

## 農林関連文書からの情報抽出とその応用

劔持 和宏      山下 竜之      志賀 正裕      藤畑 勝之      森 辰則

横浜国立大学工学部電子情報工学科

近年、農林水産業の分野においても知識を共有していくという試みが始まっている。特に、農林水産業に関連する情報が書かれた大量の電子化テキストから必要な情報を容易に得たいという利用者の要求が生じている。この種の文書においては、生産物名などの特定の分野の用語や収穫量、耕地面積などの数値情報に関する情報が頻出するために、これらを考慮した情報検索システムが望まれる。本稿では、他の固有表現に加えて予め生産物名と数値情報を文書から抽出しておき、これを文書検索に用いる手法ならびに実験システムについて報告する。

## Information Extraction for documents in agriculture and forestry and its Application

Kazuhiro Kenmochi, Tatsuyuki Yamashita, Masahiro Shiga, Katsuyuki Fujihata and Tatsunori Mori

Division of Electrical and Computer Engineering,  
Faculty of Engineering, Yokohama National University

Recently, it has begun in agriculture and forestry to share knowledges about technologies. Especially, one of the largest demands of users is to retrieve information easily from huge amount of document. In the kind of documents, it is necessary to properly treat product names, numerical information and its related information, because those expressions appear frequently and are important in those domain.

In this paper, we propose a method and its experimental system, which uses extraction of product names and numerical information as well as other named entities to retrieve relevant documents.

### 1 はじめに

近年、電子化されたテキストが広く利用されるようになってきている。既に電子化されたテキストには多種多様なものが大量に存在し、また、今後増え続けるであろうことが予想される。

現在、農林水産業の分野においても、農林水産業の従事者に対して農林関連情報を積極的に公開し、知識を共有していくという試みが始まっており、農林水産業に関連する情報が書かれた大量の電子化テキストから必要な情報を容易に得たいという

ユーザーの要求が生じている。

農林水産業がトピックとなっているテキストは次のような特徴を持つ。

1. 生産物名などの特定の分野の用語が頻出する。
2. 収穫量、耕地面積などの数値情報が頻出する。

さらに、それぞれの情報がユーザーにとって重要な情報である。

このような特徴をもつテキスト群に対して検索を行う時、ある特定の生産物名に関する情報を入手

したい場合などは、従来のキーワード検索などで十分な結果が得られるであろう。しかし、例えば“神奈川県内で収量何トン以上の生産物”の検索などといった、数値情報が検索のキーになり得るものなど、従来の検索の枠組では不可能なこともある。なぜならば、検索結果のテキストが“神奈川県内で収穫された生産物”について書かれてあったとしても収量の条件が合わないテキストである場合もあり、数値情報の吟味をしないうちに、このように誤って検索される例は無くならない。また、ユーザーの要求が“ある作物を栽培している組織か生産者を知りたい”といった、検索結果として情報の要素そのものを必要とした場合、必要な情報の要素そのものはユーザーが検索結果のテキストから探し出す必要がある。

テキストから有用な情報を得るための様々な研究が行われている中で、我々は、テキストから予め特定の情報を抽出し、それを索引情報として用いた検索により農林関連文書から有用な情報を得ることを目的とした。

情報抽出に関する研究は、米国では Tipster や MUC(Message Understanding Conference) などにおいて多くの研究がなされた。日本語における情報抽出の研究も多数行われている。文献 [1] では、新聞記事のように表記の形式が定まっているテキストに対しては、記事の構造に基づくパターンマッチが有用であることを示している。文献 [2] では、テンプレートを用いた表層的なパターンマッチングで新聞記事から新製品の情報を抽出している。

情報検索にも、既に様々な手法が提案されている。例えば、よく用いられる単純な全文検索では、ユーザーが与えた語が出現するテキストを全て検索結果とする。この方法では、ユーザーが与えた語が、検索結果のテキストの主題であるかどうか分からず、検索精度が十分とはいえない。ユーザーが与えた語が全て現われないテキストでも適合するテキストとして検索結果とするベクトル空間法に代表される手法もある。また、複合語の重要性に注目した手法としては文献 [3] などがある。他には、利用者が与えた検索要求文が不十分であるときに、この検索要求文を拡張する研究も行われている。これは、利用者が与えた語と関連する語を検索要求文に付加して検索を行ない検索洩れを少なくする手法である。文献 [4] では検索対象文献データベースから自動的に類義語辞書を作成し検索要求文を拡張している。

## 2 システムの概要

テキストから情報を得るための様々な研究が行われている中で、我々は、テキストから予め特定の

情報を抽出し、それを索引情報として用いた検索により農林関連文書から有用な情報を得ることを目的とし、以下の項目について研究、検証を行った。

まず、テキストから自動的に特定の情報を収集する情報抽出システムを作成した。抽出の対象としたのは、固有表現や農林関連文書において特に重要な、生産物名と数値情報である。次に、それらを索引情報として用いる情報検索システムを構築した。

これらの検索要求を満たすために、本システムでは次の方法を用いた。

1. テキストから、検索の際の手がかりとなりうる情報の各要素を抽出しておく。
2. 抽出した情報を索引情報として検索を行う。

抽出する情報は以下の3種類の情報である。

- 固有表現
- 数値情報
- 生産物名

固有表現はそれを含むテキストを特徴付ける重要な概念となり得、数値情報は物事の推移や状態を示す重要な情報である。さらに、農林関連文書において生産物名が重要な情報であるのは言うまでもない。システム構成を図1に示す。

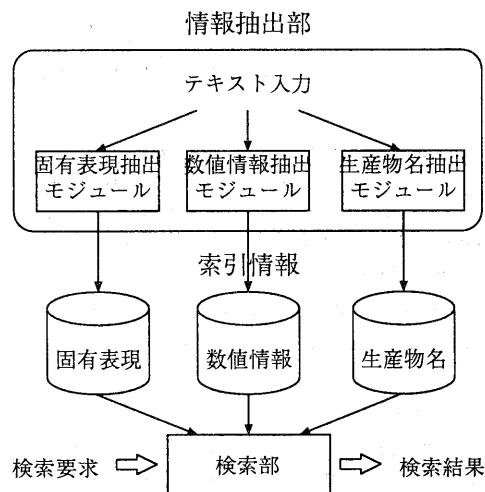


図1: システムの構成

また、固有表現、数値情報の抽出にはボタン駆動型のシステムを用いた。ボタンの作成にはいくつ

かの手法があるが、今回は大規模な正解データが存在しなかったため、バタンの作成は人手で行った。

なお、本研究の実験に用いた電子化テキストは、農業研究センターより提供して頂いた農林水産現地情報の1994年から1998年度版である。農林水産現地情報とは、農林水産業に関する新しい取り組みや、農業に従事している人々に有用な情報を全国各地の農林水産情報センターなどが収集したものである。

### 3 情報抽出部

#### 3.1 固有表現抽出

固有表現とは、テキスト中に現れる組織名、人名、地名など、ただ一つの実体を指し示す表現のことである。固有表現抽出とはこれらの表現をテキストから自動的に抽出する技術のことであり、また、情報検索、情報抽出の基礎技術として位置付けられる。抽出の対象となる概念はその目的によって異なり、本システムでは以下の項目を抽出対象としている。

**組織名** 複数の人間で構成され、共通の目的を持った組織などの名称

**人名** 人の名前、役職名、敬称などは含めない。

**地名** 場所の名前、行政区分によるもの、地形によるものを含む。

**日付表現** 時間を表す表現であり、年月日などその単位が24時間以上のもの。実時間の時間軸上を明確に指し示せる表現。

**時間表現** 時間を表す表現であり、時分秒などその単位が24時間以下のもの。実時間の時間軸上を明確に指し示せる表現。

なお、本システムの固有表現抽出モジュールは、1999年に日本で行われた情報検索、情報抽出のコンテストであるIREX(Information Retrieval and Extraction Exercise)[5]に参加したシステムを使用している。紙面の都合上この部分の詳細な説明はIREXワークショップ予稿集[6]を参照して頂きたい。

#### 3.2 数値情報抽出

数値情報とは、テキスト中に現れる数詞、及びそれに付随する情報のことである。数値情報はテキスト(特に新聞記事や報告書形式のテキスト)中に頻繁に現れ、また、それが指し示す内容が情報として重要であることが多い。しかし、キーワードマッ

チングや通常の情報検索では数値情報に関わる検索は不十分な結果に終わることもある。例えば「栽培面積100a以上でいちごの養液栽培を行っている農業者」を検索するとき、「いちご」、「養液栽培」などをキーワードにして「Aさんは1haの規模でいちごの養液栽培を行った。」などというテキストを検索できるが、同時に「Bさんは10aの広さのハウスでいちごの養液栽培を行った。」などといった、誤ったものも検索結果となるであろう。これは、数値情報の解析を行っていないからである。前述のような例において、数値に関連した情報が文意の主要な部分であるから、それらを抽出しておけば情報を検索する際の重要な手がかりになると考えている。

数値情報の核となる数詞は形態素解析によって容易に判別することができるが、もともと数詞は名詞のような非数詞情報と関連付けないとほとんど意味を持たないため、数詞とそれに関連する非数詞情報を組み合わせて抽出する必要がある。

通常、数値情報といえば前節で述べた時間表現のような、数値とその単位といった情報のことを指す。しかし、本論文で述べる数値情報抽出とは、テキスト中に現れる数値がテキストで記述されている事実とどのような関わりがあるかといったことを含めて抽出するものである。例えば、「農地Aでの収量が10%増加した。」といったテキストがあった場合、通常の情報抽出では「10%」という表記が“割合”を意味するといったことしか認識しない。一方、本研究は「収量」が「10%」の変化量で「増加」といった、テキストに記述されている事実を正確に認識することを目的とする。

情報抽出において様々な研究は行われているが、このように、数値情報に十分な注意を払った研究は少ないが、数値情報を抽出している例もある。文献[2]では製品情報の一部として“価格”という数値情報を抽出している。しかし、この場合抽出している数値情報はあくまで“価格”という特定の項目だけである。

そこで、文献[7]で述べられているように一般的に利用できる形で数値情報を抽出することを目的とした研究を行った。本研究で用いている数値情報抽出モジュールは、このシステムを利用したものである。

本モジュールで数値情報として抽出するのは以下の項目である。

**対象物名** 数詞情報によって修飾される対象物の名前。

**変化前** イベントが起こる前の状態の数詞情報。テキスト中での初期状態。

**変化後** イベントが起こった後の状態の数詞情報。  
テキスト中での終了状態。

**変化量** イベントによって変化した量の数詞情報。

**方向** イベントによる変化の方向。

イベントとはテキスト内で記述されている事象の変化のことである。なお、変化前、変化後、変化量の数詞情報は、内部でさらに次のような構造を持つ。

**見出し** 数詞の前で修飾する語

**値** 数詞

**単位** 単位

**度合** 数詞の後で修飾する語

これら全ての属性から成る情報のフォーマットと、それに対応する値を今後基本意味ユニットと呼び、それぞれの属性とその値をスロットと呼ぶ。

図2に示したのはモジュールに入力するテキストの一部であり、表1にそのテキストにモジュールを適用した際の出力結果の一部である。

近在の農家の受託作業等も含め、32haの茶園で8年4月の一番茶から操業を始めるが、12年をめどに34haまで拡大する予定である。

図2: 数値情報抽出モジュールへの入力例

表1: 数値情報出力モジュールの出力例

対象物名	変化前	変化後	変化量	方向
茶園	32ha	34ha		拡大

まず、JUMAN3.61[8]によってテキストを形態素解析する。

次に、形態素解析の結果を用いて全ての語を以下の七種のカテゴリに分類する。

**数詞** 数詞と、その単位等数詞を修飾する語。

**時間** 時間についての表現。時相名詞や、数詞と時間を表す名詞性名詞助数辞との組がこれに分類される。

**名詞** 単名詞および複合名詞。

**変化** 変化についての表現。「増」「減」等を含む語。

**助詞**

**動詞** '変化'に対応する動詞を除いた動詞。例えば、「する」。

**その他**

この際に、語の前後関係のみを参照するボタンマッチングにより、複数の語を一つの語と見なして分類を行うこともある。以下に例を示す。

- 一つの'数詞'とするもの。

- 数詞 + 名詞性名詞助数辞
- 数詞 + 名詞性名詞接尾辞
- 接頭辞 + 数詞
- 数詞 + 記号 + 数詞

例: 数詞 { 見出し = 約, 値 = 100, 単位 = 人 }

- 一つの'名詞'とするもの。

- 連続した名詞(間に「・」や「の」、「と」等をはさむものも含む)
- 接頭辞 + 名詞, 名詞 + 名詞性名詞接尾辞
- 周囲に数詞が存在しない名詞性名詞助数辞

例: 名詞 { 希望退職者の募集 }

次に、基本意味ユニットのそれぞれの属性と一致するスロットを満たすために、前段階の結果と抽出用の言語ボタンとを比較して情報を抽出する。ある一つの数詞情報に対して適用される言語ボタンは一つであり、前述した固有表現抽出システムと違って複数のボタンが適用されることはない。

最後に、基本意味ユニットを作成する。

### 3.3 生産物名抽出

#### 3.3.1 基本方針

既に行われている固有表現抽出において、商品名や小説、絵画などの作品名などの一般的な固有物名名の抽出が困難であることは、IREXなど過去に行われた情報抽出のコンテストの結果などでも良い結果が得られていないことがよく示している。

本稿では、生産物名とは農林水産業による生産物、及びそれらの加工品と定義している。我々は前

述の IREX に参加した経験もあることから、その抽出が困難であると予想していた。実際、局所的なパターンマッチを用いた固有表現抽出モジュールで抽出試験を行ったところ、その抽出精度は非常に低いものであった。これは、生産物名自身の表現に明確な特徴を見い出せることが少ないこと、さらに生産物名の近傍の表層表現を手がかりにできることが少ないことが原因となり、当初用いた局所的なパターンマッチは生産物名の抽出には有効ではなかったと判断した。そのため、我々は局所的なパターンマッチ以外の方法で生産物名を抽出することを検討した。

実際に生産物名を抽出するにあたり、テキストの以下の情報を用いた。

- 文節間の係り受け
- 文節内の表現

まず、文節間の係り受け関係を用いることにより生産物名を含む可能性のある文節（以後、候補文節と呼ぶ）を抽出する。次に、候補文節内の接続関係に着目することで生産物名を抽出する。ただし、文節の係り受けを用いずとも生産物名を抽出できる場合もあるので、これについては、局所的なパターンマッチで抽出する。

図 3 に本モジュールの構成を示す。

まず入力されたテキストを一文ずつ読み込み、形態素解析を行ったのち構文解析を行う。形態素解析には JUMAN3.61[8] を構文解析には KNP2.0-beta6[9] を用いた。

### 3.3.2 係り受け関係に基づく候補文節の抽出

我々は生産物名を含む文節を同定するにあたり、その文節の周囲のみに着目するのではなく、係り受け関係を用いてそれよりも広い範囲から生産物名を抽出する手がかりを得ることを考えた。その手がかりは、生産物に関連する特徴的な表層表現である。すなわち、注目する表現の範囲は異なるものの、特徴的な表層表現に注目する点は局所的なパターンマッチによる抽出と同じ考え方である。以下ではこの表層表現を単に「キーワード」と呼ぶ。

今回の実験では農林水産現地情報 1996 年 1 月分より人手で抽出した以下の語をキーワードとした。

サ変名詞型 生産, 栽培, 飼育, 販売, など計 29 個

動詞型 作る, 育てる, 植える, など計 10 個

名詞 1 型 特産, 商品名, 品種, など計 19 個

名詞 2 型 苗, 種, ひな, など計 21 個

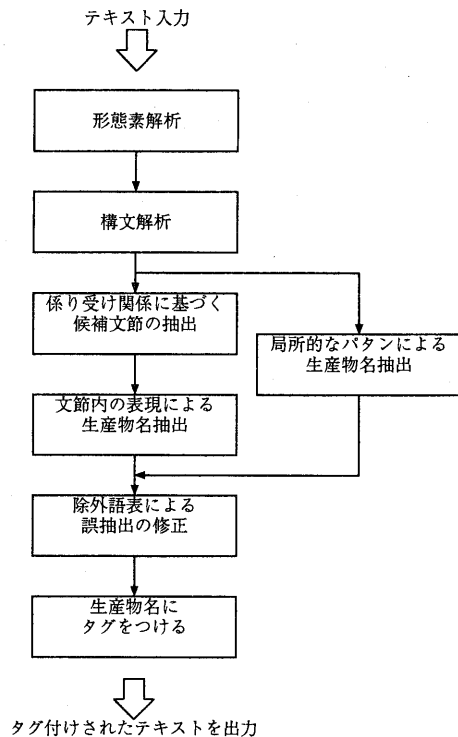


図 3: 生産物名抽出モジュール

候補文節は上記キーワードを用いて次のように抽出される。まず、生産物名抽出の手がかりとなるキーワードを含む文節を探索する。次に、そのキーワードと表 2 に示す係り受け関係を持つ文節を候補文節とする。

### 3.3.3 文節内の表現による生産物名抽出

前節の処理により抽出した候補文節から動詞・助詞・形容詞・特殊文字を除いた文字列について以下の処理を行い生産物名として抽出する。

**主辞** キーワードの探索時に、それが文節内の主辞である場合のみ生産物名抽出のための手がかりとする<sup>1</sup>。

**括弧** 抽出対象が括弧内にある場合、構文解析で正確に文節が区切られていない可能性を考慮し、括弧内の表現は全て抽出する。

<sup>1</sup>ただし、助詞は除く。

表 2: キーワード群の分類と抽出対象との関係

キーワード	条件	候補文節
サ変名詞型	受動態 連格	キーワードを受ける 文節
	受動態	キーワードに係る ガ格・末格
	その他	キーワードに係る ヲ格・ノ格・隣接
動詞型	受動態 連格	キーワードを受ける 文節
	受動態	キーワードに係る ガ格・末格
	その他	キーワードに係る ヲ格・隣接
名詞 1 型	同格連体 末格・ノ格	キーワードを受ける 文節
	その他	キーワードに係る ガ格・末格
名詞 2 型		キーワードに係るノ格

### 3.3.4 局所的なボタンによる生産物名抽出

先に述べたように、係り受け関係を用いずとも生産物名を抽出できる場合もあるので、これについては、局所的なボタンを用いる。具体的には、個々の文節内の表層表現において特徴的な表現を探索し、その直前の名詞連接を生産物名として抽出する。ここでは、先ほどと同様に人手で“栽培”、“畑”、“作り”など11個を選択して用いている。

### 3.3.5 除外語表による誤抽出の修正

キーワードと係り受け関係をなす表現はほとんどが生産物であるが、いくつか生産物でないものも出現する。これを除外するために処理のなかでさらに別の情報を調べることも考えられるが、我々は、これらが比較的定型の表現であることに注目して、除外語表に登録し、一連の処理の最後でこれらを削除するという方針をとった。除外語表はキーワードの抽出に用いたコーパスから人手で選択した。

### 3.3.6 生産物名抽出の一例

‘旋盤を改造した機械を使ったとろろこんぶの生産が’という文を例として用いる。この文を構文解析すると‘とろろこんぶの’と‘生産が’が係り受け関係になっていることが分かる。

“生産”というサ変名詞型のキーワードがあるた

め、これに係る文節の中で、ノ格かヲ格の文節を抽出対象とする。この場合、‘生産が’という文節に係る‘とろろこんぶの’という文節がノ格であるため、この文節から助詞の“の”を除いた“とろろこんぶ”を抽出対象とする。

## 3.4 情報検索部

情報検索部では、各抽出モジュールを用いて抽出した固有表現と数値情報、農産物名を索引情報として検索を行う。

表 3: 情報検索モジュールで用いる索引情報一覧

固有表現	組織名	数値情報	対象物名
	人名		変化前
	地名		変化後
	日付表現		変化量
	時間表現		方向
	農産物名		

この検索システムでは、抽出しておいた索引情報の分類自体を検索する項目として用いることができる。例えば、ある作物を栽培している会社、あるいは生産者を検索する場合、従来の手法ならば検索結果のテキストを読んでユーザーが探す必要があるが、本システムならば、会社名、あるいは生産社名などの項目を直接ユーザーに提示することを可能とする。

さらに、数値情報を用いた検索を行えるため、‘何 ha 以上の規模で栽培を行っている’などの条件にも対応できる。

### 3.4.1 検索システムの構成

まず、検索質問の作り方について説明する。検索質問は検索項目と検索条件から成る。

**検索項目** ユーザーが求める情報の分類名である。

**検索条件** 検索の手がかりとするキーワードであり、上記の分類名にしたがった入力を行う。

分類名とは、表 3 に示した、組織名、人名、地名といった固有表現や、対象物名、変化前、変化後といった数値情報である。ユーザーの検索要求として‘神奈川県でいちごを生産している組織か生産者を知りたい’というものがある時、検索質問は図 4 のようになる。

図 5 に検索の手順を示す。

まず、検索項目、検索条件を入力する。検索項目、検索条件はそれぞれ複数の指定が可能である。

検索項目：組織名，人名  
 検索条件：固有物名 = いちご  
 地名 = 神奈川県

図 4: 検索エンジンへの入力例

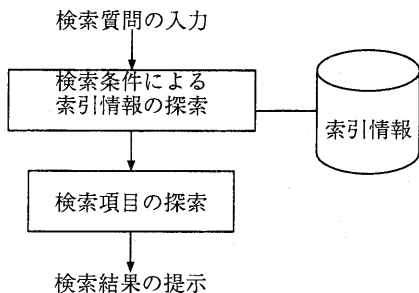


図 5: 検索エンジン

次に、検索条件に入力された固有表現、数値情報のそれぞれの検索を行い、検索条件の出現するドキュメント毎に出現頻度を求め、重み付けを行う。

さらに、検索条件の出現するドキュメント中の、検索項目に分類される情報を検索する。最後に、求めておいたドキュメントの重み順に検索結果を提示する。

検索結果の一例を図 6 に示す。

検索結果: ドキュメント ID = 199601050  
 組織名 = 神奈川県総合農業研究所

図 6: 検索エンジンの出力例

### 3.4.2 固有表現の検索

固有表現はそれぞれの分類毎に索引情報を作成する。索引情報は、固有表現をキーとし、それが出現するドキュメントの ID を値とした Inverted file により構築している。ある一つの固有表現が複数のドキュメントに出現する場合は、その全てのドキュメント ID が保存されており、ドキュメント ID 自身も複数の項目を持つ。

固有表現を検索する際には、それが出現するドキュメントの ID を即座に求めることができる。

### 3.4.3 数値情報の検索

数値情報抽出モジュールにより抽出した基本意味ユニットのうち、変化前、変化後、変化量の各数値情報は

- 見出し
- 値
- 単位
- 度合

から成り、'見出し'、'度合'には、数値に幅を持たせるような性質を持つものが抽出されることがある。そのため、検索時にはこれらの情報を考慮し、幅を持たせた検索を行わなければならない。そのため、数値による検索をする時には、曖昧性や数値の範囲を示す語などを伴った数値情報を、数値に展開した上で検索する必要がある。

## 4 試験結果及び評価

表 4 に固有表現抽出モジュールの試験結果、表 5 に数値情報抽出モジュールの結果、表 6 に正産物名抽出モジュールの結果を示す。

表 4: 農林水産現地情報からの固有表現抽出結果

タグ	再現率 (%)	適合率 (%)
ORGANIZATION	44.37	57.76
PERSON	71.21	34.81
LOCATION	66.05	58.68
DATE	91.83	72.90
TIME	100.00	28.57
All Slots	69.93	58.92
F-measure		62.98

表 5: 農林水産現地情報からの数値情報抽出結果

スロット	再現率 (%)	適合率 (%)
対象物名	87.5	40.0
変化前	100.0	44.4
変化後	90.9	62.5
変化量	90.0	60.0
方向	66.7	66.7

固有表現抽出においては、ORGANIZATION, LOCATION は精度が低い原因が互いに関連する。農林水産現地

表 6: 農林水産現地情報からの生産物名抽出結果

	再現率 (%)	適合率 (%)
	50.00	68.00
F-measure	58.00	

情報では新聞記事と比べて ORGANIZATION として抽出すべき部分の表現自体が長く複雑なものも多く、また、その内部に地名が含まれていることも多い。そのため、その部分全体を ORGANIZATION として認識することが難しくなった。全体を ORGANIZATION として抽出することに失敗した場合、その部分に含まれている地名が LOCATION と認識され、過抽出してしまう。DATE, TIME は特定の時を表さない時間表現も抽出していた例が増大したことが、適合率が低い原因である。例えば、農産物の育成期間などは、“1年間”という意味での‘1年’という表記をされることがほとんどであった。

数値情報抽出においては、全体的な傾向として適合率が低い。これは、本来そのスロットは空であるのが正解であるのにテキスト中の何らかの語を結果としてしまうことが多いからである。また、‘A から B’ という表記に対して、変化という意味と範囲という意味があり、それらを混同することにより誤抽出、あるいは過抽出する傾向がある。

## 5 おわりに

本稿では、大量のテキストから利用者が求める情報を効率的に得る手段として、固有表現、数値情報、生産物名をテキストから抽出しておき、それらを索引情報としてテキスト内容の検索を行う手法を提案した。

我々が今回情報の抽出対象とした農林関連文書では、固有表現、数値情報、農林水産業による生産物名が重要な情報として位置付けられる。

生産物名の抽出においては、今まで我々が用いていた局所的なパターンマッチによる抽出手法はあまり有効ではないことが判明したため、構文解析を行い文節間の係り受け関係を用いることによる抽出を試みたところ、精度が大幅に向上した。しかし、抽出精度、及び抽出失敗例を見る限り、各抽出モジュールには未だ改善の余地があると考えている。

情報検索部については、インターフェースの問題、自然言語による問い合わせの実現など、考慮すべき点も残したシステムとなってしまったが、数値情報を用いた検索や、検索結果として情報の要素を提示することを可能にした点では、現在の検索エン

ジンにより基本的な枠組を構築できたと考えている。

## 謝辞

本研究は農林水産省の一般別枠研究「増殖情報ベースによる生産支援システム開発のための基盤研究」の一環として行なわれた。

## 参考文献

- [1] 西野文人, 落谷亮. 新聞記事からの人物・企業情報の抽出. 情報処理学会研究報告 98-NL-127-17, 情報処理学会, Sep 1998.
- [2] 井出裕二, 永井秀利, 中村貞吾, 野村浩郷. 単一項目テンプレートによる新聞記事からの製品情報抽出. 情報処理学会研究報告 97-NL-122-10, 情報処理学会, Nov 1997.
- [3] 山田剛一, 斉藤公一, 森辰則, 中川裕志. 複合語マッチングによる情報検索. 言語処理学会第3回年次大会発表論文集, March 1997.
- [4] 斉藤公一, 森辰則, 中川裕志. 概念に基づく検索要求文の拡張. 情報処理学会研究報告 97-FI-47-10, 自然言語処理情報学基礎合同研究会, 情報処理学会, Sep 1997.
- [5] IREX 実行委員会. IREX. <http://cs.nyu.edu/cs/projects/proteus/irex/>, 1998.
- [6] 剣持和宏, 宮本昌幸, 松尾衛, 内間圭介, 森辰則. バタンによるシステムと EDR 概念辞書と決定木学習を用いたシステム. IREX ワークショップ予稿集, pp. 211-218, Sep 1999.
- [7] 斉藤公一, 迫田昭人, 中江富人, 岩井禎広, 田村直良, 中川裕志. 数値情報をキーとした新聞記事からの情報抽出. 情報処理学会研究報告 98-NL-125, 自然言語処理研究会, 情報処理学会, May 1998.
- [8] 黒橋禎夫, 長尾真. 日本語形態素解析システム JUMAN version 3.6 使用説明書, 1998.
- [9] 黒橋禎夫. 日本語構文解析システム KNP version 2.0 b6 使用説明書, 1998.