

# 半自動学習データ構築による固有表現認識の改善

澤山 熱気<sup>1,a)</sup> エリック・ニコルズ<sup>2</sup> 中野 幹生<sup>2</sup> 高橋 達二<sup>3</sup>

**概要:** レシピの推薦や、患者の病状に基づいた病気の詳細情報を提示する対話システムや情報検索システムの構築には、レシピ、医療などの特定ドメインを対象とした言語処理技術が必要である。そのような技術のベースとして、食事や疾病名など、対象とするドメインに特有の固有表現を高い精度で認識できる固有表現認識器が必要である。そのためには、正解ラベルの付いた大量の追加学習データが必要であるが、その構築には高いコストがかかる。本稿では、自動的に固有表現のタグ付けがされた対象ドメインの Web テキストから、人手のフィードバックを用いて少量の学習データを半自動的に構築し、追加学習データの構築のコストを削減する枠組みを提案する。提案手法では手動のタグ修正を用いることで、もともと NER モデルの精度が向上する可能性の高い文章を選択する幾つかの手法を提案し、提案手法がランダムに修正された手法とベースラインよりも良い性能につながったことを実証的に評価できたことを示した。さらに、信頼度の高い固有表現タグを含む文を自動的に選択した実験では、再現率の低さとトレードオフで高い適合率と F 値を得られることがわかった。

## Named Entity Recognition Improvement through Semi-supervised Data Construction

ATSUKI SAWAYAMA<sup>1,a)</sup> ERIC NICHOLS<sup>2</sup> MIKIO NAKANO<sup>2</sup> TATSUJI TAKAHASHI<sup>3</sup>

**Abstract:** Many tasks require natural language processing technology for specific domains, such as dialogue systems that make recipe recommendations or provide detailed information about illnesses based on user symptoms, and high-performance named entity recognition (NER) of cuisines and illnesses is the foundation of such technologies. However, constructing high-performance NER for new domains requires large quantities of labeled training data, which is expensive to construct. In this paper, we present a framework to reduce the cost of data construction by using small amounts of human feedback to semi-automatically extend a small core of hand-constructed training data with in-domain Web texts that have been automatically tagged with named entities. We propose several methods for selecting the sentences most likely to strengthen our NER model through manual tag correction, and empirical evaluation shows that our proposed methods lead to greater performance than a random selection baseline. Further experiments fully-automatically selecting sentences with high-confidence NE tags as additional training data show that models with higher precision and F1-scores can be trained for a small trade-off in recall.

### 1. はじめに

自分の好みに近い料理や、病状と関連する病気の情報を提示する対話システムや情報検索システムの構築には、レ

シピ、医療などの特定ドメインを対象とした言語処理技術が必要である。そのような技術のベースとして、料理名や疾病名など、対象とするドメインに固有の固有表現を高い精度で認識できる固有表現認識器が必要である。固有表現認識 (Named Entity Recognition, NER) の認識精度の向上に最も効果的なのが、正解ラベルの付いた追加の学習データを大量に用意することであるが、それには高いコストがかかる。

固有表現認識器の学習データの構築コストを削減する

<sup>1</sup> 奈良先端科学技術大学院大学

Nara Institute of Science and Technology

<sup>2</sup> ホンダ・リサーチ・インスティテュート・ジャパン

Honda Research Institute Japan

<sup>3</sup> 東京電機大学

Tokyo Denki University

a) sawayama.atsuki.ro2@is.naist.jp

方法として、文集合の中から、文中の単語に付与されたタグ情報を基に、より学習効率の高いデータを選択し、モデルの再学習に役立てる手法が提案されている。文書情報が Web 上に溢れるようになった現在では、Web 上の大量のデータから、いかに低コストで学習効率の高いデータを選択し、モデルの追加学習データを構築するかが課題となっている。

本稿では、固有表現認識の追加学習データ構築コストを削減するため、学習効率の良いデータの選択方法に着目した手法を提案する。本研究では、固有表現認識をおこなった際に付与されるタグの曖昧性に注目する。これは、ある単語に対して、文集合中で付けられた固有表現タグの種類が多いほど、その単語が対象ドメインで出現する回数が多く、かつ、固有表現認識の初期モデルがその単語へのタグ付けが苦手であることから、優先してタグ修正をおこなう必要があると考えられるためである。

提案手法では、予め文集合から対象ドメインと関連のある文を抽出し、少量の正解データから学習した固有表現認識器を用いて固有表現タグづけをおこなう。そして、タグ付けされた文を、タグの曖昧性から算出したスコアと人間からのフィードバックを用いてフィルタリングすることで、半自動で追加学習データを構築し、固有表現認識モデルの再学習をおこなう。

料理とレストランのドメインのデータを用いて提案手法を評価した結果、ベースラインの認識精度よりも精度が改善されるかを調査する。

## 2. 先行研究

固有表現認識とは、文中の重要語（固有表現）から、人名、地名、組織名、時間表現などのクラスの固有表現を認識する技術であり、情報抽出や言語理解や対話システムの基盤となる言語処理技術である。一般的に、固有表現認識では、文章中の各単語に適切なラベルを付与する系列ラベリング問題として解かれており、系列ラベリングの方法として、隠れマルコフモデル (Hidden Markov Model, HMM) [1]、条件付き確率場 (Conditional Random Fields, CRF) [2]、サポートベクターマシン (Support Vector Machine, SVM) [3]、ニューラルネットワーク (Neural Network) [4] [5] [6] などを用いた手法が提案されている。

従来の固有表現抽出は、MUC[7] で定義された、情報としての単位がはっきりとしている人名、組織名、地名、時間、日時、金額表現、割合表現の 7 種の表現を抽出するものであった。日本でも、1998-99 年に IREX [8] の一つのタスクとして採用され、固有表現認識のタスクの認識が広がった [9]。IREX [8] では MUC [7] の 7 種類に固有物名が加わり、8 種類の固有表現が設定されたが、情報抽出の広がり、質問応答という新たなアプリケーションの誕生で必要とする固有表現カテゴリの数は飛躍的に大きくなって

おり [10]、それらに対応するために 200 種類の固有表現クラス数に拡張した拡張固有表現が構築された [11]。

一方で、特定ドメインを対象とした固有表現認識においては、レストラン情報 [12] やレシピ [13] に関する研究が行われている。このように、特定ドメインでは必要とする情報単位が異なるために、別のクラスでの定義が必要であり、IREX [8] などの既存のクラス分類の正解データを用いることができない。そのため、特定ドメインの正解データを人手で構築する必要があり、大規模な文集合すべての正解データを人手で構築するにはコストが高いという問題がある。

現在、固有表現認識の学習データ構築コストを削減する方法として、大きく二つの方法が用いられている。一つは、学習方法を改良して判別性能を高める方法で、もう一つは固有表現認識器の学習データの量を増やし、効率のよい学習データを選択するという方法である。学習方法を改良する方法として、少量の正解データと大量の未知のデータを用いて学習効率を高める半教師あり学習と呼ばれる手法 [14] がある。また、学習データを増加させる方法として、膨大なデータの中から学習効率の高いデータを選択し、選択したデータのみ人手で正解を付与して学習データに追加する、能動学習を用いた手法 [15] [16] が提案されている。齋藤らの手法 [17] では、能動学習のデータ選択基準に、タグ単位の事後確率に基づいたタグ信頼度を導入し、システム出力の信頼度が閾値よりも低いタグのみを検出して人手修正をおこなう。

## 3. 提案手法

### 3.1 提案手法の概要

本研究では、学習データ構築コストを削減するため、齋藤らのタグ単位の信頼度 [17] とともに、タグの曖昧性を用いたフィルタリングを導入し、固有表現認識モデルの追加学習データを構築する。

提案手法では、あらかじめ文集合から対象ドメインの文を抽出して固有表現認識をおこない、その後、タグ信頼度に加え、タグの曖昧性を評価尺度の一つとして追加したフィルタリングをし、学習データの構築をおこなう。

提案手法は、大量の文集合から対象ドメインの文（これを関連文と呼ぶ）の抽出をおこなう『関連文抽出』と、集めた関連文から、固有表現を獲得し、半自動的に学習データを構築して、NER モデルの更新をおこなう『半自動学習データ構築と NER モデル更新』の大きく二つからなる (図 1)。以下に手順を示す。

### 3.2 関連文抽出

『関連文抽出』では、あらかじめ文集合から関連文の抽出をおこなう。これは、固有表現認識のモデル学習の際に、他のドメインの文（これを非関連文と呼ぶ）が含まれて

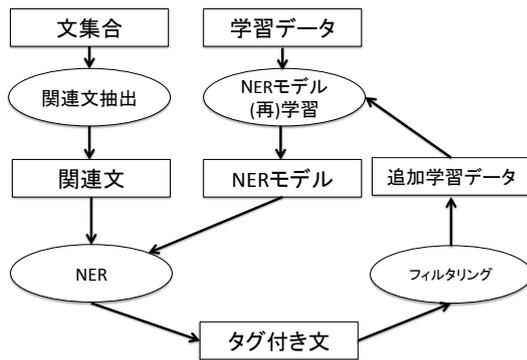


図 1 提案手法の概要

いと、固有表現認識の再学習と評価に悪影響を及ぼし、固有表現タグをつけるコストが余分にかかるためである。先行研究では、関連文抽出が行われておらず、非関連文を含む、全ての文に対して固有表現認識が行われ、タグの修正を行う際に非関連文までチェックする必要があったために、大きなコストがかかっていた。あらかじめ対象ドメインの文だけを抽出することができれば、追加学習データを構築する際に、学習の妨げになるデータを減らすことができ、より洗練した追加学習データの構築につながる可能性がある。

提案手法では、関連文、非関連文のラベルを手でつけた少量の文で教師あり学習でモデル学習をおこない、形式が正しくなかったり、一定の文字を超える文、句点で文が終わっていない文をあらかじめ取り除いたのち、ラベルの付いていない大量の文をモデルの学習をした分類器を用いてラベル付けをし、関連文を抽出する。

### 3.3 半自動学習データ構築と NER モデル更新

『NER モデル学習』では、正しい固有表現タグがつけられた少量の文を用いて、固有表現認識器のモデル学習をおこなう。

『NER』では、学習した NER モデルを用いて、『関連文抽出』で抽出した関連文に固有表現認識をおこない、固有表現タグをつける。

『フィルタリング』では、タグ付けされた文に対して、対象ドメインに関連性の薄いと思われる文を取り除く目的で、文中の単語に固有表現タグが全く付いていない(後述の Outside タグのみがつけられた)文を取り除いたのち、次節に述べる三つの手法のいずれかで文のフィルタリングをおこない、追加学習データを構築する。

『NER モデル再学習』では、初期のモデル用学習データに、『フィルタリング』で構築した追加学習データを加え、固有表現認識モデルの再学習をおこなう。

### 3.4 フィルタリング手法

本手法では、以下に述べる三つのフィルタリング手法のいずれかで、固有表現認識によってタグ付けされた文にフィルタリングをおこなう。また、今回用いる手法は、先行研究 [17] で用いられたタグの信頼度を利用する。

文信頼度下位修正手法では、文の信頼度をフィルタリングの基準とした。これは、固有表現認識された文中の単語に付与されるタグの信頼度を確認し、文中でもっとも低いタグの信頼度をその文の文信頼度として定義している。もっとも低いタグの信頼度が高い値であれば、文の修正の必要性が低いと考えられるため、文信頼度が低い順に文を選択し、修正対象文として出力する。また、リジェクトスコア手法では文の信頼度に加え、タグの曖昧性を用いたスコアを用いる。文中のある単語について、文集合全体に含まれるその単語につけられた固有表現タグの種類を確認し、固有表現クラスの総数で割った値をタグの曖昧性として定義した。これは、ある単語に付けられた固有表現タグの種類が多いほど、その単語が対象ドメインで出現する回数が多く、モデルがその単語へのタグ付けが容易でないことから、優先してタグ修正をおこなう必要があると考えられるためである。今回は、文中にあるタグの中で、最も低い信頼度をその文の信頼度とし、タグの曖昧性と文信頼度を組み合わせて、文のスコア(これをリジェクトスコアと定義する)を算出することで、フィルタリングする際の文の抽出基準とした。リジェクトスコアは以下のように定義した。

$$ambiguity(x) = \frac{\text{number of unique tag}(x)}{\text{total number of tag types}} \quad (1)$$

$$reject(x) = w_c(1 - confidence(x)) + w_a * ambiguity(x) \quad (2)$$

$$w_c + w_a = 1 \quad (3)$$

ある単語  $x$  について、 $confidence(x)$  は文中の単語につけられた最も信頼度がつけられたタグのある単語  $x$  の文信頼度、 $ambiguity(x)$  はタグの曖昧性、 $reject(x)$  はその文のリジェクトスコアを示している。 $w_c$ 、 $w_a$  はそれぞれ文信頼度とタグの曖昧性の重みとなっている。

このスコアが上位となる文を抽出し、人手で修正対象文中のタグ修正をし、追加データとする。

また、自動的に学習データを構築する手法として、文信頼度が高い順から文を一定数取り出して追加学習データとする手法(文信頼度上位自動追加手法とする)も算出した。

## 4. 実験

### 4.1 対象ドメインと固有表現クラス

本研究では、特定のドメインにおける固有表現認識を想定し、料理・レストランを対象ドメインとした。固有表現

表 1 対象ドメインと固有表現クラス

タグ	意味	例
SHOP	お店の名前	マクドナルド, 東武ストア
CUISINE	料理のジャンル	和食, イタリアン, 中華
FOOD	料理・食材名	焼肉, らーめん, トマト
MEAL	食事の種類	朝食, 昼食, おやつ
PLACE	場所	駅近, 横浜
TIME	時間	夏, 今週, 今日, 五時

クラスは、料理、レストランに関する対話や情報検索を想定し、対話システムを利用してレストラン検索をする際に重要な情報となる、レストランやお店の名前、料理のジャンル、料理の内容、レストランやお店の場所、食事の種類、時間の表現を認識できるよう、六つのクラス (SHOP, CUISINE, FOOD, PLACE, MEAL, TIME) を定義した。定義したタグとその意味を表 1 に示す。固有表現認識を行った際、これらの分類に含まれる単語に対しては、それぞれに相当するタグと、タグ付けに効果的な設定 [18] である、固有表現の先頭の単語に B (Begin)、その語に続く語は I (Inside)、固有表現の末尾の単語に E (End)、語が単体の固有表現であれば S (Single)、固有表現に該当しない語に対しては O (Outside) のタグ情報を付与する BIOES 方式に従って実験をおこなった。

#### 4.2 関連文抽出

まず、モデルの学習データとして関連文 (6,878 文)、関連文以外の文 (6,890 文) のセットを作成し、教師あり学習器 Vowpal Wabbit [19] を用いて二値分類のモデルの学習をおこなった。モデル学習後、実験用に Web 上から収集した文 (20,922,904 文) に対し、二値分類を行い、関連文抽出を行った結果、2,039,276 文が抽出された。分類対象の中から 990 文を選んで、関連文抽出の F 値を算出することで評価した結果、適合率 37.37 %、再現率 82.69 %、F 値 51.80 % となった。

#### 4.3 半自動学習データ構築と NER モデル更新

料理・レストランに関する Web から収集した文と対話文 2,177 文に固有表現タグを手手で付けた。そこから無作為に取り出した 1,000 文を評価データに用い、残りの 1,177 文を NER モデルの初期学習データとした。

初期学習データでモデル学習した固有表現認識器 DeepNNER [6] を用いて、関連文抽出で得られた文を定義した固有表現クラスでタグ付け (固有表現認識) したのうち、関連性の薄い文を取り除く目的で、文中の各単語に固有表現タグのつかなかった (O タグのみ付けられた) 文を除いた 117,832 文を用いた。その後、文信頼度手法では、文信頼度が低い順に、リジェクトスコア手法では、タグの曖昧性の値を大きくしすぎないように、文信頼度とタグの曖昧性の重みの値を 7 対 3 と 8 対 2 の二つの設定し、文の

表 2 各修正手法の F 値の変化

追加文数 (文)	文信頼度下位 (%)	標準偏差	リジェクトスコア $w_c=0.7,$ $w_a=0.3(\%)$	標準偏差
0(baseline)	77.1	0.90	77.1	0.90
100	78.7	0.21	79.0	0.25
200	78.9	1.14	78.6	0.64
300	79.6	0.59	78.8	1.47
400	79.6	0.34	78.7	0.23
500	79.8	0.86	78.9	0.42
600	79.7	0.32	<b>79.8</b>	0.56
700	79.3	0.45	79.0	0.54
800	79.2	0.62	79.4	0.38
900	<b>80.0</b>	0.41	79.0	0.82
1000	79.5	0.45	79.6	0.69

リジェクトスコアが高い順に、それぞれの手法で 100 文から 1,000 文までの 100 文きざみで修正対象文として取り出した。その後、修正対象文中のタグを手手で修正し、追加学習データを構築した。

また、文信頼度上位自動追加手法では、文信頼度が高い順に 100 文から 25,000 文まで一定数ずつ取り出して、追加学習データを構築した。これは、自動的に追加する際に、どの程度の量を追加学習データとすれば、モデルの改善につながるかを検証するためである。モデルの改善に役に立つ適切な量が発見できれば、少量の人手修正文だけでなく、大量の追加学習データを構築することができ、モデルを大幅に改善させることができる可能性がある。構築した追加学習データをモデルの初期学習データに追加して再学習をおこない、評価データを用いてモデルの評価をおこなった。

モデルの精度を比較するため、初期学習データで構築したモデル、文信頼度下位修正手法で構築した追加データを初期学習データに追加して再学習したモデル、リジェクトスコア手法で構築した追加学習データを初期学習データに追加して再学習したモデル、タグ付けされた文集合からランダムに 100 文から 1,000 文まで 100 文ずつを取り出して人手でタグ修正をし、初期学習データに追加した。これら、四つの手法で各手法ごとに五回ずつモデル学習をおこない、評価データに固有表現認識を際の F 値の平均値を算出した。実験の結果を以下に表 2 から表 4 に示す。

## 5. 考察

関連文抽出では、文集合から、不要な文として非関連文の大半を取り除くことができ、固有表現認識をおこなうために必要な関連文を抽出するコストを予め軽減することができた。しかしながら、関連文抽出の適合率が低かった。関連文抽出の改善方法として、教師あり学習器のモデルのパラメータの調整、モデルの学習に用いる非関連文を追加、モデルが判断しにくい、複数トピックを含むような文を学

表 3 各修正手法の F 値の変化

追加文数 (文)	リジェクトスコア $w_c=0.8,$ $w_a=0.2(\%)$	標準偏差	ランダム (%)	標準偏差
0(baseline)	77.1	0.90	77.1	0.90
100	78.7	0.62	77.3	0.45
200	78.5	0.52	77.3	0.46
300	78.9	0.57	76.9	0.87
400	<b>79.6</b>	0.57	77.4	0.52
500	79.4	0.76	77.3	0.90
600	79.6	0.53	77.4	0.57
700	78.9	0.41	77.6	0.51
800	79.4	0.42	77.8	0.29
900	79.5	0.22	<b>78.0</b>	0.89
1000	79.4	0.30	77.0	0.60

表 4 文信頼度上位自動追加手法の F 値の変化

追加文数 (文)	適合率 (%)	再現率 (%)	F 値 (%)	標準偏差
0(baseline)	75.6	78.6	77.1	0.90
100	78.6	77.5	78.0	0.47
200	78.4	77.6	78.0	0.41
500	79.0	77.6	78.3	0.78
1000	80.1	76.5	78.3	0.47
2500	79.7	76.8	78.2	0.69
5000	80.4	76.3	78.3	0.31
10000	80.8	75.2	77.9	0.22
25000	80.8	76.5	<b>78.6</b>	0.17

習データから予め除外するなどの処理をおこなうことが考えられる。

学習データの構築では、文信頼度下位修正手法、リジェクトスコア手法、文信頼度上位自動追加手法の全てが、ベースラインよりも F 値が向上し、ランダムに修正した手法よりも F 値が高くなった。そのため、提案した手法は、固有表現認識の精度の改善に効果があると考えられる。しかし、先行研究 [17] で用いられた指標である信頼度を用いた手法と、今回新たに提案した文信頼度とタグの曖昧性を使ったリジェクトスコア手法は同程度の精度となり、信頼度を用いる手法よりも精度が向上しなかった。

一方で、どの手法においても、必ずしも追加文数が多いモデルの認識精度が、追加文数が少ないモデルの認識精度よりも向上したわけではなかった。そのため、明らかに優れている手法や、データ量であるとは言にくいと考えられる。これは、IREX[8] のクラス分類に比べて、特定ドメインでは、分類するクラスの幅が繊細になり、分類が難しい固有表現が増えたことが原因なのではないかと考えられる。同じ単語でも複数の固有表現クラスが許される場合など、モデルの判断の難しいタグであれば、タグを修正しても分類の難しさは改善されず、大きな認識精度の改善に繋がらなかったのではないかと考えられる。

データを自動追加した信頼度上位自動追加手法では、ベー

スラインと比べて文を追加するごとに適合率と F 値が上昇傾向となったが、再現率は、文が増えるごとに下降傾向となった。

また、データを自動追加した信頼度上位自動追加手法に比べて、タグの信頼度が低いものを少量取り出し、タグの修正を文信頼度下位修正手法、リジェクトスコア手法のほうが F 値が高く、自動的手法よりも、半自動的手法でデータ構築をおこなったほうが、固有表現認識の改善に効果があると考えられる。

## 6. おわりに

本稿では、関連文抽出によって、あらかじめ文集合から関連文を抽出して固有表現認識をし、文信頼度下位修正手法、リジェクトスコア手法を用いた文のフィルタリングによって追加学習データを構築し、これによって、一般的な固有表現認識や特定ドメインの固有表現認識の学習データ構築に必要なコストを削減する手法を提案した。その結果、フィルタリングによって追加学習データを構築して再学習した固有表現認識器の精度が上昇傾向になることを確認できた。しかし、関連文抽出では、適合率が低く、関連文の中に非関連文がかなり含まれてしまったため、修正・抽出方法に改善の余地があったこと、文のフィルタリングを用いて追加データを構築し、モデルを再学習させたが、提案手法が従来の指標を用いた手法と同程度の精度を得るという結果となった。

今後の課題として、関連文抽出での抽出精度の向上による、質の高い関連文を抽出し、非関連文を用いないことで、コストを削減させること、フィルタリングにおいて、タグの信頼度やタグの曖昧性だけでなく、さらなる効果的なデータ選択基準を選定すること、ならびに対象ドメインにおける最適な固有表現クラスを定義することで、より低コストで十分な量の追加学習データを構築できるフィルタリング方法を検討すること、また、人手修正文から、固有表現認識された文を自動的に修正したり、修正された文のパターンを認識させることで、人工的に追加学習データを構築することで、より効果的な追加学習データを増やし、人手コストをさらに軽減できる方法を検討すること、さらに、追加学習データの構築と、固有表現認識モデルの再学習を交互に繰り返すことで、ブートストラップ的に固有表現認識モデルの認識精度を改善させることができるかを検証していく。

## 参考文献

- [1] Manning, Christopher D and Schütze, Hinrich. *Foundations of statistical natural language processing*, Vol. 999. MIT Press, 1999.
- [2] John Lafferty, Andrew McCallum, and Fernando CN Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. pp. 282–289,

- 2001.
- [3] 山田寛康, 工藤拓, 松本裕治. Support vector machine を用いた日本語固有表現抽出. 情報処理学会論文誌, Vol. 43, No. 1, pp. 44–53, 2002.
  - [4] Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. Natural language processing (almost) from scratch. *The Journal of Machine Learning Research*, Vol. 12, pp. 2493–2537, 2011.
  - [5] James Hammerton. Named entity recognition with long short-term memory. In *the seventh conference on Natural language learning at the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2003-Volume 4*, pp. 172–175. Association for Computational Linguistics, 2003.
  - [6] Jason PC Chiu and Eric Nichols. Named entity recognition with bidirectional LSTM-CNNs. *arXiv preprint arXiv:1511.08308*, 2015.
  - [7] Ralph Grishman and Beth Sundheim. Message understanding conference-6: A brief history. In *International Conference on Computational Linguistics*, Vol. 96, pp. 466–471, 1996.
  - [8] Sekine Satoshi and Isahara Hitoshi. IREX: Ir and ie evaluation project in japanese. In *the 2nd International Conference on Language Resources & Evaluation*, 2000.
  - [9] Sekine Satoshi and Isahara Hitoshi. IREX project overview. In *the IREX Workshop*, pp. 7–12. Citeseer, 1999.
  - [10] 関根聡. 固有表現から専門用語. 言語処理学会第 10 回年次大会「固有表現と専門用語」ワークショップ, 2004.
  - [11] Satoshi Sekine and Chikashi Nobata. Definition, dictionaries and tagger for extended named entity hierarchy. In *the International Conference on Language Resources and Evaluation*, pp. 1977–1980, 2004.
  - [12] 新里圭司, 関根聡, 吉永直樹, 鳥澤健太郎. 固有表現抽出手法を用いたレストラン属性情報の自動認識. 言語処理学会第 12 回年次大会発表論文集, pp. 93–96, 2006.
  - [13] 森信介, 山肩洋子, 笹田鉄郎, 前田浩邦. レシピテキストのためのフローグラフの定義. 情報処理学会研究報告. 自然言語処理, Vol. 2013, No. 13, pp. 1–7, 2013.
  - [14] 貞光九月, 齋藤邦子, 今村賢治, 松尾義博, 菊井玄一郎. トピック情報を用いたブートストラップ法に基づく語彙獲得. 自然言語処理, Vol. 19, No. 2, pp. 89–106, 2012.
  - [15] Dan Shen, Jie Zhang, Jian Su, Guodong Zhou, and Chew-Lim Tan. Multi-criteria-based active learning for named entity recognition. In *the 42nd Annual Meeting on Association for Computational Linguistics*, No. 589. Association for Computational Linguistics, 2004.
  - [16] Florian Laws and Hinrich Schätze. Stopping criteria for active learning of named entity recognition. In *the 22nd International Conference on Computational Linguistics-Volume 1*, pp. 465–472. Association for Computational Linguistics, 2008.
  - [17] 齋藤邦子, 今村賢治. タグ信頼度に基づく半自動自己更新型固有表現抽出. 自然言語処理, Vol. 17, No. 4, pp. 3–21, 2010.
  - [18] Lev Ratinov and Dan Roth. Design challenges and misconceptions in named entity recognition. In *the Thirteenth Conference on Computational Natural Language Learning*, pp. 147–155. Association for Computational Linguistics, 2009.
  - [19] Alekh Agarwal, Olivier Chapelle, Miroslav Dudík, and John Langford. A reliable effective terascale linear learning system. *The Journal of Machine Learning Research*, Vol. 15, No. 1, pp. 1111–1133, 2014.