

Tinjauan Literatur Named Entity Recognition pada Resep Makanan Indonesia

by Febrianto Eko Saputro

Submission date: 09-Jun-2021 09:22PM (UTC+0700)

Submission ID: 1603483602

File name: ed_Entity_Recognition_pada_Resep_Makanan_Indonesia_17523165.docx (51.97K)

Word count: 1881

Character count: 11823

Tinjauan Literatur : *Named Entity Recognition* pada Resep Makanan Indonesia

Abstract— Dalam kehidupan kita sehari-hari makanan merupakan hal yang sangat penting. Masakan juga merupakan pencerminan budaya dan tradisi dari beberapa daerah dan hampir seluruh masakan kaya akan bumbu yang berasal dari rempah-rempah seperti: kemiri, cabai, kunyit, kelapa, kencur dan gula aren. Tidak semua masyarakat tahu tentang cara memasak suatu makanan ataupun mengenali resep suatu masakan. Hal tersebut terjadi karena mereka kekurangan pengetahuan untuk mengolah bahan masakan yang sudah tersedia di rumah. Metode *Named Entity Recognition* adalah salah satu sub tugas ekstraksi informasi yang berfokus pada mengenali dan mengidentifikasi unit informasi seperti nama orang dan nama makanan. Literatur ini mengkaji 8 referensi literatur mengenai NER pada domain masakan yang didapat melalui *Google Scholar* dengan kata kunci pencarian “*Named Entity Recognition for Food Ingredients*”. Tinjauan literatur NER terbanyak ditemukan di bidang kuliner dengan menggunakan model CNN (*Convolutional Neural Network*). Hasil dari penelitian ini diharapkan dapat membantu pengembangan NER pada bidang wisata.

Keywords— NER, kuliner, CNN

I. PENDAHULUAN

Dalam kehidupan kita sehari-hari makanan merupakan hal yang sangat penting. Masakan juga merupakan pencerminan budaya dan tradisi dari beberapa daerah dan hampir seluruh masakan kaya akan bumbu yang berasal dari rempah-rempah seperti : kemiri, cabai, kunyit, kelapa, kencur dan gula aren dengan diikuti penggunaan teknik-teknik masakan menurut bahan dan tradisi adat yang berasal dari resep tersebut[1].

Resep memasak secara tradisional telah dianggap sebagai bagian dari bahasa instruksional dan teknik khusus telah dikembangkan untuk memodelkan representasi mereka. Namun, teknik ini membutuhkan data berlabel. Bagian bahan sangat penting dan tidak boleh dimodelkan seperti teks instruksional lainnya untuk mencegah kehilangan data. Informasi bahan itu sendiri dapat memiliki kasus penggunaan seperti pemasangan makanan, prediksi rasa dan prediksi masakan [9].

Named Entity (Entitas bernama) adalah kata atau suatu ekspresi yang secara unik menggambarkan suatu elemen di antara sekumpulan elemen lain dengan atribut yang serupa. Organisasi, nama orang, nama tempat dan penyakit yang ada di bidang biomedis termasuk dari entitas bernama[2]. *Named Entity Recognition* adalah salah satu sub tugas ekstraksi informasi yang berfokus pada mengenali dan mengidentifikasi unit informasi seperti nama orang dan nama makanan. Kategorisasi entitas bernama dapat berbeda berdasarkan tujuan tugas NER. Sebagai contoh, diperlukan *preprocessing* seperti *Information Retrival*[3]. Dalam masalah ini, informasi seperti bahan, bumbu, dan alat untuk

memasak dapat membantu pengguna dalam mendapatkan informasi itu.

Oleh karena itu, tinjauan literatur ini bertujuan untuk melakukan perbandingan beberapa metode pada NER, agar mengetahui apa yang sudah digunakan pada metode NER penelitian sebelumnya dan melihat hasilnya. Tinjauan literatur ini dilakukan untuk mengetahui perkembangan penelitian di bidang NER, terutama NER yang diterapkan pada data kuliner dan resep. Selain itu, di penelitian ini juga akan mengungkap tentang metode apa saja yang digunakan untuk melakukan NER. Tinjauan literatur ini dapat mempunyai kontribusi serta menunjang peneliti untuk memastikan metode-metode yang baik untuk NER dan yang paling utama di bidang kuliner.

II. METODOLOGI PENELITIAN

Dalam proses pencarian literatur, ada beberapa tahapan yang dilakukan. Pertama, mencari literatur melalui *Google Scholar* dengan batasan tahun dari tahun 2017 hingga 2021. Kedua, kata kunci yang digunakan untuk mencari literatur adalah “*Named Entity Recognition for food ingredients*”. Pada Tabel 1, dapat dijelaskan untuk literatur yang telah di dapatkan menggunakan bahasa inggris dan metode yang digunakan dalam literatur tersebut menggunakan deep learning dan rule-based.

TABLE I. TABEL NAMA PENELITI DAN JUDUL PENELITIAN

Nama Peneliti	Judul
(Willis et al., 2017) [4]	Forage: Optimizing Food Use With Machine Learning Generated Recipes
(Goswami & Liu, 2017) [5]	Deep Dish: Deep learning for classifying food dishes
(Pan et al., 2017) [6]	DeepFood: Automatic Multi-Class Classification of Food Ingredients Using Deep Learning
(Popovski et al., 2019) [7]	Foodie: A rule-based named-entity recognition method for food information extraction
(Mahalakshmi et al., 2020) [8]	Exploiting Bi-LSTMs for Named Entity Recognition in Indian Culinary Science
(Diwan et al., 2020) [9]	A named entity based approach to model recipes

(Popovski et al., 2020) [10]	A Survey of Named-Entity Recognition Methods for Food Information Extraction
(Zhu & Dai, 2021) [11]	Food Ingredients Identification from Dish Images by Deep Learning

III. PEMBAHASAN

Bab ini membahas mengenai hasil kajian literatur serta ulasan terhadap literatur-literatur yang akan dikaji. Bersumber pada penelitian yang dilakukan, implementasi NER membutuhkan beberapa tahapan untuk menciptakan hasil yang diinginkan. Berikut adalah tahapan yang dilakukan pada proses NER.

1) Pengumpulan Data

Data yang akan diolah menggunakan *dataset* yang baru atau *dataset* yang sudah ada sebelumnya dan diberi label. Dapat dilihat rincian *dataset* yang digunakan pada setiap literatur pada Tabel 2 yang menunjukkan *dataset* apa saja yang digunakan pada penelitian sebelum-sebelumnya dan jumlah datanya.

TABLE II. TABEL DATASET DAN JUMLAH DATA

Judul	Dataset	Jumlah
Forage: Optimizing Food Use With Machine Learning Generated Recipes	Meal-Master[4]	60.000 recipes
Deep Dish: Deep learning for classifying food dishes	Google Image Search, Bing Image Search API[5]	26,984 images
DeepFood: Automatic Multi-Class Classification of Food Ingredients Using Deep Learning	Mealcome (MLC dataset) [6]	41 label dan 100 images
Foodie: A rule-based named-entity recognition method for food information extraction	Allrecipes[7]	1000 recipes
Exploiting Bi-LSTMs for Named Entity	Hebbers kitchen (https://hebbarskitchen.com/) [8]	3550 recipes

Recognition in Indian Culinary Science		
A named entity based approach to model recipes	RecipeDB[9]	118.000 recipes
A Survey of Named-Entity Recognition Methods for Food Information Extraction	FoodBase[10]	1000 recipes
Food Ingredients Identification from Dish Images by Deep Learning	Xiachufang.com[11]	Salient Ingredient Dataset : 6662 images Segmented-based ingredient dataset: 7803 images

2) Preprocessing

Tahapan *preprocessing* digunakan untuk mengolah data yang mentah menjadi data yang siap diolah atau siap digunakan pada proses selanjutnya. Proses *preprocessing* ini bermacam-macam dan proses *preprocessing* data tergantung pada kebutuhan penelitian. *Preprocessing* memiliki keuntungan seperti mengklasifikasi istilah bahan dengan benar yang berbeda dalam pluralitas, kapitalisasi, kehadiran tanda hubung sebagai entitas yang identitik. Misalnya, baik "tomat" dan "Tomat" adalah output sebagai "tomat", sehingga membuat data yang dihasilkan lebih berguna untuk tujuan yang berbeda [9].

3) Pelabelan Data

Setelah melakukan *preprocessing*, tahapan selanjutnya adalah pelabelan data. Pelabelan data sangat berguna karena pelabelan data sendiri adalah proses untuk mengidentifikasi data mentah dan menambahkan informasi yang berarti dan informatif untuk menyediakan konteks kepada data sehingga model bisa mempelajari data tersebut. Contoh pada penelitian [9] mereka menggunakan Stanford POS Twitter untuk melakukan *POS Tagging* ke semua data yang ada di dataset mereka.

4) Ekstraksi Fitur

Tahapan ekstraksi fitur sangat membantu dalam pengolahan data karena *deep learning* tidak bisa mengolah semua data terutama jika data tersebut berupa teks maka data mentahan diubah menjadi vector sehingga model dapat bekerja dengan lebih efektif. Penelitian [4] menggunakan *Word2Vec* untuk mengonversi token kata dalam data

pelatihan ke fitur yang direpresentasikan vektor untuk model LSTM. Ini sangat membantu untuk model LSTM untuk mempelajari arti kata dan menghasilkan hasil yang memuaskan. Di dalam model *Word2Vec*, mereka menggunakan pendekatan *Skip-gram* karena skip-gram bisa memprediksi sumber konteks kata-kata dari kata target.

5) Pemodelan Named Entity Recognition

Langkah selanjutnya adalah membangun model NER yang terdiri dari banyak metode di dalamnya. Model yang digunakan pada literature sebelumnya bisa lihat di Tabel 3. Model yang paling banyak digunakan adalah CNN. CNN singkatan dari *Convolutional Neural Network*. Model CNN adalah multilayer *artificial neural network* yang menggabungkan *unsupervised learning* dan klasifikasi yang dimana model CNN mengambil gambar mentah sebagai input dan menghasilkan skor klasifikasi akhir [6].

6) Evaluasi

Pada tahapan ini, proses evaluasi dilakukan kepada data pengujian untuk mengukur kinerja model NER yang sudah dibuat dan dapat diukur dengan nilai *Precision*, *Recall* dan *F1 Score* [10].

IV. ANALISIS

Dari tabel 3, dapat diketahui ada 9 model yang digunakan untuk NER. Penggunaan Model CNN (*Convolutional Neural Network*) adalah yang paling banyak digunakan. Beberapa ada yang mengkombinasikan metode lain seperti BiLSTM, LSTM, CRF dan RNN serta dapat ditambahkan fitur seperti *word embeddings* dan *Pos Tagging*. Model machine learning juga di gunakan seperti SVM. NER pada domain masakan yang menggunakan *deep learning* telah dilakukan oleh [4], [5], [6], [8] dan [11]. Penelitian [4] mengambil dataset dari Meal-Master. Penelitian ini menggunakan LSTM dan RNN untuk melakukan NER. Entitas yang ada di penelitian ini adalah *Ingredients*, *Steps* dan *Category*. Kemudian penelitian [5] menggunakan CNN dan SVM untuk melakukan NER. Mereka mengambil dataset dari *Google Search Image* dan *Bing Image Search API*. Pada penelitian [6] mereka menggunakan CNN yang dimana CNN tersebut terdiri dari tiga yaitu AlexNet, CaffeNet dan ResNet. Dataset mereka diambil dari Mealcome (MLC dataset). Penelitian [8] menggunakan BiLSTM dan CRF untuk menggunakan NER dengan menggunakan dataset dari website Hebbers Kitchen. Penelitian [11] menggunakan CNN untuk melakukan NER dengan data yang diambil dari website *xiachufang.com* dan mengenali entitas *Crops*, *livestock products* dan *Aquatic Products*.

Ada juga yang menggunakan rule based untuk NER. Penelitian pada domain masakan yang menggunakan rule-based telah dilakukan oleh [7], [9] dan [10]. Penelitian [7] menggunakan FoodIE untuk melakukan NER. Dataset mereka ambil dari Allrecipes. Kemudian [9] menggunakan Stanford NER dengan fitur *K-means Clustering*. Mereka menggunakan dataset dari RecipeDB. Penelitian [10] menggunakan dua metode yaitu FoodIE dan NOCB yang dimana NOCB ada tiga ontologi. Data yang mereka gunakan dapat dari FoodBase.

Temuan selanjutnya yaitu tentang hasil dari penelitian yang sudah ada. Dapat diketahui pada tabel 3, bahwa penelitian yang menggunakan *deep learning* rata-rata

mendapatkan hasil yang cukup tinggi. Dari 5 literatur yang menggunakan *deep learning*, penelitian [11] mendapatkan hasil yang tertinggi dengan akurasi sebesar 94.81% sedangkan penelitian [5] yang mendapatkan hasil terendah dengan hasil sebesar 0.40. Kemudian penelitian yang menggunakan rule based juga mendapatkan hasil yang cukup tinggi. Dari 3 literatur yang menggunakan rule based, penelitian [9] mendapatkan hasil tertinggi dengan hasil 0.9611. Sedangkan penelitian [10] yang mendapatkan hasil terendah dengan hasil 32.62%. Dari tabel 3, diketahui bahwa penelitian [9] mendapatkan hasil tertinggi dari semua literatur sedangkan literatur [5] mendapatkan hasil terendah.

TABLE III. TABEL MODEL DAN HASIL

Judul	Model	Hasil
Forage: Optimizing Food Use With Machine Learning Generated Recipes	LSTM RNN[4]	LSTM dengan RNN memperoleh hasil sebesar 0.885
Deep Dish: Deep learning for classifying food dishes	CNN SVM[5]	Mendapatkan hasil sebesar 0.40
DeepFood: Automatic Multi-Class Classification of Food Ingredients Using Deep Learning	CNN (AlexNet, CaffeNet, ResNet) [6]	AlexNet mendapatkan hasil average tertinggi dengan 86.78%
Foodie: A rule-based named-entity recognition method for food information extraction	FoodIE [7]	Mendapatkan hasil F1 score 0.9605
Exploiting Bi-LSTMs for Named Entity Recognition in Indian Culinary Science	BiLSTM CRF[8]	Mendapatkan hasil F1 score 94.66%
A named entity based approach to model recipes	Stanford NER <i>K-means Clustering</i> [9]	Mendapatkan hasil F1 score 0.9611
A Survey of Named-Entity	FoodIE	F1 score FoodIE : 96.05%

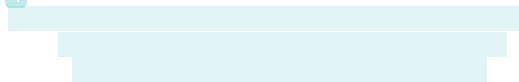
7 Recognition Methods for Food Information Extraction	NCBO(FoodOn, OntoFood, SNOMED CT) [10]	7 F1 score SNOMED CT : 63.75% F1 score OntoFood : 32.62% F1 score FoodOn : 63.90%
Food Ingredients Identification from Dish Images by Deep Learning	CNN (ResNet50) [11]	Salient Ingredients mean accuracy 85.96% Segmented-based accuracy 94.81%

masakan masaih tergolong sedikit dan model yang paling banyak digunakan adalah CNN (*Convolutional Neural Network*). CNN paling banyak digunakan karena CNN sangat efektif dalam mengklasifikasi dokumen karena dapat memilih fitur yang menonjol contoh tokens atau *sequence of tokens*. Penelitian [9] mendapatkan hasil tertinggi dengan F1 score 0.9611 dengan metode Stanford NER dan fitur *K-menas Clustering*. Penerapan NER pada domain kuliner harus dikembangkan lagi, pengembangan riset ini bisa menggunakan *deep learning* ataupun rule-based yang lebih bermacam-macam, perbanyak data sehingga dapat mengidentifikasi entitas yang lebih luas dan banyak.

V. KESIMPULAN

Penelitian ini mengkaji 8 literatur pada domain masakan. Yang didapatn melalui *Google Scholar* dalam rentang waktu tahun 2017 hingga 2021. Dari hasil kajian yang didapatkan, informasi bahwa saat ini penelitian NER pada domain

4



Tinjauan Literatur Named Entity Recognition pada Resep Makanan Indonesia

ORIGINALITY REPORT

12%

SIMILARITY INDEX

11%

INTERNET SOURCES

8%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

1	www.scribd.com Internet Source	3%
2	users.cs.fiu.edu Internet Source	2%
3	cs229.stanford.edu Internet Source	1%
4	www.ichve2020.org Internet Source	1%
5	www.scitepress.org Internet Source	1%
6	researchr.org Internet Source	1%
7	Gorjan Popovski, Barbara Korousic Seljak, Tome Eftimov. "A Survey of Named-Entity Recognition Methods for Food Information Extraction", IEEE Access, 2020 Publication	1%
8	scitepress.org Internet Source	

1 %

9

Aries Adiyanto, Dene Herwanto. "Tinjauan Kapasitas Persediaan Produk Fuji Seat PT. Tri Jaya Teknik Karawang", Jurnal Media Teknik dan Sistem Industri, 2021

Publication

<1 %

10

jurnal.una.ac.id

Internet Source

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography On