NAMED ENTITY RECOGNITION PADA TEKS BERBAHASA INDONESIA UNTUK PEMBANGKIT PERTANYAAN OTOMATIS

Rusliani[1], Ken Kinanti Purnamasari[2]
[1,2] Universitas Komputer Indonesia
Jl. Dipati Ukur 102, 112, 114, 116 Bandung
E-mail : liaein.ein@gmail.com[1], guinev_gs@yahoo.co.nz[2]

ABSTRAK

Pembangkit pertanyaan merupakan sebuah tugas yang secara otomatis membangkitkan pertanyaan dari berbagai macam masukan seperti teks, database, atau aliran semantic. Pada penelitian sebelumnya, pembangkit pertanyaan otomatis dengan menggunakan template berdasarkan analisis sintaksis telah berhasil dilakukan untuk kalimat tanya "siapa", "apa", "mana", dan "kapan". Penelitian ini menghasilkan akurasi yang masih rendah yaitu sebesar 14,53%. Hal ini dikarenakan belum adanya pendeteksian nama, ambiguitas yang belum bias teratasi, kamus yang digunakan masih belum lengkap, dan pencarian fungsi sintaksis yang belum sempurna.

Dalam mengatasi masalah tersebut, pada penelitian ini menggunakan NER untuk mengatasi pendeteksian nama dengan metode HMM. Proses yang dilakukan terdiri dari *preprocessing*; tokenisasi kalimat, tokenisasi kata, dan POS tag. Kemudian data dilatih dengan metode HMM. Dalam penelitian ini, konfigurasi yang dilakukan ada 3 (tiga), yaitu konfigurasi kelas kata terhadap kata tersebut, kelas kata terhadap kata tersebut dan kata sebelumnya, dan kelas kata terhadap kata tersebut, kata sebelum, dan setelahnya.

Berdasarkan hasil pengujian menggunakan data sebanyak 28992 tag entitas, didapatkan hasil nilai recall sebesar 0.0043, precision sebesar 0.1249, dan f-measure sebesar 0.0083, dan akurasi relevansi pelabelan kata sebesar 42,54% dan relevansi pertanyaan sebesar 19,71%.

Kata kunci : *named entity recogniton* (NER), *hidden markov model* (HMM), *part-of-speech* (POS) tag, pembangkit pertanyaan otomatis.

1. PENDAHULUAN

Pembangkit pertanyaan merupakan sebuah tugas yang secara otomatis membangkitkan pertanyaan dari berbagai macam inputan seperti teks, *database*, atau aliran semantik [1]. Banyak penelitian yang sudah dilakukan sebelumnya mengenai pembangkit pertanyaan dengan metode yang berbeda-beda baik pada teks berbahasa Indonesia maupun berbahasa Inggris.

Pada penelitian sebelumnya, pembangkit pertanyaan otomatis dengan menggunakan *template* berdasarkan analisis sintaksis telah berhasil dilakukan untuk kalimat tanya "siapa", "apa", "mana", dan "kapan". Penelitian ini menghasilkan akurasi yang masih rendah yaitu sebesar 14,53%. Hal tersebut disebabkan karena belum adanya pendeteksian nama, ambiguitas yang masih belum bisa teratasi , kamus yang digunakan masih belum lengkap, dan pencarian fungsi sintaksis yang belum sempurna [2]. Dalam penelitian ini, *Named Entity Recogniton* digunakan untuk mengatasi pendeteksian nama pada aplikasi pembangkit pertanyaan otomatis. *Named Entity* Recognition (NER) merupakan komponen pokok pada banyak aplikasi NLP (*Natural Language processing*) seperti *question answering*, *information extraction*, *clustering*, pendeteksian topik, dan peringkasan [3][4][5]. NER dapat digunakan untuk mengidentifikasi dan mendeteksi entitas seperti nama orang, lokasi, organisasi, waktu, dan lain-lain [1][4][5]. Banyak penelitian yang sudah menggunakan NER pada berbagai macam bahasa diantaranya bahasa Indonesia [6][7], Turki [2], Malaysia [3], Arab [8], Persia [9], India [1], dan Korea [10] dengan metode yang berbeda-beda.

Algoritma untuk NER diklasifikasikan menjadi 3 kategori; *rule-based*, *machine learning*, *dan hybrid* [4]. Dalam penelitian ini, algoritma yang digunakan adalah salah satu dari *machine learning* yakni *Hidden Markov Model* (HMM). Hal ini dikarenakan HMM mudah dimengerti, diimplementasi, dan dianalisis oleh sistem jika dibandingkan dengan *rule-based* dan *maximum entropy* [1].

Penelitian NER sebelumnya pernah dilakukan, yakni penelitian yang dilakukan oleh Muhammad Fachri mengenai NER berbahasa Indonesia dengan menggunakan metode HMM yang menghasilkan akurasi sebesar 88,4% [7], nilai tersebut menunjukkan bahwa NER dengan metode HMM terbukti ketepatannya. Oleh karena itu, dalam penelitian ini NER akan digunakan untuk pembangkit pertanyaan otomatis.

Berdasarkan uraian tersebut maka pada penelitian ini akan dilakukan Implementasi *Named Entity Recognition* pada Teks Berbahasa Indonesia menggunakan metode *Hidden Markov Model* untuk Pembangkit Pertanyaan Otomatis.

2. ISI PENELITIAN

Data masukan yang digunakan pada penelitian ini dibagi menjadi dua, yaitu data pelatihan dan data pengujian berupa file dengan format *.txt. Masukan yang diharapkan adalah teks berbahasa Indonesia yang memiliki tata bahasa yang baku. Berikut adalah bentuk data masukan yang digunakan, bersumber dari Universitas Gajah Mada.

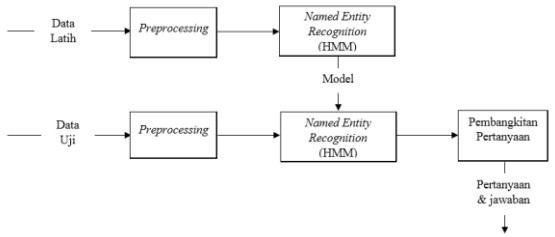
> Pengamat politik dari Universitas Gadjah Mada, Arie Sudjito, menilai, keinginan Ketua Umum Partai Golkar Aburizal Bakrie untuk maju kembali sebagai ketua umum merupakan pemaksaan kehendak. Menurut dia, ada kesan bahwa Aburizal menggunakan segala cara untuk memuluskan jalannya kembali menduduki Golkar 1. Hal ini, kata Arie, berpotensi menimbukan perpecahan di kalangan kader Golkar di daerah. "Tentu ini akan merugikan Golkar karena Golkar akan pecah dan konflik di daerah dipastikan akan terjadi," ujar Arie, saat dihubungi Kompas.com, Selasa (2/12/2014). Arie mengatakan, sikap Aburizal tersebut ditentang oleh sebagian kalangan di internal Golkar yang melakukan perlawanan dengan

Gambar 1. Data Masukan

2.1. Proses

Tahapan sistem NER ini dibagi menjadi 2 (dua) tahap, yaitu tahap pelatihan dan pengujian. Tahap pelatihan dimulai dengan pengguna mengunggah data latih yang berupa teks berisi korpus NER berbahasa Indonesia. Kemudian teks tersebut akan diproses, proses tersebut dinamakan tokenisasi. Tokenisasi yang dilakukan adalah tokenisasi terhadap kalimat dan kata. Setelah melalui proses tokenisasi, kata akan diklasifikasikan menurut jenisnya (kata kerja, kata benda, kata depan, dll), tahap ini dinamakan POS Tag. Setelah itu masuk ke tahap pemodelan dengan menggunakan Hidden Markov Model yang di dalamnya terdapat operasi perhitungan probabilitas awal, probabilitas transisi, dan probabilitas emisi. Setelah operasi perhitungan selesai, kemudian masuk ke algoritma viterbi untuk mencari kesimpulan berdasarkan pemodelan yang telah dilakukan sebelumnya.

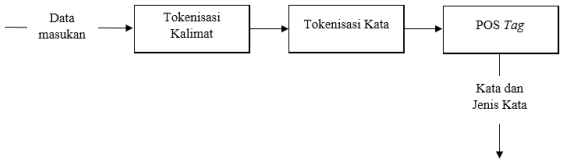
Pada tahap pengujian, data yang dimasukkan berupa data uji. Data uji merupakan korpus NER berbahasa Indonesia. Sama dengan data latih, data uji pun melalui tahap preprocessing hanya saja data tersebut akan diuji dengan pemodelan yang telah dilakukan pada proses pelatihan. Setelah itu, hasil dari pengujian akan diproses lagi ke tahap pembangkitan pertanyaan sehingga menghasilkan output berupa pertanyaan dan jawaban.



Gambar 2. Alur Proses

2.2. Preprocessing

Preprocessing pada penelitian ini terdiri dari beberapa tahap, yaitu tokenisasi kalimat, tokenisasi kata, dan POS Tag.



Gambar 3. *Preprocessing*

2.3. Tokenisasi Kalimat

Teks yang menjadi masukan berupa kumpulan dari beberapa kalimat, sehingga diperlukan tokenisasi kalimat untuk memisahkan kumpulan kalimat tersebut. Tokenisasi kalimat dilakukan dengan cara memisahkan kalimat dengan tanda titik (.) sebagai delimeter. Berikut adalah tokenisasi kalimat yang dilakukan terhadap data masukan.

Tabel 1. Tokenisasi Kalimat

No.	Kalimat
1	Pengamat politik dari Universitas Gadjah Mada, Arie Sudjito, menilai, keinginan Ketua Umum Partai Golkar Aburizal Bakrie untuk maju kembali sebagai ketua umum merupakan pemaksaan kehendak
2	Menurut dia, ada kesan bahwa Aburizal menggunakan segala cara untuk memuluskan jalannya kembali menduduki Golkar 1
3	Hal ini, kata Arie, berpotensi menimbukan perpecahan di kalangan kader Golkar di daerah.
4	"Tentu ini akan merugikan Golkar karena Golkar akan pecah dan konflik di daerah dipastikan akan terjadi," ujar Arie, saat dihubungi Kompas.com, Selasa (2/12/2014)
5	Arie mengatakan, sikap Aburizal tersebut ditentang oleh sebagian kalangan di internal Golkar yang melakukan perlawanan dengan membentuk Presidium Penyelamat Partai Golkar

2.4. Tokenisasi Kata

Setelah kalimat ditokenisasi, diperlukan proses tokenisasi kata untuk mendapatkan token dalam bentuk kata per kata. Proses ini dilakukan dengan cara memisahkan kalimat dengan delimeter spasi. Berikut adalah hasil dari tokenisasi kata.

Tabel 2. Tokenisasi Kata

No.	Kata		
1	Pengamat	keinginan	kembali
	politik	Ketua	sebagai
	dari	Umum	ketua
	Universitas	Partai	umum
	Gadjah	Golkar	merupakan
	Mada	Aburizal	pemaksaan
	Arie	Bakrie	kehendak
	Sutedjo	untuk	
	menilai	maju	
2	Menurut	menggunakan	kembali
	dia	segala	menduduki
	ada	cara	Golkar
	kesan	untuk	1
	bahwa	memuluskan	
	Aburizal	jalannya	
3	Hal	menimbulkan	Golkar
	ini	perpecahan	di
	kata	di	daerah
	Arie	kalangan	
	berpotensi	kader	
4	Tentu	pecah	ujar
	ini	dan	Arie
	akan	konflik	saat
	merugikan	di	dihubungi
	Golkar	daerah	Kompas.com
	karena	dipastikan	Selasa
	Golkar	akan	2/12/2014
	akan	terjadi	
5	Arie	sebagian	perlawanan
	mengatakan	kalangan	dengan
	sikap	di	membentuk
	Aburizal	internal	Presidium
	tersebut	Golkar	Penyelamat
	ditentang	yang	Partai
	oleh	melakukan	Golkar

2.5. Part-of-Speech Tagging

POS Tagging merupakan klasifikasi jenis kata. POS Tagging yang digunakan pada sistem ini adalah POS Tagging dari Alfan [11]. Sehingga dihasilkan sebagai berikut.

1.

Pengamat	politik	dari	Universitas	Gadjah
NN	**NN**	**IN**	**NN**	**NN**
Mada	Arie	Sudjito	menilai	keinginan
NNP	**NNP**	**NNP**	**VBT**	**NN**
Ketua	Umum	Partai	Golkar	Aburizal
NN	**JJ**	**NN**	**NN**	**NN**
Bakrie	untuk	maju	kembali	sebagai
NNP	**IN**	**VBI**	**RB**	**IN**
ketua	umum	merupakan	pemaksaan	kehendak
NN	**JJ**	**VBT**	**NNP**	**NN**

2.

Menurut	dia	ada	kesan	bahwa
VBT	**PRP**	**VBI**	**NN**	**SC**
Aburizal	menggunakan	segala	cara	untuk
NN	**VBT**	**NN**	**NN**	**IN**
jalannya	kembali	menduduki	Golkar	1
NNG	**RB**	**VBT**	**NN**	**CDP**

3.

Hal	ini	kata	Arie	berpotensi
NN	**DT**	**VBT**	**NNP**	**NN**
menimbulkan	perpecahan	di	kalangan	kader
VBT	**NN**	**IN**	**NN**	**NN**
Golkar	di	Daerah		
NN	**IN**	**NN**		

4.

Tentu	ini	akan	merugikan	Golkar
NN	**DT**	**MD**	**VBT**	**NN**
karena	Golkar	akan	pecah	dan
CC	**NN**	**MD**	**VBT**	**CC**
di	daerah	dipastikan	akan	terjadi
IN	**NN**	**VBT**	**MD**	**VBI**
ujar	Arie	saat	dihubungi	Kompas.com
VBT	**NNP**	**SC**	**VBT**	**NNP**
Selasa	2/12/2014			
NN	**NN**			

5

Arie	mengatakan	sikap	Aburizal	tersebut
NNP	**VBI**	**NN**	**NN**	**DT**
ditentang	oleh	sebagian	kalangan	di
NN	**IN**	**CDI**	**NN**	**IN**
Golkar	yang	melakukan	perlawanan	dengan
NN	**SC**	**VBT**	**NN**	**IN**
membentuk	Presidium	Penyelamat	Partai	Golkar
VBT	**NN**	**NN**	**NN**	**NN**

Gambar 4. *POS Tag*

2.6. Pelatihan

Data masukan diproses agar setiap kata yang ada pada teks ditandai dengan entitas yang sesuai. Proses ini dikerjakan secara manual [7]. Berikut adalah contoh penandaan entitas yang digunakan.

1.

Pengamat	politik	dari	Universitas	Gadjah
OTHER	**OTHER**	**OTHER**	**ORG**	**ORG**
Mada	Arie	Sudjito	menilai	keinginan
ORG	**PER**	**PER**	**OTHER**	**OTHER**
Umum	Partai	Golkar	Aburizal	Bakrie
OTHER	**ORG**	**ORG**	**PER**	**PER**
untuk	maju	kembali	sebagai	ketua
OTHER	**OTHER**	**OTHER**	**OTHER**	**OTHER**
umum	merupakan	pemaksaan	kehendak	
OTHER	**OTHER**	**OTHER**	**OTHER**	

2.

Menurut	dia	ada	kesan	bahwa
OTHER	**OTHER**	**OTHER**	**OTHER**	**OTHER**
Aburizal	menggunakan	segala	cara	untuk
PER	**OTHER**	**OTHER**	**OTHER**	**OTHER**
memuluskan	jalannya	kembali	menduduki	Golkar

	OTHER 1	**OTHER**	**OTHER**	**OTHER**	**ORG**
3.	**OTHER** Hal	ini	kata	Arie	berpotensi
	OTHER menimbulkan	**OTHER** perpecahan	**OTHER** di	**PER** kalangan	**OTHER** kader
	OTHER Golkar	**OTHER** di	**OTHER** daerah	**OTHER**	**OTHER**
4.	**ORG** Tentu	**OTHER** ini	**OTHER** akan	merugikan	Golkar
	OTHER karena	**OTHER** Golkar	**OTHER** akan	**OTHER** pecah	**ORG** dan
	OTHER konflik	**ORG** di	**OTHER** daerah	**OTHER** dipastikan	**OTHER** akan
	OTHER terjadi	**OTHER** ujar	**OTHER** Arie	**OTHER** saat	**OTHER** dihubungi
	OTHER Kompas.com	**OTHER** Selasa	**PER** 2/12/2014	**OTHER**	**OTHER**
5	**OTHER** Arie	**TIME** mengatakan	**TIME** sikap	Aburizal	tersebut
	PER ditentang	**OTHER** oleh	**OTHER** sebagian	**PER** kalangan	**OTHER** di
	OTHER internal	**OTHER** Golkar	**OTHER** yang	**OTHER** melakukan	**OTHER** perlawanan
	OTHER dengan	**ORG** membentuk	**OTHER** Presidium	**OTHER** Penyelamat	**OTHER** Partai
	OTHER Golkar	**OTHER**	**ORG**	**ORG**	**ORG**
	ORG				

Gambar 5. Penandaan Entitas

Setelah data diberi label secara manual [7], selanjutnya masuk ke proses pelatihan. Proses pelatihan pada sistem ini digunakan untuk mendapatkan nilai probabilitas awal, matriks probabilitas transisi dan matriks probabilitas emisi. Berdasarkan penelitian dari Sudha Morwal, Nusrat Jahan, dan Deepti Chopra [1] probabilitas awal dapat menggunakan persamaan (1).

(1)

$$\pi = \frac{tag\ entitas\ di\ awal\ kalimat}{jumlah\ kalimat}$$

Tabel 3. Probabilitas Awal

PER	LOC	ORG	TIME	QTY	OTHER
1/5= 0.2	0	0	0	0	4/5= 0.8

Setelah mendapatkan probabilitas awal, selanjutnya adalah mencari probabilitas transisi yang berdasarkan penelitian dari Sudha Morwal, Nusrat Jahan, dan Deepti Chopra [1] dapat menggunakan persamaan (2).

(2)

$$A = \frac{jumlah\ urutan\ tag\ dari\ t_i\ ke\ t_j}{jumlah\ tag\ t_i}$$

Tabel 4. Probabilitas Transisi

	PER	LOC	ORG	TIME	QTY	OTHER
PER	0.2222	0	0	0	0	0.7778
LOC	0	0	0	0	0	0
ORG	0.1538	0	0.4615	0	0	0.3846
TIME	0.5	0	0	0.5	0	0
QTY	0	0	0	0	0	0
OTHER	0.0548	0	0.1096	0.0137	0	0.8082

Setelah probabilitas awal dan probabilitas transisi terpenuhi, langkah berikutnya yaitu mencari probabilitas emisi yang berdasarkan penelitian dari Sudha Morwal, Nusrat Jahan, dan Deepti Chopra [1] dapat menggunakan persamaan (3).

(3)

$$B = \frac{kelas\ kata\ dari\ tag\ entitas}{jumlah\ tag\ entitas}$$

Dalam penelitian ini terdapat tiga konfigurasi urutan kelas kata yang akan dilakukan. Konfigurasi tersebut adalah sebagai berikut.

1. Kelas kata pada kata tersebut
2. Kelas kata pada kata tersebut dan pada kata sebelumnya
3. Kelas kata pada kata tersebut, kata sebelum, dan sesudahnya

Tabel 5. Probabilitas Emisi Konfigurasi 1

	PER	LOC	ORG	TIME	QTY	OTHER
NNP	0.6667	0	0.0769	0	0	0.0274
VBI	0	0	0	0	0	0.0137
NN	0.3333	0	1	1	0	0.0548
DT	0	0	0	0	0	0.0137
IN	0	0	0	0	0	0.0274
CDI	0	0	0	0	0	0.0137
JJ	0	0	0	0	0	0.0137
SC	0	0	0	0	0	0.0137
VBT	0	0	0	0	0	0.0137

2.7. Pengujian

Pengujian pada sistem ini menggunakan algoritma Viterbi, dengan persamaan (4), (5), (6), dan (7).

1. Inisialisasi

$$\delta(i) = \pi_i b_t(O_1) \tag{4}$$

$$\varphi_1(i) = 0$$

2. Rekursi

$$\delta_t(j) = \max_{1 \le 1 \le N} \left[\delta_{t-1(i)} a_{ij}\right] b_j(O_t) \tag{5}$$

$$\psi_t(j) = \operatorname*{argmax}_{1 \le 1 \le N} \left[\delta_{t-1(i)} a_{ij}\right]$$

3. Terminasi

$$P^* = \max_{1 \le I \le N} \left[\delta_{T(i)}\right] \tag{6}$$

$$Q^*T = \operatorname*{argmax}_{1 \le I \le N} [\delta_{T(i)}]$$

4. Lacak Balik

$$Q_t^* = \varphi_{t+1}(Q_{t+1}^*) \tag{7}$$

Dari proses-proses tersebut, dapat disimpulkan bahwa state yang mungkin dengan peluang terbesar untuk observasi {Arie, mengatakan, sikap, Aburizal, tersebut, ditentang, oleh, sebagian, kalangan, di, internal, Golkar, yang, melakukan, perlawanan, dengan, membentuk, Presidium, Penyelamat, Partai, Golkar} adalah {PERSON, OTHER}. Dari proses-proses tersebut, dapat disimpulkan bahwa *state* yang mungkin dengan peluang terbesar untuk observasi {*Arie, mengatakan, sikap, Aburizal, tersebut, ditentang, oleh, sebagian, kalangan, di, internal, Golkar, yang, melakukan, perlawanan, dengan, membentuk, Presidium, Penyelamat, Partai, Golkar*} adalah {*PERSON, OTHER*}.

2.8. Pembangkit Pertanyaan

Setelah tiap kata terdeteksi tag entitasnya, proses berikutnya adalah mencari fungsi-fungsi sintaksis seperti subjek, predikat, objek, dan keterangan. Kalimat tanya akan dibangkitkan menggunakan *template* berdasarkan fungsi-fungsi sintaksis yang telah ditemukan. Proses pembangkitan pertanyaan dalam penelitian ini dibagi menjadi dua proses yaitu analisis sintaksis untuk mencari fungsi sintaksis dan pembangkitan pertanyaan untuk membangkitkan pertanyaan berdasarkan template sintaksis.

Pada penelitian sebelumnya, pencarian fungsi sintaksis dilakukan dengan membagi klausa yang bersifat predikatif. Pada penelitian ini, pencarian fungsi sintaksis tersebut dilakukan berdasarkan POS *Tag* yang telah didapat sebelumnya. Sehingga untuk menemukan klausa yang bersifat predikatif, cukup dengan menemukan jenis kata (kata kerja). Berikut adalah hasil pertanyaan yang berhasil dibangkitkan.

Tabel 6. Pembangkitan Pertanyan

Kalimat	Pertanyaan yang dibangkitkan
Arie mengatakan, sikap Aburizal tersebut ditentang oleh sebagian kalangan di internal Golkar yang melakukan perlawanan dengan membentuk Presidium Penyelamat Partai Golkar	1. Siapa yang mengatakan sikap Aburizal tersebut ditentang oleh sebagian kalangan di internal Golkar? 2. Apa yang ditentang oleh sebagian kalangan di internal Golkar? 3. Apa yang sebagian kalangan lakukan pada Aburizal? 4. Apa yang dibentuk oleh sebagian kalangan di internal Golkar?

3. Hasil Pengujian

Pengujian performansi dilakukan dengan menghitung recall, precision, dan f-measure yang ada pada persamaan (8), (9), (10).

$$Recall = \frac{Correct \ + \ 0.5 * Partial}{Possible} \tag{8}$$

$$Precision = \frac{Correct \ + \ 0.5 * Partial}{Actual} \tag{9}$$

$$F - measure = \frac{Recall * Precision}{0.5 * (Recall + Precision)} \tag{10}$$

Tabel 7. Pengujian Perfomansi

Recall	Precision	F-measure
0.0043	0.1249	0.0083

Akurasi yang dihasilkan dari data yang diuji sebanyak 28992 tag dengan 12333 tag yang relevan. Total pertanyaan yang dapat dihasilkan oleh sistem adalah sebanyak 137 pertanyaan dengan 27

diantaranya adalah pertanyaan yang relevan sesuai dengan teks masukan.

Tabel 8. Pengujian Akurasi

	Total	Relevan	Akurasi
Tag	28992	12333	**42,54%**
Pertanyaan	137	27	**19,71%**

4. PENUTUP

Berdasarkan hasil penelitian yang telah telah dilakukan, dapat diketahui bahwa pembangkit pertanyaan otomatis menggunakan NER lebih baik dibandingkan dengan pembangkit pertanyaan otomatis tanpa menggunakan NER, dengan menghasilkan akurasi 19,71% dari akurasi sebelumnya yaitu 14,53%.

Named Entity Recogniton sudah mampu membantu meningkatkan akurasi pada pembangkit pertanyaan otomatis. Saran untuk penelitian selanjutnya adalah menambahkan ambiguitas pada NER sehingga dapat mengurangi kata-kata yang ambigu. Kombinasi berbagai metode juga dapat dilakukan untuk mengetahui metode mana yang paling tepat untuk NER.

DAFTAR PUSTAKA

[1] S. Morwal, N. Jahan, & D. Chopra, "Named Entity Recognition using Hidden Markov Model (HMM)", *International Journal on Natural Language Computing (IJNLC)*, Vol. 1 No. 4, 2012.

[2] M. Raihan. I, "Pembangkit Pertanyaan Otomatis dari Teks Berbahasa Indonesia Berdasarkan Template Sintaksis", Bandung, 2016.

[3] S. K. Metin, T. Kisla. T, & B. Karaoglan, "Named Entity Recognition in Turkish Using Association Measures", *Advanced Computing: An International Journal (ACIJ)*, Vol. 3 No. 4, 2012.

[4] R. Alfred, L. Leong, On. C. K., & P. Anthony, "Malay Named Entity Recognition Based on Rule-Based Approach", *International Journal of Machine Learning and Computing*, Vol. 4 No. 3, 2014.

[5] Handi, "Text Pre-Processing pada Text To Speech Synthesis System untuk Penutur Berbahasa Indonesia", ITS Surabaya, 2011.

[6] I. Budi, S. Bressan, G. Wahyudi, Z. Hasibuan, & B. Nazief A, "Named Entity Recognition for the Indonesian Language: Combining Contextual, Morphological and Part-of-Speech Features into a Knowledge Engineering Approach", 2005.

[7] M. Fachri, "Pengenalan Entitas Bernama Pada Teks Bahasa Indonesia Menggunakan Hidden Markov Model", Yogyakarta, 2014.

[8] Y. Benajiba & P. Rosso, "Arabic Named Entity Recognition Using Conditional Random Fields".

[9] M. K. Khormuji, & M. Bazrafkan, "Persian Named Entity Recognition based with Local Filters", *International Journal of Computer Applications* (0975 8887), Vol. 100 No.4, 2014.

[10] C. N. Seon, Ko. Y., Kim. J. S, & J. Seo, "Named Entity Recognition using Machine Learning Methods and Pattern-Selection Rules".

[11] A. F. Wicaksono & A. Purwarianti, "HMM Based Part-of-Speech Tagger for Bahasa Indonesia", *Fourth Int. MALINDO Workshop*, 2010.