

Implementasi *Named Entity Recognition* Pada *Factoid Question Answering System* Untuk Cerita Rakyat Indonesia

Yulia Kurniawati¹, Indriati², Putra Pandu Adikara³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹yuliakurniawati@ub.ac.id, ²indriati.tif@ub.ac.id, ³adikara.putra@ub.ac.id

Abstrak

Legenda daerah atau cerita rakyat merupakan cerita yang berkembang diantara ragam budaya Indonesia yang telah turun-menurun diwariskan. Cerita rakyat merupakan sebuah cerita dari masa lampau yang dipercayai sebagai peristiwa yang benar terjadi, biasanya cerita rakyat tersebut berupa asal mula dari suatu tempat. Kurangnya sarana yang menarik dalam memperkenalkan cerita rakyat tersebut menjadi salah satu alasan kurang diminatinya cerita rakyat Indonesia. Selain itu tingkat pemahaman anak yang masih kurang dibanding orang dewasa menyebabkan mereka mudah melupakan cerita rakyat tersebut ketika mereka kurang mengerti terhadap cerita rakyat Indonesia tersebut. Penelitian ini bertujuan mempermudah anak-anak dalam memahami cerita tersebut. Oleh karena itu peneliti membuat *question answering system* dengan menggunakan metode *Named Entity Recognition*. Pengklasifikasian *named entity* pada penelitian ini menggunakan metode *naive bayes*. Pada penelitian ini digunakan empat *named entity* untuk mengenali kata yang selanjutnya akan dijadikan kandidat jawaban antara lain *product*, *person*, *location* dan *none*. Di mana *none* bukan merupakan entitas. Selain itu tipe pertanyaan yang dapat diajukan pada *question answering system* ini adalah *Factoid Question* yaitu berupa pertanyaan yang jawabannya berupa fakta yang singkat dan padat bukan berupa uraian. Data yang digunakan merupakan lima cerita rakyat Indonesia yang diperoleh dari internet dan hasil klasifikasi *Named Entity* memiliki nilai *precision* sebesar 34,22%, *accuracy* pengklasifikasian NE sebesar 64,65% dan *recall* sebesar 13,13% sedangkan untuk *question answering accuracy sistem* didapatkan akurasi sebesar 16,7%.

Kata kunci: *Question Answering System, Named Entity Recognition, Naive Bayes Classification*

Abstract

Regional legend or folklore is a story that developed among the various cultures of Indonesia that have been down and down inherited. Folklore is a story of the past that is believed to be the true event, usually the folktale is the origin of a place. The lack of an attractive means of introducing folklore is one of the reasons for the lack of interest of Indonesian folklore. In addition, the level of understanding of children who are still less than adults cause them to easily forget the story of the people when they are less understood to the story of the people of Indonesia. This study aims to facilitate the children in understanding the story. Therefore, the researcher makes question answering system by using Named Entity Recognition method. The classification of named entity in this research using naive bayes method. In this research used four named entity to recognize the next word will be candidate for answers such as product, person, location and none. Where none is an entity. In addition, the type of question that can be asked on the question answering system is Factoid Question is a question that the answer is a short and solid fact not a description. The data used are five folk stories of Indonesia obtained from the internet and the classification of Named Entity has a precision value of 34.22%, the accuracy of NE classification of 64.65% and recall 13.13% while for question answering accuracy system obtained accuracy of 16.7%.

Keywords: *Question Answering System, Named Entity Recognition, Naive Bayes Classification*

1. PENDAHULUAN

Cerita rakyat Indonesia merupakan cerita legenda daerah yang berkembang di Indonesia

yang dipelihara oleh warga dan tidak jarang dipercayai kebenarannya sebagai peristiwa yang nyata, namun tidak jarang pula yang menganggapnya hanya sebagai mitos dan dongeng belaka. Pada zaman sekarang ini semakin canggihnya teknologi membuat cerita rakyat tersebut terlupakan. Generasi muda zaman sekarang lebih memilih membaca novel modern maupun menonton film animasi luar negeri seperti Cinderella, Putri Salju dan sebagainya, sehingga tidak jarang generasi muda khususnya anak-anak lebih mengetahui cerita rakyat negara lain dibanding negara sendiri.

Kurangnya sarana yang menarik dalam memperkenalkan cerita rakyat tersebut menjadi salah satu alasan kurang diminatinya cerita rakyat Indonesia. Selain itu tingkat pemahaman anak yang masih kurang dibanding orang dewasa menyebabkan mereka mudah melupakan cerita rakyat tersebut ketika mereka kurang mengerti terhadap cerita tersebut.

Oleh karena itu dibutuhkan sebuah sistem untuk membantu generasi muda khususnya anak-anak untuk menemukan informasi yang relevan dengan informasi yang tengah dicari oleh anak-anak tersebut secara tepat dan efisien. Salah satu cara untuk mewujudkan hal tersebut adalah dengan membangun *Question Answering System* dengan obyek cerita rakyat Indonesia sehingga dapat memudahkan anak-anak tersebut untuk menemukan jawaban dari informasi yang ingin ia cari dengan tepat dan efisien, sehingga dapat meningkatkan pemahaman anak-anak terhadap cerita rakyat yang tengah dibaca.

Dari penelitian yang telah dilakukan sebelumnya dilakukan oleh Ivin Andhika Zulen dan Ayu Purwarianti (Zulen & Purwarianti, 2011). Pada penelitian tersebut dalam menentukan jawaban dari *factoid question* digunakan metode *named entity recognition* dengan menghitung jarak terdekat antara kata kunci pertanyaan dengan kandidat jawaban yang berupa kata yang memiliki *named entity* yang sesuai dengan *expected answer type* yang diperoleh dari pertanyaan masukan sistem. Dalam penelitian ini diperoleh MMR sebesar 0.6191 dari perhitungan jarak antara kandidat jawaban dengan kata kunci pertanyaan.

Dalam pencarian *named entity* pada penelitian ini digunakan metode *naive bayes* dikarenakan pada penelitian sebelumnya yang dilakukan oleh Mahalaksmi G.S. dalam

mengklasifikasikan *named entity* berbasis dokumen dengan *naive bayes* diperoleh akurasi sebesar 79% (G.S., et al., 2016), sehingga dapat dikatakan bahwa metode *naive bayes* cukup baik untuk digunakan sebagai metode klasifikasi *named entity*.

Pada penelitian yang dilakukan oleh Gautama dibandingkan perhitungan jarak antara metode *Manhattan* dengan *Euclidean*, hasilnya diperoleh kesimpulan bahwa perhitungan menggunakan metode *Manhattan* memiliki waktu eksekusi lebih singkat dibandingkan dengan *Euclidean* (Gautama, et al., 2015). Sehingga pada penelitian ini digunakan metode *Manhattan* sebagai metode perhitungan jarak antara kandidat jawaban dengan kata kunci pertanyaan yang ada.

Berdasarkan penelitian sebelumnya menunjukkan bahwa *named entity recognition* sesuai untuk digunakan sebagai metode dalam pencarian jawaban pada *factoid question answering*, sehingga dalam membangun *factoid question answering system* ini penulis menggunakan metode *named entity recognition* dalam menemukan jawaban yang sesuai dengan pertanyaan yang diberikan oleh pengguna.

2. DASAR TEORI

2.1 Pemrosesan Bahasa Alami (*Natural Language Processing*)

Bahasa alami adalah bahasa yang telah berevolusi secara alami dan digunakan oleh manusia untuk tujuan komunikasi, misalnya Hindi, Inggris, Prancis, Jerman merupakan contoh dari bahasa alami. Bahasa alami juga disebut sebagai Computational Linguistic merupakan studi ilmiah bahasa dari perspektif komputasi. Natural Language Processing adalah bidang komputer dan linguistik manusia yang berkaitan dengan interaksi antara komputer dan bahasa manusia (Kumar, 2010).

2.2 Temu Kembali Informasi (*Information Retrieval*)

Sistem temu kembali informasi merupakan teknologi untuk mencari informasi yang sesuai dengan apa yang diinginkan pengguna dari berbagai sumber dan menampilkan semua hasil

yang sesuai dengan yang diinginkan oleh pengguna (Mansouri, et al., 2008).

2.3 Question Answering System (QAS)

Question Answering System (QAS) atau sistem tanya jawab merupakan sistem yang dapat memberikan jawaban langsung kepada pengguna, berbeda dengan sistem yang berbasis *Information Retrieval* (IR) seperti mesin pencari yang memberikan hasil berupa daftar tautan yang *relevan*. Sistem tanya jawab memberikan jawaban yang benar dan tepat sehingga dibutuhkan usaha yang tidak mudah untuk membangun sistem tersebut. Secara umum, sistem tanya jawab terdiri dari tiga komponen utama yaitu *Question Analysis*, *Passage Retrieval* dan *Answer Extraction* (Hartati & Hartati, 2015).

Dalam proses untuk menemukan jawaban *question answering system* memiliki tiga tahapan proses yang harus dilalui yaitu *question analyzer*, *passage retriever* dan *answer finder* untuk lebih jelasnya akan dijelaskan pada subbab selanjutnya.

2.4 Factoid Question

Factoid question sering dianggap setara dengan pencarian informasi. Dalam *factoid question answering* diberikan *knowledgebase* dan sebuah *query* yang digunakan dalam menemukan jawaban dari *factoid question* tersebut (Iyyer, et al., 2014).

Factoid Question merupakan jenis pertanyaan yang jawabannya merupakan sebuah fakta yang diketahui dan terdapat didalam koleksi pustaka. Adapun jenis pertanyaan yang dapat digunakan dalam *factoid question* diantaranya siapa, di mana, ke mana, dari mana, kapan, apa dan berapa. Oleh karena itu *factoid question* dianggap sesuai digunakan dalam sistem tanya jawab ini. Salah satu metode yang dapat digunakan untuk menjawab pertanyaan bertipe *factoid question* adalah *named entity recognition* (NER).

2.5 Named Entity Recognition

Named Entity Recognition (NER) merupakan inti dari ekstraksi informasi dan melibatkan pemrosesan dokumen terstruktur serta tidak terstruktur. Pengindentifikasian eksperesi mengacu pada orang, tempat, organisasi, perusahaan, tanggal dan waktu. Dimana tujuan dari *Question Answering System*

tersebut adalah untuk mendapatkan dan mengembalikan jawaban yang benar dan spesifik untuk menjawab pertanyaan dari pengguna (Mansouri, et al., 2008)

2.6 Naive Bayes

Naive Bayes merupakan pengklasifikasi sederhana berdasarkan penerapan teorema Bayes dengan asumsi independen. Klasifikasi berdasarkan asumsi merupakan prinsip hipotesis posterior maksimal untuk mengidentifikasi benda yang paling mungkin tergolong dalam kategori. Teorema Bayes menunjukkan hubungan antara satu kondisional probabilitas dan kebalikannya (An, et al., 2017).

Dalam pengelompokan *Naive Bayes*, digunakan persamaan untuk menghitung probabilitas masing-masing kelas A yang diberikan nilai B_i dari semua atribut untuk contoh yang harus diklasifikasikan. Persamaan yang dapat digunakan dalam klasifikasi *naive bayes* dapat dilihat pada Persamaan 1.

$$P(A|B_1 \dots B_n) = P(A) \prod_i \frac{P(A|B_i)}{P(A)} \quad (1)$$

2.7 Manhattan Distance

Manhattan Distance/ City Block Distance, merupakan salah satu teknik yang sering digunakan untuk menentukan kesamaan antara dua buah obyek. Pengukuran ini dihasilkan berdasarkan penjumlahan jarak selisih antara dua buah obyek dan hasil yang didapatkan dari *Manhattan Distance* bernilai mutlak. Dimana *Manhattan Distance* melakukan perhitungan jarak dengan cara tegak lurus (Gautama, et al., 2015).

Manhattan Distance merupakan persamaan yang digunakan untuk menghitung jarak diantara dua titik dengan cara mengambil nilai absolut dari jarak kedua koordinat. Perhitungan jarak menggunakan *Manhattan Distance* dapat diperoleh dengan Persamaan 2 berikut:

$$Distance = |a - c| + |b - d| \quad (2)$$

Keterangan:

a: Koordinat titik a

b: Koordinat titik b

c: Koordinat titik c

2.8 Confusion Matrix

Confusion Matrix merupakan metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. *Confusion Matrix* mengandung informasi yang membandingkan

hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya. Untuk menghitung *Confusion Matrix* dapat digunakan Tabel 1.

Tabel 1. *Confusion Matrix*

Kelas	Terklasifikasi Positif	Terklasifikasi Negatif
Positif	True Positive	False Negative
Negatif	False Positive	True Negative

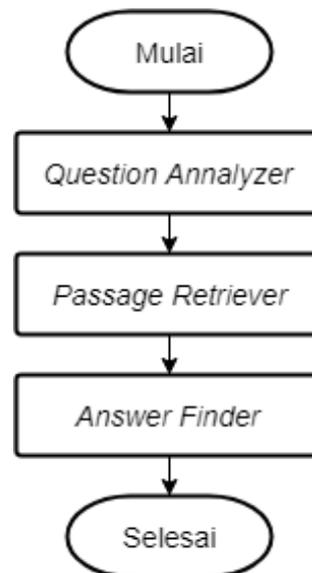
Sumber: (Leidiyana, 2013)

True positives adalah jumlah record positif yang diklasifikasikan sebagai positif, false positives adalah jumlah record negatif yang diklasifikasikan sebagai positif, false negatives adalah jumlah record positif yang diklasifikasikan sebagai negatif, true negatives adalah jumlah record negatif yang diklasifikasikan sebagai negative, kemudian masukkan data uji. Setelah data uji dimasukkan ke dalam confusion matrix, hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah sensitivity(recall), specificity, precision dan accuracy (Leidiyana, 2013).

3. PERANCANGAN DAN IMPLEMENTASI

Perancangan sistem ini dimulai dari pertanyaan yang diberikan oleh pengguna yang kemudian akan diproses dalam *question analyzer* untuk menentukan EAT yang terdapat dalam pertanyaan tersebut dan kemudian kandidat jawaban dicari melalui tahapan *passage retriever*, setelah ditemukan kandidat jawaban kemudian ditentukan jawaban yang tepat melalui tahapan *answer finder*.

Gambaran umum sistem ini dapat dilihat pada Gambar 1.

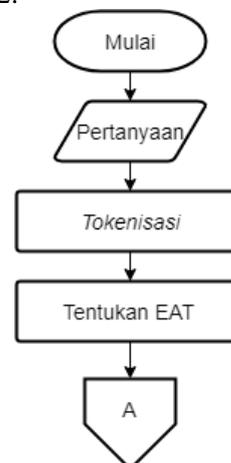


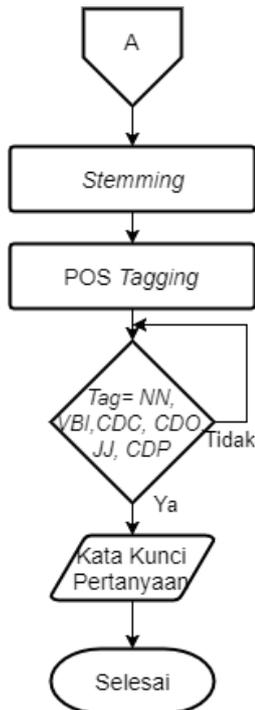
Gambar 1. Diagram Alir Sistem

3.1. Question Analyzer

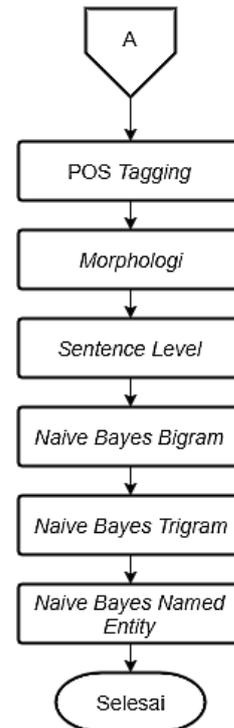
Question Analyzer merupakan tahapan dimana pertanyaan yang dimasukkan oleh pengguna di proses dengan menggunakan *Expected Answer Type (EAT)* dengan mempertimbangkan kata kunci pertanyaan dan kata kunci pendukung dari pertanyaan yang diberikan oleh pengguna (Zulen & Purwarianti, 2011).

Dalam sistem ini tahapan *question analyzer* dilakukan dengan melakukan tokenisasi terhadap pertanyaan yang diberikan oleh pengguna, kemudian dari pertanyaan tersebut ditentukan *expected answer typenya*. Dari pertanyaan tersebut kemudian dilakukan stemming untuk mendapatkan kata kunci pertanyaan yang diperlukan dalam sistem. Diagram alir *question analyzer* ditunjukkan pada Gambar 2.





Gambar 2. Diagram Alir *Question Analyzer*



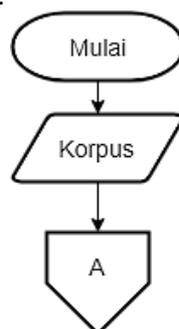
Gambar 3. Diagram Alir *Passage Retriever*

3.2. Passage Retriever

Pada tahap ini korpus yang ada diolah dengan beberapa tahap, yaitu *pre-processing*, *POS Tag*, *NE Tag*. Pemberian *POS Tag* yang digunakan untuk korpus diambil dari *IPOS Tagger* (Wicaksono, dan Purwarianti, 2010).

Dalam proses *passage retriever* pada sistem ini, pada awalnya dilakukan *POS tagging* terhadap korpus yang kemudian dilanjutkan dengan pengklasifikasian kata berdasarkan aturan *morphologi* dan *sentence level* dengan menggunakan metode *naive bayes*. Selanjutnya dari hasil *POS tagging*, *morphologi*, dan *sentence level* dilakukan klasifikasi penggabungan kata menjadi bigram dan trigram yang kemudian selanjutnya dilakukan klasifikasi *named entity* terhadap kata dengan menggunakan metode *naive bayes*.

Diagram alir *passage retriever* dapat dilihat pada Gambar 3.

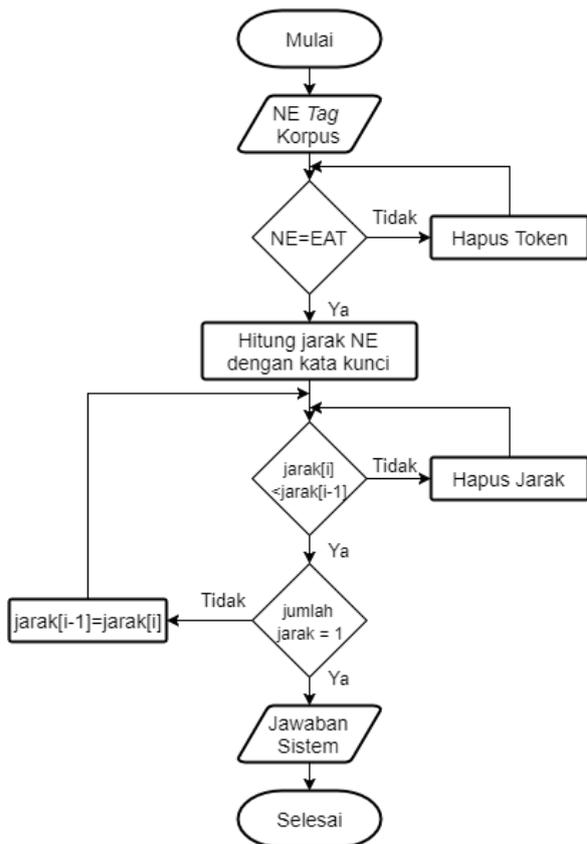


3.3 Answer Finder

Tahap ini merupakan tahap akhir dari *question answering system*. Dimana pada tahap ini semua kandidat jawaban yang telah terpilih sebelumnya akan diurutkan berdasarkan jarak antara kandidat jawaban dengan kata kunci dari pertanyaan yang ada, kandidat jawaban dengan jarak terkecil kemudian akan dipilih sebagai jawaban yang akan diberikan kepada *user*.

Pada sistem ini tahapan *answer finder* dimulai dengan pemilihan *named entity* yang sesuai dengan *expected answer type* yang diharapkan. Selanjutnya dilakukan perhitungan jarak antara kata kunci pertanyaan dengan kandidat jawaban. Kemudian jarak dari kandidat jawaban tersebut diurutkan dan diambil nilai jarak terkecil untuk dijadikan jawaban dari sistem.

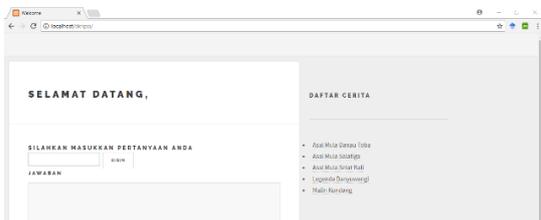
Diagram alir *answer finder* dapat dilihat pada Gambar 4.



Gambar 4. Diagram Answer Finder

3.4 Halaman Antar Muka Sistem

Halaman utama sistem berisi *textbox* yang digunakan oleh pengguna untuk memberikan pertanyaan kepada sistem dengan menekan tombol kirim. Selain itu di sebelah kanan halaman utama terdapat daftar cerita yang dijadikan koleksi pustaka dalam sistem dan merupakan batasan pertanyaan yang dapat diberikan kepada sistem. Implementasi halaman utama sistem dapat dilihat pada Gambar 5.



Gambar 5. Halaman Antar Muka

4. PENGUJIAN DAN ANALISIS

4.1 Pengujian *Confusion Matrix*

Confusion Matrix merupakan metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. *Confusion Matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem

dengan hasil klasifikasi yang seharusnya. Untuk menghitung *Confusion Matrix* dapat digunakan Tabel 2.

Tabel 2 Pengujian *Confusion Matrix*

	Person	Location	Product	Quantity	None
Person	74	22	13	0	216
Location	105	145	40	0	60
Product	27	30	91	0	4
Quantity	0	0	0	0	4
None	154	89	85	0	1160

Analisis yang dilakukan pada pengujian ini adalah untuk menghitung akurasi *named entity* yang tidak sesuai dengan kelas yang seharusnya, yang disebabkan oleh kemiripan antar data yang cukup besar terutama pada *named entity person* yang mana terdapat beberapa *named entity* yang seharusnya termasuk dalam *named entity* lain akan tetapi terklasifikasi sebagai *named entity person*.

4.2 Pengujian *Accuracy Named Entity Recognition*

Pengujian *accuracy* pada *named entity recognition* dilakukan untuk mengetahui evaluasi dari sebuah sistem dengan menghitung probabilitas pengkategorian *named entity* (sesuai atau tidak sesuai) dengan seluruh dokumen. Skenario pengujian *accuracy* pada *named entity recognition* ditunjukkan pada Tabel 3.

Tabel 3 Pengujian Akurasi

No	Kata	NE System	NE Pakar	Validitas
1	Sumatera	Location	Location	Valid
2	Hiduplah	Person	None	Tidak Valid
3	Seorang	Person	Person	Valid
4	Petani	Person	Person	Valid
5	Ikan	Product	Product	Valid

Dari pengujian akurasi NER yang telah dilakukan didapatkan akurasi sebesar 61.71% dengan jumlah *named entity* yang sesuai sebanyak 1447 dari 2345 kata. Analisis yang dilakukan pada pengujian ini akurasi *named entity recognition* dipengaruhi oleh metode dan fitur yang digunakan dalam klasifikasi *named entity* tersebut. Dikarenakan dalam sistem ini fitur yang digunakan hanyalah terbatas pada *morphologi*, *POS Tagging* dan *sentence level*, sehingga akurasi yang didapatkan tidak terlalu tinggi.

4.3 Pengujian Akurasi Jawaban Sistem

Pengujian akurasi jawaban ini digunakan untuk melihat seberapa besar tingkat akurasi sistem dalam menjawab pertanyaan yang diberikan oleh pengguna. Dalam skenario pengujian akurasi jawaban diberikan 30 pertanyaan kepada sistem. Contoh pengujian akurasi jawaban sistem yang dilakukan ditunjukkan pada Tabel 4.

Tabel 4 Pengujian akurasi jawaban sistem

Pengujian ke	Pertanyaan	Jawaban Sistem	Jawaban Pakar	Validitas
1	Siapa Yang Iba Mendengar Cerita Surati ?	Ageng	Raden Banterang	Tidak Valid
2	Siapa Yang Memberitahukan Sebuah Rahasia Kepada Raden Banterang ?	Kail	Pengemis	Tidak Valid
3	Apa Yang Dihunuskan Raden Banterang Ketika Mendekati Istrinya?	Pantangan	Keris	Tidak Valid
4	Siapa Yang Ingin Membunuh Raden Banterang?	Bernama	Kakak Surati	Tidak Valid

Analisis dari pengujian yang dilakukan, ketika jawaban yang diberikan oleh sistem tidak sesuai dengan semestinya *named entity* dari jawaban tersebut telah sesuai dengan *Expected Answer Type* yang ditentukan sistem, sehingga dapat disimpulkan bahwa hasil klasifikasi *named entity recognition* mempengaruhi akurasi dari jawaban *question answering system* dikarenakan jawaban yang seharusnya didapatkan tidak termasuk pada *named entity* yang seharusnya sehingga tidak masuk ke dalam kandidat jawaban.

4.4 Pengujian Precision Pada K Teratas

Pengujian *precision* pada *k* teratas atau dapat disebut dengan *Precision@K* merupakan pengujian yang bertujuan untuk menghitung persentasi dokumen teratas sejumlah *K*. Dimana dalam pengujian yang dilakukan ditentukan nilai *threshold* atau nilai *K* sebesar 10.

Precision pada peringkat ke *k* didapatkan dengan menentukan *k* peringkat sebagai *threshold*, kemudian menghitung persentase jawaban relevan sebanyak *k* peringkat dan mengabaikan jawaban dengan peringkat lebih dari *k* (Stanford, 2013). Contoh skenario pengujian *Precision@10* ditunjukkan pada Tabel 5.

Tabel 5 Skenario pengujian precision@10

No	Judul Cerita	Pertanyaan	Precision@10
1	Legenda Banyuwangi	Siapa yang Iba Mendengar Cerita Surati?	0,3
2		Siapa yang Memberitahukan Sebuah Rahasia Kepada Raden Banterang?	0
3		Apa yang Dihunuskan Raden Banterang Ketika Mendekati Istrinya?	0,1
4		Siapa yang Ingin Membunuh Raden Banterang?	0
5		Siapa yang Ditemui Surati?	0

Dari hasil pengujian yang dilakukan terhadap 5 cerita rakyat Indonesia dengan 30 pertanyaan didapatkan nilai *precision@10* dalam rentang antara 0 sampai 0,4. Hal yang mempengaruhi nilai *precision@K* dalam pengujian ini diantaranya adalah kesalahan pengklasifikasian *named entity* dan metode dalam pencarian jawaban yang hasilnya kurang baik karena hanya mempertimbangkan jarak antara kata kunci pertanyaan dengan kandidat jawaban.

5. KESIMPULAN

Berdasarkan hasil penelitian Implementasi *Named Entity Recognition* Pada *Factoid Question Answering System* Untuk Cerita Rakyat Indonesia, maka dapat diambil kesimpulan bahwa:

1. Penerapan *Named Entity Recognition* pada Cerita Rakyat Indonesia dapat di implementasikan dengan menggunakan algoritme klasifikasi *Naive Bayes*. Dalam pengklasifikasian *named entity recognition* dengan *naive bayes* didapatkan akurasi sebesar 61,71% hasil yang didapatkan kurang sesuai dikarenakan banyaknya kesalahan pengklasifikasian *named entity* yang disebabkan oleh kemiripan antar data dan parameter yang digunakan.
2. Hasil klasifikasi *Named Entity* memiliki nilai *precision* sebesar 34,22%, nilai *accuracy* *NE* sebesar

64,65% dan recall sebesar 13,13% sedangkan untuk question answering system pada penelitian ini memiliki akurasi sebesar 16,7%.

3. Question Answering System untuk Cerita Rakyat Indonesia dapat di Implementasikan dengan menggunakan algoritme perhitungan jarak Manhattan antara kata kunci pertanyaan dengan NER kandidat jawaban. Dari implementasi question answering system tersebut didapatkan akurasi sebesar 16,7%. Dalam penerapan question answering system dengan metode NER yang didapatkan kurang sesuai disebabkan banyaknya kesalahan pengklasifikasian NER yang telah dilakukan pada proses sebelumnya sehingga mempengaruhi tingkat akurasi dari question answering system.

Question Answering System ini masih memiliki beberapa kekurangan, saran yang diberikan untuk pengembangan sistem berikutnya adalah sebagai berikut:

1. Untuk klasifikasi Named Entity pada penelitian selanjutnya diharapkan dapat memperhatikan parameter-parameter yang lain selain yang digunakan oleh penulis untuk meningkatkan akurasi klasifikasi named entity tersebut.
2. Untuk tahap answer finder pada penelitian selanjutnya diharapkan dapat menambahkan parameter atau algoritme lain untuk meningkatkan akurasi sistem dikarenakan penggunaan metode Manhattan Distance tidak cukup baik untuk mendapatkan hasil yang baik.

6. DAFTAR PUSTAKA

- A. & Hartati, S., 2015. Analisis Ekstraksi Pengetahuan Eksternal Untuk Question Answering System. Seminar Nasional Teknologi Informasi (SNATi), pp. B-29 - B-3.
- An, Y., Sun, S. & Wang, S., 2017. Naive Bayes Classifiers for Music Emotion. *IEEE*, pp. 635-638.
- Gautama, A. W., Purwanto, Y. & Purboyo, T. W., 2015. Analisis Pengaruh

Penggunaan Manhattan Distance Pada Algoritma Clustering Isodata (SelfOrganizing Data Analysis Technique) Untuk Sistem Deteksi Anomali Trafik. *e-Proceeding of Engineering*, Volume 2, pp. 7404-7411.

G.S., M., J, B. A., Akshaya, K. & S, B. R., 2016. Domain Based Named Entity Recognition using Naive Bayes Classification. *Australian Journal of Basic*.

Iyyer, M. et al., 2014. A Neural Network for Factoid Question Answering over. *Empirical Methods in Natural Language Processing*.

Kumar, E., 2010. Natural Language Processing. New Delhi: I.K. International Publishing House Pvt. Ltd.

Leidiyana, H., 2013. Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko. *Jurnal Penelitian Ilmu Komputer, System Embedded & Logic*, pp. 65-76.

Mansouri, A., Affendey, L. S., Mamat, A. & Kadir, R. A., 2008. Semantically Factoid Question Answering Using Fuzzy SVM Named Entity Recognition. *IEEE*.

Stanford, 2013. *Stanford University*. [Online] Available at: <https://web.stanford.edu/class/cs276/handouts/EvaluationNew-handout-6-per.pdf> [Diakses 22 Januari 2017].