

PENERAPAN METODE *TERM FREQUENCY (TF) - INVERS DOCUMENT FREQUENCY (IDF)* UNTUK PENCARIAN SINONIM KATA DALAM KAMUS BAHASA DAYAK NGAJU KALIMANTAN TENGAH

Devi Karolita¹, Ade Chandra Saputra²

^{1,2}Jurusan Teknik Informatika, Fakultas Teknik, Universitas Palangka Raya

¹devikarolita@it.upr.ac.id, ²adechandra@it.upr.ac.id

ABSTRACT

The available Bahasa Dayak Ngaju dictionary only in hardcopy form and only consisted of the list of words and the meaning of them. The dictionary didn't provide the synonym of the word we are looking for. If we want to know the paraphrase of one particular word, we have to do it manually by searching it on all of the list of words in the dictionary. Therefore, this research proposed to develop a Dayak Ngaju dictionary that provide word synonym search. We use *Term Frequency (TF) - Invers Document Frequency (IDF)* algorithm that will help synonym searching by comparing the term emergence in the meaning of the word that we are going to look the synonym for in all of the meaning of the words listed on the dictionary. This method will calculate the weight in all the documents based on the input query. To evaluate the methodology accuracy, we used precision and recall method. The results showed that after searching synonym of 100 queries, we founded that 33 queries had synonyms with precision = 1 and recall = 1 and all the synonyms that are found are relevant based on expert's justification.

Keyword: *TF-IDF, dictionary, Bahasa Dayak Ngaju*

1. PENDAHULUAN

Salah satu media komunikasi yang saat ini berkembang pesat adalah media internet. Karena dengan mengakses internet informasi apapun yang dibutuhkan dapat diperoleh dengan cepat. Sehingga tidak heran jika aktivitas sehari-hari masyarakat pada umumnya sekarang bergantung pada internet. Internet dimanfaatkan sebagai media untuk belajar, bisnis, gaya hidup dan lain sebagainya. Untuk pembelajaran di internet tersedia begitu banyak *e-book*, dan jurnal yang dapat di diperoleh atau di *download* tanpa harus membeli buku karena lebih murah dan cepat mendapatkannya.

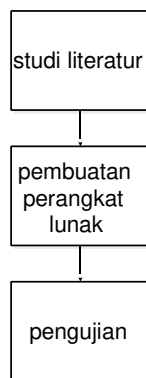
Banyak *website* yang menyediakan jenis kamus-kamus *online* yang jauh lebih efektif dan cepat daripada menggunakan kamus berupa buku. Kamus adalah buku acuan yang memuat kata dan ungkapan yang biasanya disusun menurut abjad berikut keterangan maknanya, pemakaiannya dan terjemahannya. Kamus biasanya berguna untuk seseorang yang ingin belajar bahasa atau kata-kata baru. Kamus Dayak Ngaju yang banyak diterbitkan hanya berupa buku kamus. Dan kamus tersebut hanya berisi daftar arti-arti kata atau terjemahannya ke bahasa Indonesia. Demikian juga dalam mencari sinonim sangat sulit karena harus mencari satu per satu di dalam

kamus yang isinya tidak teratur. Sinonim kata adalah ungkapan (kata, frasa, atau kalimat) yang kurang lebih sama maknanya dengan ungkapan yang lain.

Dari permasalahan diatas maka perlu dibuat aplikasi (*software*) bahasa Dayak Ngaju yang dapat melakukan pencarian sinonim. Dalam pencarian sinonim kata diperlukan sistem yang dapat menentukan dokumen yang paling tepat di antara dokumen-dokumen yang mengandung kata yang dicari. Dalam melakukan pencarian sinonim kata ini digunakan metode *Term Frequency (TF) - Invers Document Frequency (IDF)*. Dimana sistem ini memperhitungkan bobot pada setiap dokumen berdasarkan query yang dimasukan oleh pengguna sistem dan menampilkan hasil yang paling tepat.

2. METODOLOGI PENELITIAN

Dalam melakukan penelitian ini, metode yang dilakukan dapat dilihat pada Gambar 1.



Gambar 1. Metode Penelitian

1. Studi Literatur

Dasar teori dalam penelitian ini diperoleh dari referensi: buku, jurnal, artikel laporan penelitian, dan situs-situs di internet yang berkaitan dengan penelitian.

2. Pembuatan perangkat lunak

Adapun metodologi perancangan perangkat lunak yang dilakukan adalah metode pengembangan *Waterfall*.

3. Pengujian

Setelah melakukan perhitungan TF-IDF, maka diperoleh dokumen yang merupakan yang paling mirip atau sinonim query yang di *input* oleh pengguna pada sistem. Hasil pencarian yang ditampilkan sebagai sinonim kemudian dilakukan pengujian apakah sudah relevan. Cara mengukur kualitas dari pencarian sinonim ini adalah dengan menggunakan *precision* dan *recall*. Dalam uji coba dilakukan pencocokan *query* pada data di *database*. Pengujian dilakukan terhadap 500 buah *query* (kata) bahasa Dayak Ngaju. Untuk mengetahui kemampuan sistem yang diusulkan dilakukan perhitungan nilai presisi dan *recall* terhadap 100 kali uji coba dengan *query* yang berbeda. Dan pengujian hasil pencarian juga akan dibantu oleh ahli bahasa Dayak ngaju untuk memperoleh hasil yang benarbenar akurat.

3. HASIL DAN PEMBAHASAN

3.1. Tahapan Pencarian Sinonim

Langkah-langkah pencarian sinonim pada kamus Bahasa Dayak Ngaju adalah sebagai berikut.

1. Menerapkan *Text Processing*

a. *Tokenization*

Pada proses *tokenization* setiap kata dipisahkan dan menghilangkan tanda baca atau karakter serta mengubah setiap huruf besar menjadi

huruf kecil. Contoh hasil *tokenization* dapat dilihat pada Tabel 1.

Tabel 1. Contoh Hasil *Tokenization*

D1	bumbu penyedap masakan yang dibuat dari ikan kecil kecil atau udang yang dilumatkan halus halus belacan
D2	bumbu penyedap masakan yang dibuat dari ikan kecil kecil atau udang yang dilumatkan halus halus belacan
D3	bumbu penyedap masakan yang dibuat dari ikan kecil kecil atau udang yang dilumatkan halus halus belacan
D4	bumbu penyedap masakan yang dibuat dari ikan kecil kecil atau udang yang dilumatkan halus halus belacan

b. *Stopwords Removal*

Langkah berikutnya adalah menghilangkan *stopwords* atau kata-kata yang tidak terlalu penting. Adapun daftar *stopwords* yang dihapus sesuai pada *stopwords list* yaitu: yang, dari, dan atau. Tabel 2 menunjukkan hasil proses penghilangan *stopwords*.

Tabel 2. Contoh Hasil *Stopwords Removal*

D1	bumbu penyedap masakan dibuat ikan kecil kecil udang dilumatkan halus halus belacan
D2	bumbu penyedap masakan dibuat ikan kecil kecil udang dilumatkan halus halus belacan
D3	bumbu penyedap masakan dibuat ikan kecil kecil udang dilumatkan halus halus belacan
D4	bumbu penyedap masakan dibuat ikan kecil kecil udang dilumatkan halus halus belacan

c. *Stemming*

Pada proses *stemming*, setiap kata akan di kembalikan ke kata akar dengan menggunakan aturan Porter Stemmer untuk Bahasa Indonesia. Contoh hasil *stemming* menggunakan aturan Porter Stemmer untuk Bahasa Indonesia dapat dilihat pada Tabel 4.

Tabel 4. Contoh Hasil *Stemming*

D1	bumbu sedap masa buat ikan kecil kecil udang lumat halus halus belac
D2	bumbu sedap masa buat ikan kecil kecil udang lumat halus halus belac
D3	bumbu sedap masa buat ikan kecil kecil udang lumat halus halus belac
D4	bumbu sedap masa buat ikan kecil kecil udang lumat halus halus belac

2. Pembobotan TF/IDF

Tabel 5 menunjukkan hasil pembobotan menggunakan algoritma TF-IDF

Tabel 5. Perhitungan TF-IDF

term	TF				D F	D/ DF	IDF+1	W=TF*(IDF +1)			
	D 1	D 2	D 3	D 4				D1	D2	D3	D4
sedap	1	1	1	1	4	1	1	1	1	1	1
buat	1	1	1	1	4	1	1	1	1	1	1
halus	2	2	2	2	8	0.5	0.699	1.398	11.398	1.398	1.398
bumbu	1	1	1	1	4	1	1	1	1	1	1
masa	1	1	1	1	4	1	1	1	1	1	1
ikan	1	1	1	1	4	1	1	1	1	1	1
kecil	2	2	2	2	8	0.5	0.699	1.398	11.398	1.398	1.398
udang	1	1	1	1	4	1	1	1	1	1	1
lumat	1	1	1	1	4	1	1	1	1	1	1
balac	1	1	1	1	4	1	1	1	1	1	1

bobot (W) untuk D1 = 10.796
 bobot (W) untuk D2 = 10.796
 bobot (W) untuk D3 = 10.796
 bobot (W) untuk D4 = 10.796

Dari hasil pengurutan bobot dokumen diatas diperoleh bobot yang sama untuk setiap dokumen. Dokumen yang memiliki sinonim “acan” yang berada pada dokumen 1. sehingga diurutkan berdasarkan abjad yang ditampilkan sebagai sinonim dokumen 1, yaitu dokumen 2,3 dan 4.

3.3. Analisis Hasil Pengujian Kerja Sistem

Untuk perhitungan nilai *Precision* dan *Recall* terhadap 100 kali uji coba dengan *query* yang berbeda dapat dilihat secara lengkap pada Tabel 7.

Tabel 7. Perhitungan Nilai *Precision* dan *Recall* 100 Kali Uji Coba

No	input	sinonim	precision	recall
1	acan	blasas balasan malasan	P=3/3 =1	R=3/3 = 1
2	ajong	jukong arut lasang alut banama	P=5/5 =1	R=5/5 = 1
3	iwan	ayub sindah daup	P=3/3 =1	R=3/3 = 1
4	lilir	rilir	P=1/1 =1	R=1/1 = 1
5	jadi	jari	P=1/1 =1	R=1/1 = 1
6	jakah	abir	P=1/1 =1	R=1/1 = 1
7	bahua	taheta	P=1/1 =1	R=1/1 = 1
8	balaman	kasintu	P=1/1 =1	R=1/1 = 1
9	tabe	tabi	P=1/1 =1	R=1/1 = 1
10	rahun	ranum danum	P=2/2=1	R=2/2 = 1
11	katuk		P=0	R=0
12	nule		P=0	R=0
13	senam		P=0	R=0
14	ambon	enon	P=1/1 =1	R=1/1 = 1
15	andap	anap	P=1/1 =1	R=1/1 = 1
16	wujud		P=0	R=0
17	pander		P=0	R=0
18	pahelat		P=0	R=0
19	bau		P=0	R=0
20	bajan		P=0	R=0
21	gambuh		P=0	R=0
22	utara		P=0	R=0
23	bahekan		P=0	R=0
24	bajai		P=0	R=0
25	mujur		P=0	R=0
26	katarangan		P=0	R=0
27	umpamae		P=0	R=0
28	labih		P=0	R=0
29	uret		P=0	R=0
30	handipe		P=0	R=0
31	takiri		P=0	R=0
32	angguh	ayun	P=1/1 =1	R=1/1 = 1
33	panga		P=0	R=0
34	aduk		P=0	R=0
35	uras		P=0	R=0
36	badurok	hantuen	P=2/2 =1	R=2/2 = 1

				1
		hantimang	P=1/1 =1	R=1/1 = 1
37	gilau	gau	P=1/1 =1	R=1/1 = 1
38	ginau	ginyau	P=1/1 =1	R=1/1 = 1
39	gohap	jowoi	P=1/1 =1	R=1/1 = 1
40	gohong		P=0	R=0
41	iner		P=0	R=0
42	balawang		P=0	R=0
43	tungkeh		P=0	R=0
44	turus		P=0	R=0
45	lunju		P=0	R=0
46	tungkeh		P=0	R=0
47	sadar		P=0	R=0
48	bane	bana	P=1/1 =1	R=1/1 = 1
49	sunto		P=0	R=0
50	dawat		P=0	R=0
51	gantong		P=0	R=0
52	gawi		P=0	R=0
53	baha		P=0	R=0
54	uhat		P=0	R=0
55	umur		P=0	R=0
56	kasintu	balaman	P=1/1 =1	R=1/1 = 1
57	tunju		P=0	R=0
58	balok	kahau	P=1/1 =1	R=1/1 = 1
59	balon	galembong	P=1/1 =1	R=1/1 = 1
60	himun		P=0	R=0
61	tatawe		P=0	R=0
62	bangang	busik	P=1/1 =1	R=1/1 = 1
63	babang	ridu	P=1/1 =1	R=1/1 = 1
64	tumbuh		P=0	R=0
65	dahian		P=0	R=0
66	tatarik		P=0	R=0
67	tajaga		P=0	R=0
68	dadau	rau	P=1/1 =1	R=1/1 = 1
69	taimbing		P=0	R=0
70	damar	nyating	P=1/1 =1	R=1/1 = 1
71	adil		P=0	R=0
72	akik		P=0	R=0
73	lapik		P=0	R=0
74	capak	dampah	P=1/1 =1	R=1/1 = 1
75	dampahan		P=1/1 =1	R=1/1 = 1
76	duhi	dahiang	P=1/1 =1	R=1/1 = 1
77	riwut		P=0	R=0
78	ikau		P=0	R=0
79	gadur	kandarah	P=1/1 =1	R=1/1 = 1
80	gagelan	Kagelan	P=1/1 =1	R=1/1 = 1
81	garanat	manas	P=1/1 =1	R=1/1 = 1
82	gigir	gawing	P=1/1 =1	R=1/1 = 1
83	pantatan		P=0	R=0
84	tanggoi		P=0	R=0
85	topang		P=0	R=0
86	topeng		P=0	R=0

87	turus		P=0	R=0
88	jabong		P=0	R=0
89	laga		P=0	R=0
90	ongko		P=0	R=0
91	haga		P=0	R=0
92	bahari		P=0	R=0
93	pahir		P=0	R=0
94	antang		P=0	R=0
95	penda		P=0	R=0
96	aba	abai	P=1/1 =1	R=1/1 = 1
97	asep		P=0	R=0
98	ampit		P=0	R=0
99	sahap	tatudong	P=1/1 =1	R=1/1 = 1
100	manok		P=0	R=0

Dari Tabel 7, hasil pengujian dengan *precision* dan *recall* untuk 100 *query* berbeda yang diinput diperoleh *query* yang memiliki sinonim sebanyak 33 *query*. Sehingga diperoleh rata-rata hasil pencarian sinonim adalah:

$$\text{Rata-rata pencarian sinonim} = \frac{33}{100} \times 100 = 33\%$$

Hasil perhitungan rata-rata pencarian sinonim diperoleh 33% karena 67 *query* lain tidak ditemukan sinonimnya. Hal ini terjadi karena data didalam *database* belum lengkap.

4.KESIMPULAN

Dari hasil penelitian, dapat disimpulkan beberapa hal sebagai berikut.

1. Untuk pencarian sinonim dari 100 *query* yang dimasukkan dan yang memiliki sinonim adalah 33 *query* dan diperoleh nilai *Precision* =1 dan *Recall* = 1.
2. Hasil perhitungan rata-rata pencarian sinonim diperoleh 33%.
3. Relevansi dalam pencarian sinonim 100% tepat, artinya kualitas sistem dalam pencarian sangat baik.

DAFTAR PUSTAKA

- Indriyono Bonifacius Vicky, Utami Ema, Sunyoto Andi. 2015. *Pemanfaatan Algoritma Porter Stemmer Untuk Bahasa Indonesia Dalam Proses Klasifikasi Jenis Buku*. Program Pascasarjana Magister Teknik Informatika STMIK AMIKOM. Yogyakarta.
- Iper, Dunis.2009.*Kamus Bahasa Dayak Ngaju-Indonesia*. Palangkaraya: C.V. Anugrah Indah Mandiri. Iper, Dunis., Alexander, Darius Karel. 2016. *Kosakata Bahasa*

- Indonesia-Dayak Ngaju*. Palangkaraya: C.V. Anugrah Indah Mandiri
- Jogiyanto Hartono, 2005. *Analisis dan Desain Sistem Informasi*. Yogyakarta: Andi
- Dharwiyanti, Sri dan Satria W, Romi. 2003. *Pengantar Unified Modeling Language(UML)*. <http://www.ilmukomputer.com>. Diakses pada tanggal 26 September 2017.
- Dermawani, Novalina S. 2008. *Aplikasi Web Dinamis Menggunakan Php dan Mysql pada International Education Centre, Inc.* USU.Medan. (Http://Repository.Usu.Ac.Id/Bitstream/123456789/14004/1/09e00318.Pdf).
- Diakses pada tanggal 21 November 2016
- Sholihin, Andi. 2012. *Program Stemming Bahasa Indonesia Dengan Algoritma Porter Berbasis Web dengan PHP & MySQL*.
- Utami, Retno. 2010. *Kajian Sinonim Nomina Dalam Bahasa Indonesia*. Universitas Sebelas Maret. Surakarta. (http://eprints.uns.ac.id/9073/1/149581708201005051.pdf). Diakses pada tanggal 21 September 2017.