

ドキュメント・データを対象にした タイムコーディングシステムの実現方式

石川 幹直[†] 細川 宜秀[†] 高橋 直久[†]

[†]名古屋工業大学工学部電気情報工学科 〒466-8555 愛知県名古屋市昭和区御器所町
E-mail: †ishikawa@moss.elcom.nitech.ac.jp, ††{hosokawa,naohisa}@elcom.nitech.ac.jp

あらまし 本稿では時間的関連性を伴うドキュメント・データに対しタイムコーディングを自動で行うシステムを提案する。タイムコーディングとは、ドキュメント・データに含まれる時間表現をタグ付けすることである。時間表現の実現方式には座標表現のみにより構成される時間表現 (type-1) と座標表現に時間を特定する要素を組み合わせて構成される時間表現 (type-2) が存在し、type-2 の表現はドキュメント・データ中に多く含まれ、時間の表現能力が高いことが示されている。したがって、大量のドキュメント・データから type-2 の表現を抽出するためには自動化機構の実現が本質的に重要である。提案システムは、時間を表す単語から type-2 を決定する機構を実現することによって、ドキュメント・データ全体の文脈認識を伴わずにタイムコーディングが実現可能になる。さらに実験により、時間タグ自動抽出の可能性を示す。

キーワード 時間 DB, テキスト DB, 情報検索

A Time Coding System for Document Databases

Motonao ISHIKAWA[†], Yoshihide HOSOKAWA[†], and Naohisa TAKAHASHI[†]

[†] Department of Electrical and Computer Engineering, Nagoya Institute of Technology
Gokiso, Showa, Nagoya, 466-8555 Japan

E-mail: †ishikawa@moss.elcom.nitech.ac.jp, ††{hosokawa,naohisa}@elcom.nitech.ac.jp

Abstract In this paper, we present an implementation method of time coding systems for document databases. We define systems for extracting temporal data items from document databases as time coding systems. There are following two types of temporal data items: (type-1) a coordinate, and (type-2) a pair of a coordinate and a context. Many type-2 representations are included in document data. The main features of our method are as follows. (Feature-1) We realize a time coding system for extracting type-2 representations from document databases. (Feature-2) We realize the time coding system without complicated context recognition mechanisms. The proposed method makes it possible to apply temporal operators to document databases. We clarify feasibility of our method by showing several experimental results.

Key words Temporal DB, Text DB, Information Retrieval

1. はじめに

現在、ネットワークの発達に伴い、知識情報を含むドキュメント・データが分散配置されている。それらのドキュメント・データ群からの検索・分類による知識獲得技術が重要になっている。それらのドキュメント・データの多くが長年の知的生産活動によって蓄積されてきたことを考慮すると、時間に関するドキュメント・データ群の検索・分類機能の実現はその知識獲得技術の確立に大きく貢献する。

ドキュメント・データ群の検索・分類機能として、ドキュメント・データの作成日や更新日に基づいた時間に関する検索・分類手法が実現されている。しかし、ドキュメント・データに

含まれる時間表現を扱っておらず、現在十分な検索・分類手法を提供するには至っていない。ドキュメント・データに含まれる時間表現によってドキュメント・データ群を時間軸上に写像する機能は、ドキュメント・データを対象とした検索・分類のための基本機能として位置づけられる。ドキュメント・データに付与する時間表現を時間タグと定義する。

時間表現には、次の 2 タイプがある [3]。

type-1 時間軸上の座標で表現される時間表現 (座標表現)
type-2 座標表現に文脈を組み合わせて構成される時間表現
type-2 を対象とした時間的関連性評価機構が type-1 を対象とした時間的関連性評価機構よりドキュメント・データ検索への適用性が高いことが示されている [3] ので、type-2 を含む時間タ

生成手法を確立することが必要である。

本稿ではそのような時間タグを自動で生成することをタイム・コーディングと定義し、タイム・コーディング・システムの実現方式を提案する。本システムの特徴は、前置詞句に含まれる単語からその前置詞句が時間表現であると確定できるものを簡単なルールによって座標に変換することにより、ドキュメント・データの文脈認識を伴わずにタイム・コーディングを可能にする点にある。ここでドキュメント・データの文脈認識とは、ドキュメント・データの記述内容に基づいて単語の意味を判断することである。

手でタイム・コーディングを行った場合と比較して、タイム・コーディング・システムの自動化の可能性を検証する。さらに、提案するタイム・コーディング・システムを適用した時間的関連性評価を伴うドキュメント・データ検索システムを実装し、提案システムのドキュメント・データ検索への適用性を検証する。

提案システムの対象は、英文で記述されたドキュメント・データ群である。

2. 提案システムの実現方式

提案システムの特徴は、時間を表す単語を含む前置詞句を対象とすることによってドキュメント・データの文脈認識を伴わずに時間表現に変換することである。

さらに、提案システムの実現のポイントは、ドキュメント・データに含まれる異なる時間表現間の時間的関連性を評価可能にするために、共通の時間軸上に写像するための機能を実現する点にある。例えば、ドキュメント・データ群に含まれる時間表現には、西暦表現、月表現、曜日表現などの異なる時間表現がある。これらの表現間の時間的関連性を評価可能にするためには、これらと比較するための共通の時間軸上の値に変換する機能の実現が必要である。

提案システムは、次の step-1 から step-4 の手順にしたがって、ドキュメント・データに含まれる type-2 を抽出する。図 1 は、提案システムの実行手順を表す。ここで、提案システムは、その入力として、ドキュメント・データの識別子、そのドキュメント・データに含まれる 1 文をとる。

step-1 ドキュメント・データ(文)の構文木への変換

既存の構文解析機能 [2] を用いて、ドキュメント・データから S 式で表現された構文木を生成する。ここでドキュメント・データは、文ごとに生成する。図 1 では、“I live in Nagoya on Friday.” を構文解析することで S 式の構文木を生成している。図 1 にある構文解析結果において、S は主部、NP は名詞句、VP は動詞句、PP は名詞句を表す。

step-2 前置詞句の抽出

S 式で表現された構文木から前置詞句を抽出する。構文解析機能では、前置詞句に 'PP' のラベルが貼られているため、'PP' のラベルを持つ句をすべて抜き出す。図 1 では、“(PP IN (NP NAGOYA))” と “(PP ON (NP FRIDAY))” が抽出される。

step-3 前置詞句を type-2 に変換

複雑な文脈認識を伴わずに type-2 へ変換する。これより、時

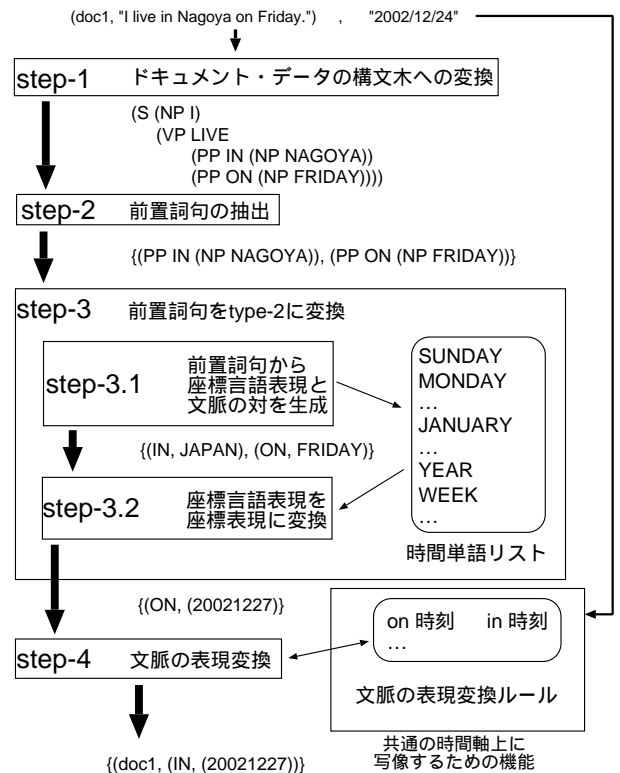


図 1 提案システムの実行手順

間単語リストと簡単な時間表現生成ルールを用いることで座標表現に変換することが可能になる。その実行手順を次に示す。

step-3.1 前置詞句から座標言語表現と文脈の対を生成

前置詞句を座標に変換できそうな言語表現（以下、座標言語表現）とその他の部分（以下、文脈）に分ける。ここで、新聞記事などのドキュメント・データにおいて、時間を表す名詞句はその末尾の単語が時間を表す単語であることが多いことに着目する。提案システムでは末尾の単語が時間単語リストに含まれるかを比較し、含まれるならば座標言語表現、そして残りの部分を文脈とする。時間単語リストに含まれない場合は時間表現と認識しない。これにより提案システムは、文脈認識のための複雑な自然言語処理を伴わずに、前置詞句が時間表現を表しているかどうかを判断することが可能になる。図 1 では、提案システムが単語リストに含まれる FRIDAY を時間を表す前置詞句として認識する。

step-3.2 座標言語表現の座標表現への変換

ここでは、座標言語表現を、共通の時間軸上の座標表現に変換する。

表 1 は、座標言語表現を西暦時間軸上の座標表現に変換するためのルールを表す。ここで、ルール 1, 4, 6 は、西暦表現を西暦時間軸上の座標表現に変換するためのルールを表す。

ルール 2, ルール 3, ルール 5, ルール 7 は、西暦以外の座標言語表現を西暦時間軸上の座標表現に変換するためのルールを表す。ここで、これらのルールは、入力として「西暦時間軸のある時刻」をとる。この時刻を入力としてとる理由は、これらのルールの適用対象となる座標言語表現が、西暦時間軸のある時刻からの相対的な時間を表すからである。

表1 座標言語表現を座標表現に変換する例（出版日を2002年4月7日とする）

ルール	変換対象時間軸	入力	末尾の名詞句	変換例
ルール-1	西暦時間軸	-	4桁整数	1999 → (19990101,19991231)
ルール-2	西暦時間軸	出版日	月	SEPTEMBER → (20020901,20020930)
ルール-3	西暦時間軸	出版日	月日	APRIL 11 → (20020411)
ルール-4	西暦時間軸	-	月年	FEBRUARY 1995 → (19950201,19950228)
ルール-5	西暦時間軸	出版日	曜日	FRIDAY → (20020405)
ルール-6	西暦時間軸	-	月日,年	ARG. 26, 1998 → (19980826)
ルール-7	西暦時間軸	出版日	数字+時間の単位	23 DAYS → (20020407,20020429)
ルール-b	野球時間軸	-	INNING(S)	EIGHT INNING → (80,82)

ルールbは、座標言語表現を野球の時間軸上の座標表現に変換するためのルールを表す。

各ルールについての詳細を次に述べる。

ルール1 年

日付に0101, 1231を付加して座標にする。名詞句に4桁整数の単語が1つだけ存在することが条件になる。

ルール2 月

記事の日付を取得し、月を数字に変換して座標にする。名詞句に時間単語リストに含まれる月の表現が1つだけ存在していることが条件になる。

ルール3 月日

記事の年を取得し、付加する。また期間で表現するため前置詞がONならばINに変換する。名詞句に単語が2つ存在し、時間単語リストに含まれる月の表現、2桁の整数と続くことが条件になる。

ルール4 月年

記事に書かれた月、1ヶ月のこととする。名詞句に単語が2つ存在し、時間単語に含まれる月の表現、4桁の整数と続くことが条件になる。

ルール5 曜日

記事の日付から曜日を計算し、そこからいちばん近い曜日の日とその記事の時間表現が示していることとすることで算出する。名詞句に、時間単語リストに含まれる曜日の表現が1つだけ存在することが条件になる。

ルール6 月日,年

表現が長くなるため月が省略されて「.」が使われることがあるので、解析前に変換する。日付の情報がきちんと与えられているので、そのとおりに座標に変換する。名詞句の単語が複数存在し、月表現、2桁整数+カンマ、4桁整数の順に続く場合が条件になる。

ルール7 数字表現+時間の単位

時間の単位にはDAY(S), WEEK(S), MONTH(S), YEAR(S)がある。これらが存在した場合、期間を表すことが考えられるので記事の日付を取り出し、そこから過去へさかのぼって算出する。名詞句の単語が複数存在し、時間の単位が時間単語リストに含まれ、その前の単語が数字を表す表現になっていることが条件になる。

図1では、“FRIDAY”は曜日の表現のためルール5が適用され、2002/12/24からもっとも近い金曜日である2002/12/27をそ

の記事の時間と認識する。

また、上記7つのルールに加え、ドキュメント・データ固有の時間表現が存在する。例えばスポーツ記事において、スポーツ固有の時間表現がある。写像対象時間軸もそれ固有のものを用意する必要がある。野球の記事に関しては、次のルールが考えられる。

ルールb inning(s)

名詞句の最後にinning(s)を含む場合、野球の試合開始から終了までの時間軸に写像する。野球の試合時間を表す時間軸は離散的なものであり、イニングとアウト数を併せて表す。EIGHT INNINGなら(80,82)と表現する。

step-4 文脈の表現変換

座標言語表現を座標表現に変換した際、本来利用するはずのない前置詞に対応される場合がある。写像対象時間軸に対応させるために前置詞と座標の対応を見て、間違っているようならば置きかえる。期間を表す座標表現でありながら前置詞がONの場合に、INへ変更する。図1では期間を表す座標表現(20021227)のため、前置詞をINに変更している。

出力で得られるIDはドキュメント・データに対してつけられている。そのためドキュメント・データが複数の時間表現を持つこともある。

3. 実験

検証項目は次の3点である。

検証1 時間タグ自動抽出システムの実現可能性検証

検証2 座標言語表現の座標表現への変換ルールの妥当性検証

検証3 検索システムに応用した際の適用性検証

時間情報を含むドキュメント・データに対し、提案システムの再現率、適合率を評価する。ここで再現率、適合率とは次の式で定義されるものである。

$$\text{再現率} = \frac{\text{提案システムが出した正解数}}{\text{手動により決定した正解数}}$$

$$\text{適合率} = \frac{\text{提案システムが出した正解数}}{\text{提案システムが出した全体数}}$$

ここでの正解とは時間表現を含む前置詞句を手動で抽出した表現に一致するものである。

本実験では実験データとして、英字新聞The Japan Times [4]の1995年から2002年における野茂英雄投手に関する新聞記事248件を使用した。

表 2 構文木生成の成功率

使用文数	構文解析成功文数	成功率
1808 文	1214 文	66.78%

表 3 時間表現を含む前置詞句抽出精度

使用文数	抽出前置詞句数	再現率	適合率
316	448	100%	70.36%

表 4 座標言語表現に対する座標変換精度

時間タグ候補数	座標変換成功数	再現率	適合率
334	293	87.72%	100%

表 5 前置詞変換精度

出現	自動総数	自動正解	再現率	適合率
80	80	80	100%	100%

表 6 提案システムを通じた結果 (ドキュメント 248 件に対して)

手動正解	自動生成	自動正解	再現率	適合率
334	293	135	40.42%	46.08%

3.1 時間タグ自動抽出システムの実現可能性検証実験

3.1.1 実験方法

本実験では、提案システムの各 step の実行結果の再現率と適合率を評価することによって、各 step の実現方式の妥当性を検証する。

さらに、提案システムにおける各 step を連動させ、その結果の再現率と適合率を評価することによって、時間タグ抽出の自動化可能性を検証する。

本実験では、ドキュメント・データ 248 件を type-2 の抽出対象とした。

3.1.2 実験結果と考察

表 2, 表 3, 表 4, 表 5 は、それぞれ提案システムの step-1, step-2, step-3, step-4 に正しい入力を与えた場合における出力結果の再現率と適合率を表す。ここで、各 step への正しい入力は、248 件のドキュメント・データから人手によって作成した。

表 3 より、step-2 がドキュメント・データに含まれるすべての type-2 を抽出することを確認した。また、step-2 の適合率が約 7 割程度に落ちた理由は、前置詞句とともに時間表現とは関係のない前置詞句もまとめて抽出されるためである。しかし、この手続きの目的は、ドキュメント・データに含まれるすべての前置詞句を抜き出すことにあるので、再現率が 1 になっているという事実のみが重要である。これより、step-2 の実現方式の妥当性を明らかにした。

表 4 より、step-3 がドキュメント・データから時間を表すすべての前置詞句を抽出することを確認した。また、step-3 の再現率が 9 割弱に低下した理由は、座標言語表現を座標表現に変換する際に、変換できない名詞句が存在したためである。その理由を次に示す。

理由 1 1 つの名詞句に複数の時間表現を含む。

“in 1952, 1953” という表現は構文解析によって (NP (NP 1952) <COMMA> (NP 1953)) と表現された。しかし本稿で示したルールは、1 つの名詞句に複数の情報をもつものを対象としていな

いたために時間表現として認識しなかった。

理由 2 時間を表す単語が含まれていない、時間を表す前置詞句が存在する。

今回使用したドキュメント・データにおいて、時間を表す単語が含まれていないが時間を表す前置詞句 (たとえば、“in the seventh”) が含まれていた。提案システムは、時間を表す単語を含む前置詞句を対象としているので、それを含む文あるいは文書の意味認識は行われない。結果として “in the seventh” は「野球の 7 回」を示すものだが、“in the seventh” を時間表現として認識せず、時間軸上の座標を確定できなかった。

しかし、248 件のドキュメント・データに対して再現率の低下が 9 割弱程度に抑えられたことを考慮すると、step-3 の実現方式の妥当性が統計的に確認できたと判断できる。

表 5 より、step-4 がドキュメント・データに含まれるすべての type-2 のみを抽出することを確認した。すなわち、今回使用したドキュメント・データ 248 件の中で前置詞変換が必要な 80 件について、すべて前置詞句を変換することができた。これより、step-4 の実現方式の妥当性が明らかとなった。

表 6 は、各 step を連動させて type-2 を抽出した結果の再現率と適合率を表す。この結果からは、時間タグ抽出の自動化は困難であるように思えるが、表 2 とこれまでの考察から、提案システムの再現率と適合率の低下は step-1 が引き起こしたものと判断できる。本実験における step-1 の動作を確認すると、野球に関する専門的な意味を表す単語や固有名詞が、step-1 で使用した構文解析機構の辞書に登録されておらず、それが、構文解析の精度を低下させた。したがって、構文解析の辞書を整備することによって、構文解析機構の精度向上が期待できる。すなわち、提案システムの再現率と適合率の向上が期待できる。

以上の考察から、提案する時間タグ抽出方式の自動化は可能であると判断できる。

3.2 座標言語表現から座標表現への変換ルールの妥当性検証実験

3.2.1 実験方法

本実験では、表 1 に示すルールの妥当性を検証する。

表 1 に示すルールは、ドキュメント・データ 44 件から人手で生成したものである。表 1 に示すルールの妥当性を明らかにするために、このルールを 248 件のドキュメント・データに適用し、提案システムの type-2 抽出結果の再現率と適合率の低下が抑えられることを示す。

3.2.2 結果と考察

表 4 は、248 件のドキュメント・データに含まれる座標言語表現を座標表現に変換した結果の再現率と適合率を表す。この結果より、再現率の若干の低下は認められたが、提案システムが高い再現率と適合率を示した。これより、表 1 に示すルールの妥当性を明らかにした。

3.3 ドキュメント・データ検索システムへの適用性検証実験

3.3.1 実験方法

本実験では、提案システムが時間的関連性評価を伴うドキュメント・データ検索システムに適用性が高いことを検証する。提案システムの適用性が高いことを検証するために、次に示す

表7 検索実験に使用した入力

提案検索システム	'nomo','since','(19970101),(20010101)'
比較対象システム1	'nomo','(19970101),(20010101)'
比較対象システム2	'nomo'

3つのドキュメント・データ検索システムによる検索結果の精度の比較を行う。

提案検索システム type-2 による時間的関連性評価を伴うドキュメント・データ検索システム

このシステムは提案システムをドキュメント・データ検索システムに適用したシステムである。このシステムに与える入力のデータ構造はドキュメント・データの識別子、ドキュメント・データ、ドキュメント・データに含まれる座標表現、および文脈の組とする。実行手続きは次のとおりである。(手続き1)ドキュメント・データに時間情報をタグ付けする。(手続き2)(手続き1)の結果得られた時間タグに時間的関連性評価機能を適用し、利用者によって与えられた時間的関連性を満足するドキュメント・データの組を抽出する。(手続き3)(手続き2)の結果得られたドキュメント・データの集合に対する検索キーワードの包含性を評価し、検索キーワードを含む組を抽出する。

比較対象システム1 type-1 による時間的関連性評価を伴うドキュメント・データ検索システム

このシステムは従来の座標表現のみの評価による時間的関連性評価機能をドキュメント・データ検索システムに適用したシステムとして位置付けられる。このシステムに与える入力のデータ構造はドキュメント・データの識別子、および、ドキュメント・データに含まれる座標表現の組とする。実行手続きは次のとおりである。(手続き1)ドキュメント・データの組集合の座標表現に時間的関連性評価機能を適用し、利用者によって与えられた時間的関連性を満足するドキュメント・データの組を抽出する。(手続き2)(手続き1)の結果得られた組集合のドキュメント・データに対する検索キーワードの包含性を評価し、検索キーワードを含む組を抽出する。

比較対象システム2 時間的関連性評価を伴わないドキュメント・データ検索システム

このシステムは現在の主要なドキュメント・データ検索システムの1つとして位置付けられる。ドキュメント・データの集合に対する検索キーワードの包含性を評価し、検索キーワードを含むドキュメント・データを抽出する。

指標には3節の始めに示した再現率、適合率を用いる。

3.3.2 実験環境

実験データには英字新聞 The Japan Times [4] の1995年から2002年までの野茂英雄投手に関する新聞記事248件を使用した。

本実験では、時間的関連性評価機構を Microsoft SQL Server 2000 上で稼働させた。キーワードの包含性評価については SQL Server 2000 標準の機能を用いた。type-1 と type-2 を対象とした時間的関連性評価機能を、ユーザ定義として実装した。

3.3.3 結果と考察

図2は、提案検索システムと、比較対象システム1,2の再現率、適合率を表す。表7は、各システムに与えた検索条件を表

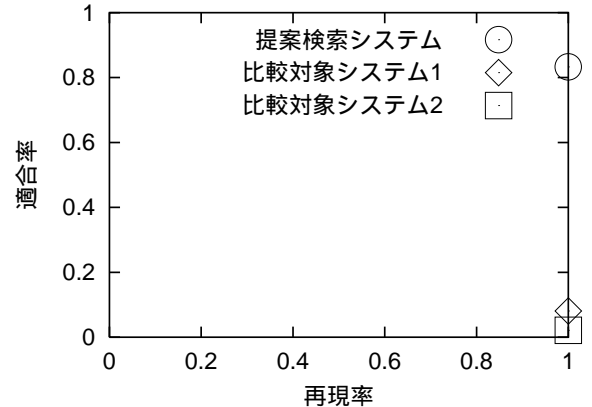


図2 検索システムに適用した際の再現率、適合率

す。ここで、検索者が次の意図を持って検索条件を与えたものとした。

「野茂選手が1997年から2001年になるまでの期間で、野茂選手が久しぶりに行ったことを示す記事を検索する。」

比較対象システム1の適合率は、比較対象システム2よりも適合率が高くなった。また、さらに提案検索システムは適合率が高くなった。これらの結果より、文献[3]と同様の結果を得ることができたため、提案システムは、時間的関連性評価を伴うドキュメント・データ検索システムにおいて高い適用性を持つことを明らかにした。

4. 関連研究

文献[7]は、ドキュメント・データに含まれる機能語から深層格の出現頻度を求め、その出現頻度を用いてドキュメント・データ間の類似性を計算するための手法を示している。その類似性の計算において、時間的関連性評価は行われていない。文献[5]は、文献[7]で示された方式を拡張した方式を示している。この方式では、ドキュメント・データに対する深層格の寄与度を簡単な構文解析を伴って計算し、その寄与度をドキュメント・データ間の類似性計算に利用している点が特徴である。これらの方式で扱う深層格の中には、TIMEなどの時間に関する深層格があり、時間に関する深層格の類似性が、ドキュメント・データ間の類似性に貢献することを考慮すると、この方式は、ドキュメント・データに含まれる時間情報を考慮したドキュメント・データ検索方式として位置付けられる。

一方、提案システムは、ドキュメント・データに含まれる時間表現を自動抽出するための機構を実現することによって、文献[1]や文献[3]で示された時間的関連性評価機能をドキュメント・データベースに適用可能にする。

文献[6]は、既存の全文検索システムを、その変更を伴わずに、時間情報を利用したドキュメント・データ検索システムに拡張するための方式を示している。その特徴は、ドキュメント・データに関連する時間を含む時間間隔を単語として表現し、その単語をドキュメント・データに追加することにある。これによって、ある期間中のドキュメント・データを検索することを可能にする。しかし、問合せの条件として任意の時間間隔を指

定可能にするためには、それに対応する単語を登録する必要があり、その登録に要するオーバーヘッドは大きいものとなる。

一方、提案システムは、そのような単語を用いることなく、ドキュメント・データに含まれる時間表現と検索条件として与えられた時間表現間の時間的関連性を計算するための機能 [1], [3] に適用可能なデータをドキュメント・データから自動抽出するものである。したがって、提案システムと時間的関連性評価機能を組み合わせて実現されるドキュメント・データ検索システムでは、時間間隔を表す単語を登録することなく、任意の時間間隔を検索条件として指定可能になる。

5. おわりに

本稿ではドキュメント・データを対象としたタイム・コーディングを自動で行うシステムを提案した。タイム・コーディングとは、ドキュメント・データに含まれる時間表現をタグ付けすることである。時間表現の実現方式には座標表現のみにより構成される時間表現 (type-1) と座標表現に時間を特定する要素を組み合わせて構成される時間表現 (type-2) が存在し、type-2 の表現はドキュメント・データ中に多く含まれ、時間の表現能力が高いことが示されている。大量のドキュメント・データから type-2 の表現を抽出するには自動化機構の実現が本質的に重要であった。提案システムは、時間を表す単語から type-2 を決定する機構を実現することによって、ドキュメント・データ全体の文脈認識を伴わずにタイム・コーディングが実現可能になった。さらに実験により、時間タグ自動抽出の可能性を示した。

提案システムの対象が前置詞句だったが、今回使用したドキュメント・データにおいては“last year”, “last week”などの副詞による時間表現が複数存在するため、今後副詞句も対象にすること、またひとつの名詞句に複数の時間を表す単語が含まれていた場合も対象にすること、そして Web 検索システムに応用していくことが課題にあげられる。

謝辞 本研究の一部は、日本データベース学会・マイクロソフト株式会社共催 2002 年度データベース研究支援プログラムの支援によるものである。

文 献

- [1] Allen, J.F.: Maintaining Knowledge about Temporal Intervals, *Comm. of the ACM*, No.26, pp.832-843(1983)
- [2] Davy Temperley, Daniel Sleator, John Lafferty, Link Grammer, <http://www.link.cs.cmu.edu/link/>
- [3] 細川 宜秀, 清水 康: 文脈認識をともなった時空間的関連性評価方式, *情報処理学会論文誌: データベース*, Vol.43, No. SIG5(TOD 14), pp.118-133
- [4] The Japan Times Monthly Bound Volume, The Japan Times. Ltd.
- [5] Liu, G. Z.: Semantic vector space model: Implementation and evaluation, *J. the American Society for Information Science*, Vol. 48, No. 5, pp.395-417 (1997)
- [6] 増井 俊之: インターフェイスの街角 63-実用的な全文検索システム-, 株式会社アスキー, *UNIX MAGAZINE Vol.18 #3*, pp.177-182(2003)
- [7] Wendlandt, E.B. 1991 Incorporating a semantic analysis into a document retrieval strategy. In *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.270-279