orientasi kedua vektor memiliki nilai dengan rentang antara 0-90. Untuk itu semakin besar derajat orientasi kedua vektor maka semakin kecil nilai *cosine similarity* yang dihasilkan[13]. Persamaan *cosine* dilihat dari persamaan (5) berikut[13]:

$$\text{Cos}(\theta) = \frac{A \cdot B}{|A| |B|} = \frac{\sum_{i=1}^n AB_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (4)$$

2.8 Wikipedia

Wikipedia adalah sebuah situs yang berisi tentang artikel ensiklopedia dan referensi online terbesar di dunia. *Wikipedia* didirikan oleh Larry Sanger dan Jim Wales yang dikelola oleh *Wikimedia Foundation*, sebuah perusahaan nirlaba yang bertempat di Amerika Serikat[14]. *Wikipedia* menggunakan *software open source* yang banyak digunakan oleh situs-situs berbasis ensiklopedia lainnya yang dapat di-download secara bebas di <https://dumps.wikimedia.org/enwiki/>[15].

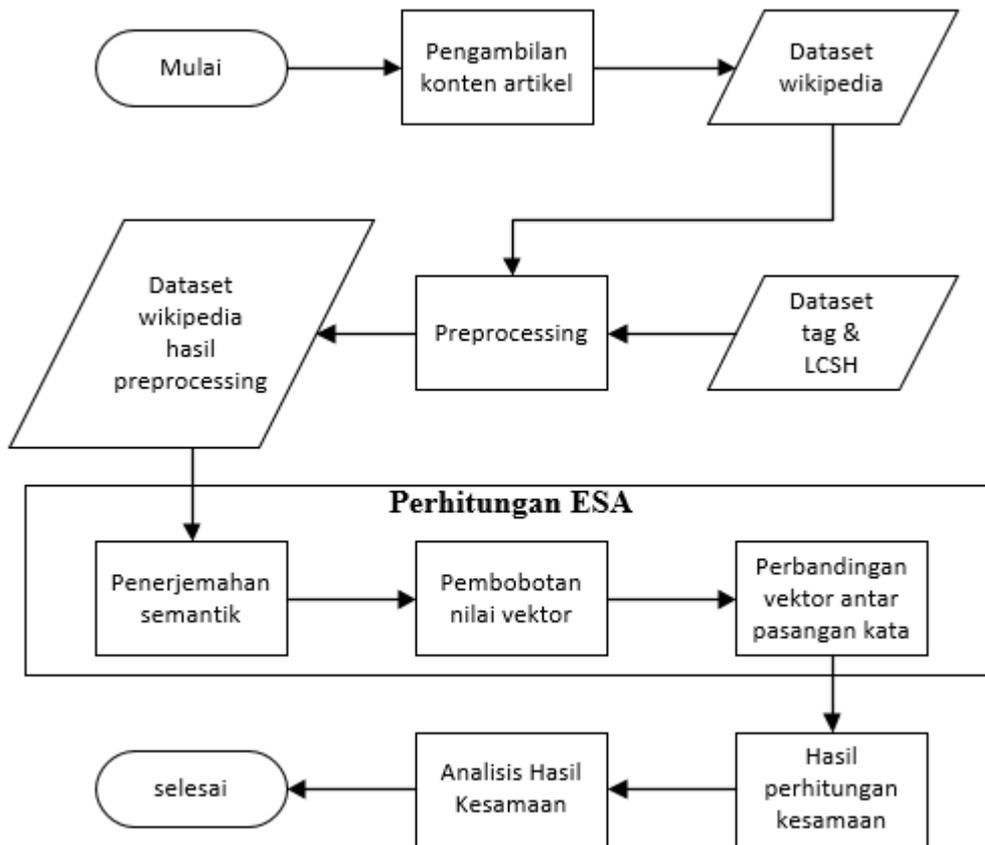
Wikipedia menyediakan artikel yang mendukung pengetahuan dasar perhitungan keterkaitan kata dengan lebih terstruktur dibanding *search engine* dan cakupan kata-nyapun lebih kompleks dibanding *WordNet*[16]. Artikel *Wikipedia* dapat didistribusikan atau diakses langsung kontennya oleh pengguna tanpa perlu dilakukan perhitungan sistematis, oleh karena itu *wikipedia* ini cocok diimplementasikan dengan menggunakan metode *ESA* yang berarti eksplisit/langsung. Judul-judul artikel yang terdapat dalam *wikipedia* dapat digambarkan berupa vektor besar[8]. Vektor dari judul-judul inilah yang nantinya akan dijadikan acuan database dalam perhitungan kesamaan antar pasangan kata dengan menggunakan metode *ESA*. Suatu kata dikatakan terkait dengan kata lain secara semantik dalam artikel *wikipedia* yaitu jika kata-kata tersebut banyak/sering terkandung dalam tiap judul artikel yang digunakan dalam perhitungan[8].

Judul-judul artikel yang menjadi acuan vektor direkap secara manual berdasarkan data *dump wikipedia*. Data *dump wikipedia* memiliki banyak struktur teks yang tidak dibutuhkan dalam perhitungan sistem, oleh karena itu untuk menyederhanakannya, isi konten artikel *wikipedia* berbahasa inggris diambil secara *online* menggunakan *API python* berdasarkan judul-judul yang sudah direkap dalam bentuk teks terlebih dahulu.

3. Perancangan Sistem

Sistem yang dibangun dalam tugas akhir ini adalah sebuah sistem yang mampu menganalisis tingkat kesamaan semantik *tag* pada *LibrayThing* terhadap kata kunci pada *LCSH* dibandingkan melalui artikel *wikipedia*. Tugas akhir ini akan membutuhkan data berupa *tag* pada *LibraryThing.com* untuk dibandingkan terhadap kata kunci atau tajuk subjek pada *LCSH* yang terlebih dahulu dipasangkan dan data artikel *wikipedia*. Data *LCSH* dan *tag* didapatkan dari *LibraryThing.com* sebanyak 150 buku dengan menggunakan *google scraper*. Data yang didapatkan dari jumlah buku tersebut adalah sebanyak 532 *tag* dan 312 *LCSH*, contoh data awal *LCSH* dan *tag* untuk 5 buku dapat dilihat pada tabel 3-1. Data artikel *Wikipedia* berbahasa inggris didapatkan dari *dumps Wikimedia* sebanyak 2000 judul artikel dari total 5.501.449 artikel per-tanggal 30 Oktober 2017. Data tersebut terlebih dahulu melalui tahap *preprocessing* untuk mendapatkan data yang mudah diproses dan dianalisis untuk menjamin tidak ada duplikasi data, menghilangkan kata yang tidak penting dan ambigu. Data-data tersebut dihitung secara semantik dengan menggunakan metode *Explicit Semantic Analysis (ESA)* untuk melihat kesamaannya. Kemudian *output* dari sistem ini berupa nilai dengan skala antara 0-1. Kemudian hasil dari kesamaan tersebut akan dianalisis sesuai kebutuhan pengujian.

Berikut ini merupakan penjelasan perancangan dari tiap tahapan dan komponen utama dalam perancangan sistem pada penelitian tugas akhir ini yang dapat dilihat pada Gambar 1.



Gambar 1 Gambaran Umum Sistem

Perhitungan nilai kesamaan semantik antara *tag* dan *LCSH* dengan menggunakan metode *ESA* melalui 3 proses. Pertama, proses penerjemahan semantik dengan menghitung jumlah kemunculan kata pada tiap artikel *wikipedia*. Kedua, dilakukan proses pembobotan nilai vektor untuk mendapatkan nilai bobot *tf.idf* dari setiap kata pada artikel. Pembobotan ini dihitung berdasarkan seberapa besar atau tinggi kesamaan kata tersebut terhadap tiap kata inputan. Proses pembobotan nilai vektor dengan menggunakan persamaan (1), persamaan (2), dan persamaan (3). Jika hasil seluruh bobot telah dihitung, kemudian akan masuk ke proses yang ketiga yaitu perbandingan vektor antara pasangan kata dari setiap perhitungan *ESA* dengan menggunakan persamaan (4) *cosine similarity* untuk mendapatkan nilai kesamaan semantiknya. Setelah itu hasil kesamaan tersebut akan dianalisis seberapa tinggi tingkat kesamaan *tag* yang dibuat oleh pengguna terhadap *LCSH* yang dibuat oleh ahli.

4. Pengujian dan Analisis

4.1 Analisis Hasil Kesamaan

Dalam sistem yang dibuat untuk mengetahui kesamaan/keterkaitan antara kata kunci yang dibuat oleh pengguna sebagai *tag* dan kata kunci yang dibuat oleh ahli sebagai tajuk subjek dengan menggunakan metode *ESA*. Metode ini juga digunakan dalam penelitian sebelumnya yang menjadi *paper* acuan[7], dimana antara *query* dengan dokumen yang dalam hal ini adalah buku perlu dihitung nilai untuk mengetahui hasil dari peran kata kunci dalam *searching* untuk menemukan koleksi buku yang dicari. Metode *ESA* digunakan karena salah satu teknik dalam melihat atau mendeteksi kesamaan yaitu dengan pendekatan secara langsung terhadap *wikipedia*.

Dalam penelitian ini didapat hasil kesamaan rata-rata semua buku sebesar 50% kata kunci tajuk subjek yang dianotasi oleh para ahli memiliki kesamaan terhadap *tag* yang dibuat oleh pengguna web *LibraryThing.com*. Karena banyaknya data hasil yang tidak memungkinkan untuk ditampilkan pada subbab ini maka hasil dari kesamaan dari buku-buku dapat dilihat pada lampiran.

Perlu diingat bahwa hasil dari penelitian ini tidak bisa sepenuhnya menjadi tolak ukur untuk memilih kata kunci yang tepat, karena dataset yang penulis gunakan hanya sebanyak 150 data buku sebagai pengujian tugas akhir sehingga tidak dapat mewakili seluruh dari total kata kunci yang dibuat oleh ahli dan pengguna. Dari hasil kesamaan ini membuktikan bahwa kata kunci yang dibuat oleh pengguna sebagai *tag* memiliki kesamaan pada nilai tengah. Hasil

dari analisis kesamaan ini berguna bagi ahli atau pustakawan untuk memberi kata kunci tajuk subjek yang tepat memperhatikan sudut pandang pengguna pada koleksi buku untuk mengatur katalog koleksi bahan pustaka.

4.2 Pengaruh Kepopuleran Buku

Pada pengujian ini melihat pengaruh kepopuleran buku terhadap seberapa banyak nilai kesamaan antara *tag* dan *LCSH* yang muncul pada tiap buku yang dijadikan data pada tugas akhir ini. Hasil dari pengujian dapat dilihat pada Tabel 1.

Tabel 1 Skenario Pengaruh Kepopuleran Buku

The City of Ember			The Neverending Story			Harry Potter and the Prisoner of Azkaban		
Tag (kata1)	LCSH (kata2)	Nilai	Tag (kata1)	LCSH (kata2)	Nilai	Tag (kata1)	LCSH (kata2)	Nilai
adventur	fantasi	0.472	centuri	Fantasi	0.216	adventur	audiobook	0.195
Chapter	fantasi	0.174	centuri	larg	0.768	adventur	adventur	1
Book	fantasi	0.359	book	Fantasi	0.359	board	audiobook	0.117
Children	fantasi	0.309	Book	Large	0.650	Board	adventur	0.228
Fiction	fantasi	0.569	Book	book	1	Board	Children	0.360
Literatur	fantasi	0.344	children	Fantasi	0.309	Board	School	0.423
Dystopia	fantasi	0.104	Childre n	Larg	0.541	school	Adventur	0.258
Dystopian	fantasi	0.126	children	Book	0.565	School	Fiction	0.330
Electr	fantasi	0.056	literatur	Fantasi	0.344	School	Children	0.599
Ember	fantasi	0	Literatu r	Larg	0.472	School	School	1
Fantasi	fantasi	1	Literatu r	Book	0.559	children	Audiobook	0.231
Friendship	fantasi	0.177	classic	Fantasi	0.230	children	Adventur	0.330
Futur	fantasi	0.188	Classic	Larg	0.547	children	Fiction	0.348
Futurist	fantasi	0.191	classic	Book	0.555	children	children	1
Juvenile	fantasi	0.184	Dragon	Fantasi	0.365	fiction	Audiobook	0.251
Mystery	fantasi	0.332	Fairi	Fantasi	0.420	fiction	Adventur	0.482
Postapocal y	fantasi	0.218	tale	Fantasi	0,517	fiction	fiction	1
Read	fantasi	0.258	Fantasi	Fantasi	1	england	Centuri	0.553
Science	fantasi	0.275	Fiction	Fantasi	0,569	england	England	1
Seri	fantasi	0.387	German	Fantasi	0.155	fantasi	Fantasi	1
Surviv	fantasi	0.152	german	Larg	0.590	Fantasi	Fiction	0.569
Teen	fantasi	0.221	juvenil	fantasi	0.184	juvenil	juvenil	1
Toread	fantasi	0	juvenil	book	0.190	Magic	fantasi	0.519
Citi	fantasi	0.191	sff	fantasi	0	Magic	Magic	1
Young	fantasi	0.290	Toread	Fantasi	0	Wizard	fantasi	0.422
adult	fantasi	0.296	unread	fantasi	0	wizard	wizard	1
Rata-rata		0.259	Rata-rata		0.508	Rata-rata		0,699

Penulis memilih 3 buku dari 150 buku dataset yang dikategorikan tingkat kepopuleran buku, ditetapkan berdasarkan pencarian terbanyak pada web *LibraryThing.com* dan pertimbangan pada hasil kesamaan yang didapat oleh sistem untuk tiap buku yang dibagi menjadi buku tidak populer, populer, dan sangat populer. Diberikan contoh sebanyak masing-masing 30 pasang kata yang terdapat pada buku “The City of Ember” sebagai buku tidak populer, “The Neverending Story” sebagai buku populer, dan “Harry Potter and the Prisoner of Azkaban” sebagai buku yang sangat populer.

Dari pengujian ini didapatkan bahwa pada buku yang sangat populer terdapat persepsi yang sama oleh pengguna dan para ahli dalam memberikan kata kunci yang tepat yang bernilai “1” (sangat mirip/terkait) sebanyak 10 pasangan kata, terlihat juga bahwa nilai kesamaan pasangan kata pada buku ini yaitu hampir sebesar 70%. Untuk buku yang populer ditemukan sebanyak 2 dari 30 pasangan kata yg sama. Walau hanya 2 kata yang ditemukan untuk tingkat kesamaan yang paling tinggi yaitu bernilai “1” namun hasil kesamaan dari rata-rata buku tersebut melebihi 50% sesuai

dengan rata-rata kesamaan buku secara total yang didapatkan oleh sistem. Untuk buku yang tidak populer, para ahli hanya memberikan 1 subjek saja yaitu fantasi, terlihat hanya sepasang kata yaitu “fantasi-fantasi” yang bernilai “1” didapatkan nilai rata-rata kesamaan pasangan kata yaitu sebesar 26%. Jadi kesimpulan yang didapat bahwa kepopuleran buku memiliki pengaruh dalam hal pemilihan kata kunci yang diberikan oleh pengguna dan para ahli. Dalam hal ini dengan semakin populernya buku maka semakin banyak yang membaca atau berkunjung ke web untuk mencari buku tersebut, sehingga akan banyak kata kunci yang diberikan yang membuat peluang kata kunci tersebut sama akan semakin besar.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan hasil analisis dari pengujian terhadap sistem yang dibuat, maka dapat diambil beberapa kesimpulan sebagai berikut :

1. Dengan data yang digunakan pada pengujian ini maka didapatkan bahwa *tag* yang dibuat oleh pengguna dan tajuk subjek yang dibuat oleh ahli menggunakan kata kunci yang sama sebesar 50%.
2. Hasil yang didapat dengan melihat kesamaan ini dapat menjadi pertimbangan ahli dalam memilih kata kunci untuk menganotasikan koleksi buku dengan memprioritaskan sudut pandang pengguna web.
3. Kepopuleran buku memiliki pengaruh dalam menentukan kata kunci yang tepat untuk digunakan pada subjek yang akan diberikan oleh para ahli karena dinilai sebagai salah satu pendorong untuk memberikan banyak kata kunci dalam hal ini *tag* oleh *pengguna* sehingga membesarkan peluang untuk memikirkan kata kunci yang sama.

5.2 Saran

Saran dari penelitian tugas akhir ini untuk penelitian selanjutnya adalah sebagai berikut :

1. Untuk penelitian selanjutnya sebaiknya dataset *tag* & *LCSH* dapat ditambahkan, serta artikel *wikipedia* yang dipilih harus berkaitan terhadap batasan permasalahan topik kata-kata yang dipakai guna mendapat akurasi yang lebih pasti untuk mewakili semua koleksi buku yang ada pada web *LibraryThing.com*
2. Ahli perlu berkolaborasi dengan orang yang memiliki disiplin ilmu terkait buku atau perpustakaan guna memilih kata kunci yang tepat untuk membuat tajuk subjek yang tepat pada buku.
3. Web *LibraryThing.com* akan tercapai memakai konsep arsitektur informasi yang terstruktur jika ahli lebih memikirkan sudut pandang pengguna, sebab sistem tersebut dibuat untuk pengguna agar tertarik menggunakan web dan tidak bingung dalam menemukan koleksi buku yang dicari.

Daftar Pustaka

- [1] J. Iqra and M. P. P. Iain-su, “Sistem Temu Kembali Informasi Dengan,” vol. 0, no. 2, pp. 1–8, 2013.
- [2] Perpustakaan RI, “tajuk subjek dalam konteks pengajaran dan penggunaannya di perpustakaan indonesia.” [Online]. Available: www.perpustakaan.go.id. [Accessed: 09-Nov-2016].
- [3] T. Bogers and V. Petras, “Tagging vs. Controlled Vocabulary: Which is More Helpful for Book Search?,” *iConference 2015 Proc.*, 2015.
- [4] LibraryThing.com, “LibraryThing.” [Online]. Available: <https://www.librarything.com/about>. [Accessed: 09-Nov-2016].
- [5] Library of Congress, “Library of Congress Subject Heading.” [Online]. Available: <https://www.loc.gov/aba/cataloging/subject/>. [Accessed: 09-Nov-2016].
- [6] P. Morville and L. Rosenfeld, *Information architecture for the world wide web*, vol. 73. 2006.
- [7] P. Heymann and H. Garcia-Molina, “Contrasting Controlled Vocabulary and Tagging: Do Experts Choose the Right Names to Label the Wrong Things?,” *Proc. 2nd ACM Int. Conf. Web Search Data Min. (WSDM-09), Late Break. Results Sess.*, pp. 1–4, 2009.
- [8] E. Gabrilovich and S. Markovitch, “Computing Semantic Relatedness using Wikipedia-based Explicit

- Semantic Analysis,” pp. 1606–1611, 2006.
- [9] P. R. A. Stegwee, “Introduction to Healthcare Standards and Health Level Seven Topics □ Need for electronic Healthcare Information Exchange □ Role of Healthcare Standards and benefits □ What is HL7 and examples of standards,” 2010.
- [10] I. T. Report, L. U. November, P. Heymann, and H. Garcia-molina, “Can Tagging Organize Human Knowledge ?,” *Science* (80-.), 2008.
- [11] E. Alhadeff and ג. בלום, *Search Engine Optimization for TF-IDF*. Hebrew University of Jerusalem, 2010.
- [12] Jana Vembunarayanan, “Tf-Idf and Cosine similarity.” [Online]. Available: <https://janav.wordpress.com/2013/10/27/tf-idf-and-cosine-similarity/>. [Accessed: 18-Oct-2017].
- [13] Wikipedia, “Cosine similarity.” [Online]. Available: https://en.wikipedia.org/wiki/Cosine_similarity. [Accessed: 18-Oct-2017].
- [14] Kamusq, “Wikipedia – Sumber artikel Terpercaya? | Pengertian dan Definisi.” [Online]. Available: <http://www.kamusq.com/2012/10/wikipedia-sumber-artikel-terpercaya.html>. [Accessed: 19-Oct-2017].
- [15] D. Vrandeć, “Semantic Wikipedia,” no. August 2007.
- [16] D. Milne and I. H. Witten, “An Effective , Low-Cost Measure of Semantic Relatedness Obtained from Wikipedia Links,” *Artif. Intell.*, vol. pp, pp. 25–30, 2007.

