自然言語処理 115－10
情報学基礎 43－2
（1996. 9. 12）

パターンマッチング手法による名称特定処理の有効性の検討

江里口 善生　　木谷 強
eriguchi@lit.rd.nttdata.jp　　tkitani@lit.rd.nttdata.jp

ＮＴＴデータ通信 情報科学研究所

あらまし

名称特定処理とは、固有名詞や日時などの表現を特定し、その種類を識別する情報抽出の基本的な処理である。名称特定の手法としてパターンマッチングが有効であることが知られているが、形態素解析の結果をパターンマッチングに使用するとき、形態素区切りと品詞がパターン開発者の期待するものと異る場合、従来は正しく名称を特定するために形態素解析ツールを修正しなければならなかった。DARPA 主催の情報抽出コンテスト MET に出展した名称特定処理システム Erie は、形態素の区切りと品詞を修正する機能をパターンマッチングエンジンに持たせ、形態素解析ツールを変更することなく、正しく名称を特定することに成功した。

和文キーワード　名称特定、情報抽出、パターンマッチング、形態素解析

Correct Name Recognition Using Pattern Matching

Yoshio Eriguchi　Tsuyoshi Kitani

Laboratory for Information Technology, NTT DATA

Abstract

Name recognition, which is essential in information extraction, is a task of identifying names and their types from a text. Pattern matching is known to be effective for name recognition from past research results. When word boundaries and parts of speech generated by a morphological analyzer do not match expectations of a pattern developer, the analyzer had to be changed to get desirable results. A name recognition system called Erie, developed for the Multi-lingual Entity Task (MET) sponsored by DARPA, can correct word boundaries and parts of speech by giving additional patterns to the pattern matching engine without changing the analyzer to meet the developer's requirements.

Key words　name recognition, informatoin extraction, pattern matching, morphological analysis

1 はじめに

1980年代後半からアメリカの政府機関DARPAが主催する情報抽出コンテストMUC(Message Understanding Conference)が開催されている[1], [2]。情報抽出とは、あらかじめ定められた項目に該当する情報をテキストから抽出し、抽出した情報間の関係を特定する処理である。大量の情報が氾濫している現代の情報化社会において、必要な情報を適切に取捨選択する有効な技術として注目を集めている。

1995年11月に開催された第6回大会(MUC-6)では、情報抽出タスクを固有名詞などの名称を特定する名称特定処理(Named Entity)、指示語等の照応関係を認定するCorefernece、基本項目の情報を抽出するTemplate Element、そして対象分野に依存する複雑な情報とその関係を抽出するScenario Templateの4つに分割し、各タスクごとに評価した[3]。その中で名称特定処理は、多くのシステムが再現率、適合率ともに90%を越え、最も精度が高いシステムは、再現率97%、適合率97%と人手による処理と遜色のない結果を得た。MUC-6では英語のみが対象であったが、名称特定処理の好成績を受け、日本語、中国語、スペイン語の3言語についての名称特定処理コンテストMET(Multilingual Entity Task)が1996年5月に開催された。

我々はパターンマッチング手法による名称特定システムErieを開発しMETの日本語部門に参加した。本稿では名称特定処理システムErieの処理の流れと使用するパターンの形式を紹介し、新聞記事に対する評価結果を報告する。

2 Multilingual Entity Task

METは英語を対象としていたMUC-6の名称特定タスクを、日本語、中国語、スペイン語に拡張した名称特定処理コンテストである。名称特定処理とは、テキストに出現する人名、地名、組織名、あるいは数量表現や時間表現などの名称について、その範囲と種別を識別する処理であり、情報抽出の精度を向上させるために必須な基礎技術である。

2.1 METの課題

METの課題は、新聞記事から固有名詞(組織名、地名、人名)、時間表現(日付、時間)、及び数量表現(金額、割合)の3カテゴリ(7サブカテゴリ)の名称を特定し、特定した名称にSGML形式のタグをつけることである。たとえば、

> マレーシアのハーバート大蔵省国庫局長は26日の記者会見で、貿易赤字拡大について言及した。

に対し、次のようなタグをつけることが要求されている。

> <location> マレーシア </location> の
> <person> ハーバート </person>
> <organization> 大蔵省 </organization>
> <organization> 国庫局 </organization>
> 長は <date>26日 </date> の記者会見で、
> 貿易赤字拡大について言及した。

タグ付けする対象や、カテゴリの定義については、METで定めた詳細なガイドラインにより規定されている。

2.2 従来の研究

MUC-5に出展したTEXTRACTの固有名詞特定処理[4],[5]や、松尾の新聞記事からの製品情報抽出システム[6]などで、パターンマッチングが名称特定処理に有効であることが報告されている。パターンマッチングによる名称特定処理は、名称の多くがパターン化できるという仮定に基づいている。"ウニタス社"のように最後が"社"で終わる単語は企業名(組織名)、"嘉数氏"のように"氏"と結び付く単語は人名であるという特徴をパターン化することができれば名称は特定できる。しかし、METの課題に取り組む場合、従来の手法では次のような問題が存在する。

TEXTRACTの固有名詞特定処理システムは、形態素解析により文を形態素に分割し、形態素を単位としたパターンマッチング処理を行っている。そのため、形態素解析ツールにより分割された形態素を、更に細かく分割する必要のある次のような名称を特定することが困難であった。

- "来日" $\Longrightarrow$ "来" と "日(地名)"
- "国防相" $\Longrightarrow$ "国防(組織名)" と "相"

一方、松尾のシステムでは形態素解析をせず、文字を単位としたパターンマッチング処理を行う。そのため、上記の分割に関する問題は生じないが、形態素の品詞とその属性を使用することができない。

MET特有の形態素分割にも対応し、かつ、形態素の情報を使ったパターンマッチングを実現するためには、MET用の形態素解析ツールを開発する方法と汎用の形態素解析ツールの結果を再加工する方法がある。

前者は、既存の形態素解析ツールの辞書を変更することで解決できる場合もあるが、一般的には接続文法、複数の分割候補に対する処理、未知語処理などの変更が必要となり開発に時間がかかる。METの参加チームではSRAやNEC & シェフィールド大学が既存の形態素解析ツールをMET用にチューニングしていた[7]。

後者は、形態素解析ツールには手を入れず、形態素解析結果を再加工する解決方法である。処理効率の面では前者の手法に劣るが、開発が容易であり、タスクによる

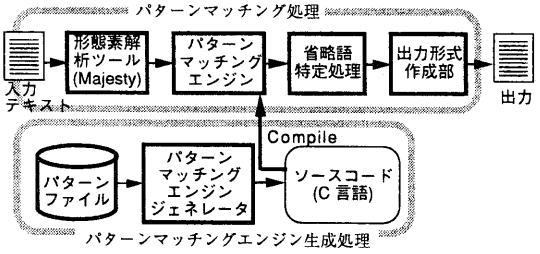

図 1: Erie のシステム構成図

名称の特定条件の違いに柔軟な対応ができるという利点がある。本稿で提案するシステムは、後者の手法を採用している。

次章では、形態素の再分割、新たな品詞付与などの機能をもつパターンマッチングエンジンについて説明する。

3 名称特定処理システム Erie

3.1 Erie の特徴

MET 用に開発した名称特定処理システム Erie は以下の特徴を有する。

- パターンマッチングエンジンは、形態素解析ツールにより分割された形態素を入力単位とし、形態素の品詞情報をパターン作成に利用する。
- 入力された形態素に対し、形態素区切りの変更と、新たな品詞を付与するパターンが記述できる。
- パターンファイルを読み込み、パターンマッチングエンジンを生成するパターンマッチングエンジンジェネレータを持つ。

Erie のモジュール構成を図 1に示す。Erie の処理は、パターンマッチングエンジンを生成するバッチ処理と、生成されたエンジンが名称を特定するパターンマッチング処理に分かれている。

パターンマッチングエンジンの生成処理では、ユーザが定義したパターンをパターンマッチングエンジンジェネレータが構文解析し、構文にごとに定められた変換規則に従って、パターンマッチングエンジンのソースコードを C 言語で生成する。生成されたソースコードをコンパイルすることで、パターンマッチングエンジンの実行形式となる。

次に、パターンマッチング処理を説明する。パターンマッチング処理では、まず、入力文を形態素解析ツールにより解析する。形態素解析には Majesty[1] を使用した。形態素解析の結果をシステムが読み込みパターンマッチング処理を行う。パターンマッチング処理は、入力された形態素列と定義されたパターンを順番に比較する原始的なアルゴリズムである。あるパターンに適合した形態素、あるいは形態素列に対し、パターンのタイプごとに定められたさまざまな処理が施される。パターンのタイプと対応する処理については次節で説明する。

パターンマッチング処理のあと、省略語特定処理によりパターンマッチング処理では特定できない組織名の省略語を特定する。省略語特定処理は、既に組織名として特定された名称と未知語を比較することにより、組織名の省略語を見つけ出す処理である。

最後に、カテゴリ (サブカテゴリ) が付与された形態素列に対し、MET で定められた SGML タグを付与したテキストデータに変換し出力する。

3.2 パターンの仕様

Erie で記述できるパターンには、品詞定義パターンと形態素分割パターンおよびカテゴリ付与パターンの 3 タイプがある。品詞定義パターンと形態素分割パターンは、形態素解析の結果を変更するために使用する。カテゴリ付与パターンは、名称を特定するために使用する。各タイプのパターン形式とその機能を説明する。

3.2.1 品詞定義パターン

品詞定義パターンは、

$$\text{品詞名} = \{Word_1, Word_2, ..., Word_n\};$$

という形式で表現する。このパターンは、右辺の形態素列の集合に対し、左辺の品詞名を与える。左辺の品詞名は、形態素解析ツールが定義する品詞名の他に独自にユーザが定義でき、アルファベットと数字からなる記号を使用する。また、品詞名を表す記号列の間にハイフンを入れ、ハイフンの左を品詞名、ハイフンの右を属性値とすることができる。属性値は品詞を更に細かく分類をしたい場合に定義する。

右辺の集合の $Word$ は、"会社" のように対象となる形態素の表記を列挙する。たとえば、

```
SUFFIX-COMPANY = {
  社 公社 研究所 銀行
}
```

は品詞定義パターンの一例で、形態素 "社"、"公社"、"研究所"、"銀行" に、品詞 SUFFIX-COMPANY(組織名につく接尾語) を新たに付与するためのパターンである。

[1] 辞書には約 93,000 語登録されており、そのうち固有名詞が 20,500 語 (組織名が約 4,000 語、人名が約 8,100 語、地名が約 8,400 語) 登録されている。

3.2.2 形態素分割パターン

形態素分割パターンは、次の形式で記述する。

$Word = Word_1$: 品詞名$_1$ $Word_2$: 品詞名$_2$
{ 分割条件 1; ...; 分割条件 n;}

形態素分割パターンは、左辺の形態素 $Word$ を右辺の形態素列 $Word1$、$Word_2$ に分割し、分割された形態素に品詞$_1$、品詞$_2$ を付与する。中括弧内に記述された条件式は、分割を行うための条件で、すべての条件を満たすときのみこのパターンは適用される。

左辺の形態素は、形態素そのものの表記を記述する他に、"_ 長官" のように、記号 "_ " を文字列の前後に組み合わせることができる。"_ " はワイルドカード的な意味をもち、"_ 長官" には、"官房長官" や "国務長官" など、最後が "長官" で終わる形態素すべてに適合する。

条件式は 0 個以上の任意の数が記述できる。条件式が 0 のときは、左辺の Word に適合する形態素に対して常に形態素分割パターンが適用される。条件式には、分割対象の形態素あるいはその前後の形態素の種類や、分割対象形態素の文字数、あるいは、例外の文字列などを指定できる。たとえば、

```
日付 = 日:SUFFIX-DATE 付:SUFFIX
{
  POS:PRE = NUM;
}
```

は形態素分割パターンの例である。このパターンは "日付" という形態素を "日" と "付" に分割し、それぞれに SUFFIX-DATE と SUFFIX という品詞を与えるパターンである。しかし、条件式により、直前 (PRE) の形態素の品詞 (POS) が数詞 (NUM) であることが定義されているため、数詞に続く "日付" に限り分割が実行される。この条件により「"3"、"日付"、"の"」という形態素の並びのときには分割し、「"今日"、"の"、"日付"」という形態素の並びのときに分割しない。

3.2.3 カテゴリ付与パターン

カテゴリ付与パターンは、形態素をグループ化し名称のカテゴリ名を付与するために使用する。一般形は次のような形式である。

$category$: $< pattern >$;

左辺の $category$ はカテゴリ名で、アルファベットと数字から構成される記号である。右辺の $pattern$ は正規表現に似たパターン表記である。本稿ではこのパターンを、形態素パターンと呼ぶ。形態素パターンの種類については本稿最後の付録に示している。

カテゴリ付与パターンは、$pattern$ に適合する形態素列をグループ化し、記号 $category$ をカテゴリ名として付与する。グループ化とは、一つまたは複数の形態素をまとまりとしてとらえるものである。各グループにはユニークなグループ ID が付与される。

カテゴリ付与パターンでは、カテゴリを付与する形態素列の前後の形態素列に対するパターンを以下のように記述できる。

$category : pattern_1 < pattern_2 > pattern_3$;

ここで、$pattern_1$、$pattern_2$、$pattern_3$ すべての形態素パターンを満たす形態素列が存在した場合、"<" と ">" に囲まれた形態素パターン $pattern_2$ に対応する形態素列に対し、グループ化とカテゴリ名の付与が行われる。

形態素パターンは正規表現を模倣して作成した。大きく異なる点は、記号':' を用いて AND 条件を記述できる点である。通常の正規表現は文字との照合を目的としているが、カテゴリ付与パターンでは文字の他に、品詞、カテゴリ名、文字種などとも照合できるようにしているため、複数の照合条件が組み合わされることがある。たとえば、N:KATAKANA は、品詞が名詞 (N) でかつ文字種がカタカナ (KATAKANA) である形態素に適合する形態素パターンである。また、

```
PERSON:
  < [N-PERSON UNKNOWN]+ > SUFFIX-PERSON;
```

は、人名を特定するカテゴリ付与パターンの例である。属性が人名である固有名詞 (N-PERSON)、あるいは未知語 (UNKNOWN) である形態素の一つ以上の繰り返しのあと、人名につく接尾語 (SUFFIX-PERSON) が続く場合、N-PERSON あるいは UNKNOWN の部分をグループ化し、カテゴリ PERSON を与える。このパターンは形態素列

鈴木	善行	氏
N-PERSON	UNKNOWN	SUFFIX-PERSON

に対して適合し、"鈴木" と "善行" をグループ化し、カテゴリ名 PERSON を与える。

4 実験データと評価項目

4.1 実験データ

実験に使用したデータは、すべて DARPA より MET 参加者に提供された新聞記事である。新聞記事は開発用に 300 記事が提供され、そのうち 100 記事には、名称に対するタグが付与されている正解データも併せて提供された。新聞記事には、タイトル、日付、本文などを区別するための SGML 形式のタグが付与されている。

開発用データ パターンの開発は、作成したパターンを使い名称特定処理を実行し、処理結果と正解データを

比較して失敗の原因を分析したのち、パターンを修正するという一連のサイクルを繰り返す。開発サイクルに使用するトレーニングデータには、開発用に提供された正解データ付きの100記事のうち60記事を使用した。60記事には、固有名詞が897、日時の表現が351、数量表現が61が含まれている。正解データがない200記事からはパターンマッチングに必要なキーワードと固有名詞を抽出し、抽出した固有名詞は形態素解析用の辞書に追加した。

評価用データ 本実験では、METコンテストで使用された99記事の新聞記事に対して評価を行った。99記事中に含まれる特定対象名称の数は2389で、そのうち固有名詞が1698、日時の表現が630、数量表現が61である。未知のデータに対する名称特定処理の効果を測定するため、評価データの記事内容は一度も見ていない。

4.2 パターンセット

実験で使用したパターンセットは、品詞パターン数が58、新たに品詞を付与した形態素の数が723(そのうち国名が204、通貨単位が94)、形態素分割パターン数が112、カテゴリ付与パターン数が188で構成されている。

4.3 評価項目

名称特定処理の評価には、METが用意したスコアリングプログラムを使用した。評価プログラムは正解データとシステムの出力を比較し、自動的に処理精度を計算する。スコアリングプログラムの評価項目は、名称を分割する位置の正しさを評価する分割精度と、名称に付与したカテゴリ(サブカテゴリ)の正しさを評価するカテゴリの付与精度、それと二つの項目の平均値である総合精度がある。

正解データにはDARPAから提供された正解タグつきデータを使用した。スコアリングプログラムでは、分割精度とカテゴリ付与精度および総合精度が出力されるが、本稿では総合精度のみ扱う。総合精度にもいくつか評価指標が用意されているが、そのうち適合率と再現率を取り上げる。

特定した名称の個数をACT、正しく分割できた名称の個数を$COR1$、正しいカテゴリの付与ができた名称の個数を$COR2$、正解データに含まれる名称の数をPOSとしたときの、総合精度における再現率(REC)、適合率(PRE)の計算方法を以下に示す。

$$REC = \frac{COR1 + COR2}{2 \times POS}$$

$$PRE = \frac{COR1 + COR2}{2 \times ACT}$$

表1: METの評価結果

システム名	再現率(%)	適合率(%)
System A	90	94
System B	80	89
System C	76	87
System D	79	84
System E	68	82
System F	65	82
System G	63	81

表2: カテゴリ別の評価結果

カテゴリ名	再現率(%)	適合率(%)
固有名詞	76	87
日時	96	97
数量表現	95	100
総合	82	90

5 評価結果と考察

5.1 METの結果

表1は、METの日本語部門の評価結果である。日本語部門には、NEC & シェフィールド大学、米国SRA、SRI、MITRE、BBN、NTTデータ通信の6チームが参加し、そのうち1チームが2システムを出展したため、合計7システムの参加となった[2]。

出展されたシステムは、すべて形態素解析をした後にパターンマッチングを行うという点では共通していたが、パターンを機械学習により作成するか、人手により作成するかの違いがあった。したがって、処理精度の差はパターン作成に使用した記事の量、システムが記述できるパターン種類、パターン作成者(あるいは機械)のパターン作成能力、形態素解析の精度などにより生じていると考えられる。

5.2 名称特定処理の効果

METに参加した時点では46記事のトレーニングデータを使用したが、その後14記事を加え60記事のトレーニングデータよりパターンを更新した。このパターンセットを使用したErieの名称特定処理の結果を表2に示す。

Majestyは固有名詞として、企業名、地名、姓、名の4種の属性を付与する。Majestyが与えた品詞だけで名称を特定した場合の固有名詞の特定精度は、再現率62%、適合率72%であった。これに対し名称特定処理を用いた場合、表2に示すように再現率76%、適合率87%と精

[2]METではシステム別の評価結果は公開するが、システムと開発チームの関係は非公開情報とする取り決めがあったため、表1にチーム名は掲載しない。

表 3: 形態素分割パターンを使わないときの評価結果

カテゴリ名	再現率 (%)	適合率 (%)
固有名詞	67	85
日時	96	97
数量表現	95	100
総合	74	89

表 4: トレーニングデータに対する評価結果

カテゴリ名	再現率 (%)	適合率 (%)
固有名詞	96	96
日時	100	99
数量表現	100	100
総合	98	97

度が15% 程度も向上し、パターンマッチング手法による名称特定処理は効果が確認できた。Majesty には日時と数量表現に対応する品詞は存在しないので、名称特定処理と比較することはできなかった。

カテゴリ別に特定精度を比較すると、日時、数量表現に対しては高精度の特定を実現できているが、固有名詞に対しては再現率が 76%、適合率が 87% である。日時の表現と数量表現はその多くが一定のパターンで表現されるのに対し、固有名詞はその表現が多様であり、パターン化しにくいことが原因である。また、固有名詞は "辜振甫台湾セメント董事長" のような、人名と企業名が連続して並ぶ場合も多いが、連続した固有名詞の片方、あるいは両方が未知語である場合は、その区切りをパターンで見分けることは難しい。

5.3 形態素分割パターンの効果

本手法の特徴である、形態素解析の結果を分割するパターンの効果について調べた。表 3は、形態素分割パターンを使用しない場合のカテゴリ別の評価結果である。形態素分割パターンを使用した場合の評価結果である表 2と比較すると、特定精度は総合で再現率は 8%、適合率は 1% 向上している。

この結果は、汎用形態素解析ツール Majesty の出力には、さらに細かく分割する必要がある形態素が多数存在し、そのような形態素に対して形態素分割パターンが有効であったことを示している。

5.4 パターンマッチング処理の限界

パターンマッチングによる名称特定処理は、パターン化できる名称についてはその効果を発揮できるが、パターン化できない名称に対しては効果がない。本実験では、トレーニングデータに対して正解に近付くようにパターンを作成している。そこで、文書中に存在するパターン化できない名称の割合を知るために、トレーニングデータに対する実行結果を分析する。トレーニングデータに対するパターンマッチング処理の結果を表 4に示す。

表 4の再現率から、固有名詞で 4% 程度パターン化できなかった名称が存在することがわかる。また適合率からは、固有名詞で 4% 程度、日時についても 1% 程度、誤った名称特定が実施されていることがわかる。本実験で特定できなかった名称と、名称特定を誤った例をいくつか示す。

- "環境保護団体グリンピース" に対し "グリンピース" をパターンがないため組織名として特定できない。"グリンピース" は一般名詞であり、単純に「"〜団体" + 一般名詞」というパターンでは、組織名でない表現とマッチすることが多いため、パターンとして定義できないからである。

- "三峡プロジェクト" に対し "三峡" を地名として特定できない。"三峡" は未知語であり、適切なパターンを作成できないためである。

- "吉井市長" に対し "吉井" を地名と誤認識する。Majesty は "吉井" を、地名と人名両方の属性で出力する。現状では、市長の前に地名が来た場合は他の属性を無視し地名と判断するパターンとなっているためである。

- "受け入れ機関" を固有名詞 (組織名) と誤認識する。機関名を特定するための「名詞 + "機関"」を組織名とみなすパターンが存在するためである。

- "「明日の内閣」" に対し "明日" を日付と認識する。明日という単語は無条件に日付としているためである。

5.5 処理時間

処理時間は記事の文字数に比例することがわかっている。実験に使用した新聞記事の一記事あたりの平均文字数は約 400 文字であった。また、処理時間はパターンセットを構成するパターンの数にも依存する。名称特定処理の処理時間は、Sun Sparc Station 10 上で、1 記事あたり形態素解析に 0.6 秒、パターンマッチング処理に 0.9 秒要し、合計 1.5 秒要した。実行時にパターンファイルをシステムが読み込むのではなく、実行前にパターンファイルからC言語のプログラムを作成し、処理を高速化している。60 記事の解析を約 1 分半で終了し、開発で使用するには充分に高速であった。

6 おわりに

形態素を入力単位とし、パターンマッチング手法により名称を特定する場合、形態素解析ツールで分割された形態素を、より細かく分割しなければ名称が特定できないという問題があった。そこで、パターンマッチングエンジンの機能を拡張し、形態素を更に分割するパターンを記述できるようにした。この結果、形態素解析ツールを名称特定処理用に改造しなくても適切に形態素を分割できるようになり、本機能を使わないときに再現率74%、適合率89%であった名称の特定精度を、再現率82%、適合率90%に向上させることができた。

固有名詞に対する名称の特定精度は再現率76%、適合率87%であるが、形態素解析単独で名称を特定した場合は再現率62%、適合率72%であり、名称特定処理の効果を確認できた。また、固有名詞以外の特定対象名称である日時の表現や数量表現は、再現率、適合率ともに95%を越える精度で特定することができた。

本実験ではパターン開発時に使用した記事が60記事と少なかった。今後の実験で、開発用の記事を増やした場合に、パターンの精度がどれくらい向上するか検討する必要がある。また、本実験ではパターンをすべて人手で作成したが、パターンの開発を、機械学習により自動化あるいは半自動化することについても、あわせて検討をしたい。

参考文献

[1] *"Proceedings of the Forth Message Understanding Conference (MUC-4)"*, Morgan Kaufman Publishers Inc., 1992.

[2] *"Proceedings of the Fifth Message Understanding Conference (MUC-5)"*, Morgan Kaufman Publishers Inc., 1993.

[3] 若尾 孝博. "英語テキストからの情報抽出", 情報処理学会研究報告, 96-NL-114, pp. 77-82, 1996.

[4] 木谷 強. "固有名詞の特定機能を有する形態素解析処理", 情報処理学会研究報告, 92-NL-90, pp. 73-80, 1992.

[5] Kitani,T., Eriguchi,Y. and Hara,M. "Pattern Matching and Discourse Processing Information Extraction from Japanese Text.", *Journal of Artificial Intelligence Research*, Vol. 35, No. 3, pp. 404–413, 1994.

[6] 松尾 比呂志, 木本 晴夫. "抽出パターンの階層的照合に基づく日本語テキストからの内容抽出法", 情報処理学会論文誌, Vol. 36, No. 8, pp. 1838-1844, 1995.

[7] Takemoto, Y., Wakao, T., Yamada, H , Gaizauskas, R. and Wilks, Y. "Description of NEC/Sheffield System Used for MET Japanese", *TIPSTER Text Phase II 24-month workshop*, 1996.

付録: パターンの仕様

Erieで使用可能な形態素パターン

品詞 $pattern$ = 品詞記号
(品詞記号を直接記述)

表記 $pattern$ = "形態素表記"
(文字列を " " で括る)

カテゴリ名 $pattern$ = $ カテゴリ名
($ の後にカテゴリ名を記述)

文字種 $pattern$ = 文字種記号
(KATAKANA,KANJI などの予約語を記述)

連結 $pattern = pattern_1 ... pattern_n$
($pattern$ を列記する)

選択 $pattern = [\ pattern_1 ... pattern_n\ ]$
([ ] 内に $pattern$ を列記)

積 $pattern = pattern_1 : pattern_2$

否定 $pattern = !pattern_1$

0 以上の繰り返し $pattern = pattern_1*$

1 以上の繰り返し $pattern = pattern_1+$