

電子メールコミュニケーションにおけるスケジュール情報抽出

長谷川 隆明 高木 伸一郎

NTT 情報通信研究所

hasegawa@isl.ntt.co.jp

一度に多数の人に情報を伝達できる電子メールは、送信する側にとって都合が良い。しかし受信する側では、人手を介さずに情報を抽出して再利用することができないという問題がある。本稿では、電子メールを介したコミュニケーションにおいて用いられる言語の特徴を分析し、それに基づいて構成したパターンを用いたパターンマッチングによって、電子メールからイベントの開催日時や開催場所、期限付きの返信依頼等のスケジュール情報を抽出する方法を提案する。任意の電子メールを対象として、スケジュール情報の抽出を行った実験の結果、十分に実用に耐え得る程度の高い精度が得られたことを報告する。

Extraction of Schedule Information in Communication through E-mails

Takaaki HASEGAWA and Shin'ichiro TAKAGI
NTT Information and Communication Systems Labs.

hasegawa@isl.ntt.co.jp

E-mail is convenient because senders can communicate information to many people at one time. A problem is that receivers cannot extract and process the information within the e-mail automatically. We analyze the characteristics of the language used in e-mails and propose a way to extract schedule information, such as event date, event location, and RSVP date, from the text of e-mails by using pattern matching. Experiments using actual e-mails confirm the accuracy of our proposal; schedule information can be extracted from any e-mail with practical levels of performance.

1 はじめに

インターネットの爆発的な普及により、多くの幅広い層の人たちが電子メールを使ったコミュニケーションを行うようになってきている。多数の人に一度に情報を伝達できるという電子メールが持つ同報性により、電子メールはコミュニケーションの手段として、電話、FAX および手紙に取って代わる機会が増えている。例えば、電子メールを使ったコミュニケーションの中には、会議や打ち合わせの連絡や、講演会、パーティー、ゴルフの案内といった様々なイベントの通知が含まれる。さらに、それらに対する出席の可否等の返答や報告の提出などの送信の依頼を期限付きで送信することも多い。このように、電子メールを用いることによって、イベントの開催日時や開催場所に関する情報を伝達したり、その出席の可否について期限付きで送信を依頼したりすることを多数の人に同時に行うことができる。しかし、その同報性のために、送信する側では有効なコミュニケーションの手段であっても、受信する側にとっては、一人当りの受信する電子メールの数が飛躍的に増え、電子メールの利用者はいわゆる情報の洪水の中に身を置くことになる。多数のイベントのスケジュール情報を受信した利用者は、手帳や携帯情報端末などに備えられているスケジューラによって利用者自身の予定を一つ一つ確認し、受信したスケジュール情報を利用者のスケジューラに入力し直さなければならないだけでなく、要求されている期限までに忘れずにそれぞれの差出人へ出席の可否についての返信を送らなければならない。

さらに携帯情報端末や携帯電話等の無線通信のインフラが整いつつあり、電子メール利用者は外出先からでも電子メールを送受信することが可能な環境になっている。そのため、今や電子メールは単なる情報伝達だけではなく電話のような即時性を持つようになり、たとえ外出先であっても素早くコミュニケーションを行うことが要求されている。このような状況では、緊急に行われる打ち合わせのスケジュール情報を含んだ電子メールを受信していたとしても、多量の電子メールの山の中からその電子メールを即座に見つけることは困難であり、ましてや外出先では自分の手帳や携帯情報端末のスケジューラに入力されているスケジュールを調べて、打ち合わせの電子メールに対して出席の可否の電子メールを送信するのはかなりの負担となる。このように、電子メールを介したコミュニケーションは、送信側にとっては同報性及び即時性の点から都合が良いが、逆に受信側にとっては情報過多と情報の再利用の点から問題がある。そのため、電子メールを介したコミュニケーションに対するフィルタリングと人手を介さない情報の

再利用のための機構が求められている。

電子メールの情報を利用するには、まず中心となる情報を抽出しなければならない。電子メールは手紙に比べればインフォーマルなコミュニケーションとして使われるという側面から、電子メールの文章には口語的で非一般的な表現が用いられ、文法的誤り、漢字変換誤り、誤字脱字等のノイズが見られたりすることが多い。このため、一般的な自然言語処理に使われるような構文解析/意味解析を用いる方法では、未知語として処理されたり解析を誤ったりするケースが多くなるという問題が生じる。一方、パターンマッチングは文法誤り等のノイズに強いだけでなく、辞書を使わないために構文解析/意味解析に比べて処理が軽いので、受信する電子メールが多量にある場合に特に有効である。本稿では、電子メールを介したコミュニケーションにおいて用いられる言語の特徴を分析し、それに基づいて構成したパターンによるパターンマッチングのみによって、任意の電子メールの文章からイベントの開催時間や開催場所、期限付きの返信依頼等のスケジュール情報を抽出する方法を提案する。

本稿の構成は以下の通りである。第2節でスケジュール情報の定義と電子メールに見られるスケジュール情報の伝達表現を対象とした言語の特徴の調査結果について述べる。第3節で電子メールからのスケジュール情報を抽出するアルゴリズムを提案する。第4節でスケジュール情報を含む電子メールのフィルタリングと情報抽出の精度についての評価を報告する。第5節で本手法の有効性と課題を考察し、第6節で結論を述べる。

2 スケジュール情報の伝達

2.1 スケジュール情報の定義

人間同士のコミュニケーションの中には、同じ時間に同じ空間を他人と共有するという実世界のイベントへの参加がある。例えば、会議や講演会等がそうである。時間と場所が正確に伝わらなければ、同じ時間に同じ空間を共有できなくなるという意味から、実世界のイベントへの参加のためのアポイントメントは重要である。また、他人から期限付きで返信や提出等を依頼された場合、それをその期限までに実行することは、その人と人間関係を維持する上で重要である。そのため、依頼された行為とその期限は正確に伝わらなければならない。つまり、人間同士のコミュニケーションを維持するためには、イベントの開催日時と開催場所、依頼された行為、行為の期限を抽出することが重要である。本稿では、開催日時と開催場所を含むイベントの通知および受

Subject: インターネット講習会
 From: Hisashi OHARA <ohara@isl.ntt.co.jp>
 To: chigi@isl.ntt.co.jp
 Date: Wed, 12 Nov 1997 13:53:44 +0900

知技G各位

小原です。

12月10日(水) 15:00~16:30に104C会議室で
 インターネット講習会があります。

未受講の人は、やむを得ない事情がない限り出席せよとのことで
 す。

各位受講実績と参加の可否(否の場合は理由を添えて)を小原まで
 11/18日中にご回答ください。
 よろしくお願ひします。

図 1: スケジュール情報を含む電子メールの例

信者がコミットする返信や提出の依頼(以下コミット
 メントとする)をスケジュール情報と定義する。

多量に受信した電子メールから、スケジュール情
 報を含む電子メールだけを選択し、各項目の情報を
 抽出することが、電子メールを介したコミュニケーション
 にとって重要である。そのためには、受信した任
 意の電子メールについて、

1. スケジュール情報のフィルタリング
2. 個々の情報の抽出

が必要となる。

我々が日常受信している電子メールのうち、スケ
 ジュール情報を含んでいる電子メールの例として、図
 1を挙げる。図1の例では、発信者はインターネット
 講習会というイベントの開催日時と開催場所を通知
 すると同時に受信者を参加の可否等について回答す
 ることにコミットさせている。

スケジュール情報を通知する電子メールは、それ
 ぞれの電子メール毎にスケジュール情報の表現形態
 が異なるが、以下の情報を抽出できればコミュニケー
 ションを維持するためには十分であると考え。

● イベント情報

イベント名 開催通知の対象となるイベントの
 名前

開催日時 イベントが開催される日時や日程

開催場所 イベントが開催される場所

● コミットメント情報

行為種別 受信者がコミットする行為の種別

期限 受信者がコミットする行為の期限

表 1: イベントの日時の表現形態

分類	メール数	割合
ラベル化	47	78.3%
文章中	13	21.7%

表 2: イベントの通知の動作表現

表現	メール数	割合
ラベル化	4	8.2%
動作表現	52	86.7%
(内訳)		
開催します	22	36.7%
行います	6	10.0%
お知らせします	5	8.3%
決定しました	3	5.0%
あります	3	5.0%
実施します	2	3.3%
させていただきます	2	3.3%
開きます	2	3.3%
その他	7	11.7%
動作なし	4	6.7%

2.2 スケジュール情報伝達の表現

電子メールからスケジュール情報を抽出するには、
 まず電子メールにどのような形式でスケジュールが
 記述されているのかを掴む必要がある。そこで我々
 は、実際に受信した電子メール80通のうち、イベン
 トの通知を含んでいる60通を調べ、表現方法を分類
 した。まず、イベントの日時を記述する形式が「日
 時:」などによってラベル付きのアイテムイゼーシ
 ョンがなされているか否かに分類した結果を表1に示す。ま
 た、イベントの通知を示す表現についても分類した。
 この結果を表2に示す。動作表現のない例としては、
 「〇月〇日の△(イベント)ですが」、「ご案内しま
 した△(イベント)の件ですが」等の表現が見られた。

また、先に触れた同じ電子メール80通のうち、
 受信者が返信や提出にコミットする表現を含む48通
 についても分類を行い、ラベル付きのアイテムイゼー
 ションがなされているか否かについて調べた結果を
 表3に示す。コミットメントの種別の内訳と返信の依
 頼を表す表現を表4に示す。また、文章中に期限を示
 す表現のあるメールは24通あり、うち6通が「締切
 りは」などのように期限を明示的に表示している。
 期限表現の分類を表5に示す。

これらの表から、ラベル付きのアイテムイゼーシ
 ョンがなされずに文章中に埋め込んで記述している表
 現は、イベントの通知よりも受信者をコミットさせ
 る通知の方に多く見られることがわかる。電子メー
 ルを対象とした情報抽出においては、ラベル付きの

表 3: 受信者にコミットさせる表現形態

分類	メール数	割合
ラベル化	3	6.3%
文章中	45	93.8%

注: 46 通のうち 2 通がコミットメントを 2 つ含むため全体を 48 通として計算した。

表 5: 期限を表す表現

表現	メール数	割合
までに (迄に)	15	62.5%
中に	2	8.3%
至急	1	4.2%
締切り (期限)	6	25.0%

表 4: コミットメントの内訳と動作表現

表現	メール数	割合
返信	38	79.2%
(内訳)		
連絡下さい	16	(42.1%)
お知らせ下さい	5	(13.2%)
申し込み下さい	3	(7.9%)
回答下さい	2	(5.3%)
送って下さい	2	(5.3%)
報告願います	2	(5.3%)
教えて下さい	2	(5.3%)
メール下さい	2	(5.3%)
リプライ下さい	2	(5.3%)
その他	2	(5.3%)
提出	10	20.8%

表 6: ラベルを検出するキーワードの例

開催日時	開催場所	期限	イベント名
日時	場所	締切り	について
日程	会場	申込期限	のお知らせ
とき	ところ	〆切	のご案内

析、形態素解析を行わず、イベントの通知の動作表現、および受信者がコミットする動作表現等で構成したパターンを用いたパターンマッチングによる手法を採る。

3.2 情報抽出の方法

3.2.1 ラベルを表すキーワードの利用

イベントやコミットメントのスケジュール情報を通知する電子メールの中には、受信者が見やすいようにラベル付けされたアイテム化がなされている情報があることが多い。開催日時や開催場所、期限を表すキーワードは「日時」「場所」「期限」等いくつか見られる。また、イベント名を表すキーワードとして、「について」「のお知らせ」「のご案内」等が見られる。そこで、電子メールのテキストに対するこれらのキーワード検索によってアイテム化がなされたラベルを検出し、ラベルに続く文字列をイベントやコミットメントのスケジュール情報として抽出する。ラベルを検出するためのキーワードの例を表 6 に示す。

3.2.2 言語の特徴からのパターン構成

実際に受信した電子メールのテキストから言語の特徴を抽出した結果に基づいて、文章中对して適用するパターンマッチングに用いるパターンを構成する。図 2 にパターン例の一覧を示す。パターンの中にはマッチした文字列を格納する変数が存在する。これらの変数は以下の通りである。

?date_pat 年月日

?time_pat 時刻

?place 開催場所

アイテム化がなされている情報のもとより、文章中に埋め込まれている情報も抽出できなければならない。

3 スケジュール情報の抽出

3.1 情報抽出の方針

情報抽出とは、テキストから中心となる情報だけを取り出すことである。情報抽出の手法としては、構文解析/意味解析を行う手法や定型的な表現からなるパターンを用いる手法がある [3]。スケジュール情報を含む電子メールテキストにおいて中心となる情報は、ラベル付きのアイテム化がなされている情報だけでなく、文章中にも埋め込まれている情報も含まれる。電子メールはコミュニケーションの手段であるので、イベントの通知の動作表現や受信者にコミットさせる動作表現を、言語行為 [5] と捉えれば、それぞれ主張型、要求型の言語行為として分類できる。それゆえ、文章中のこれらの表現、すなわち、イベントの通知を表す動作表現、受信者がコミットする動作表現を見つけることによって、イベントの通知や受信者がコミットする行為を含むスケジュール情報を抽出することが可能であると考えられる。以上の理由から、電子メールからのスケジュール情報を抽出するために、我々は構文解析/意味解

?event_name イベント名

event_pat ≡ {start_pat + (end_pat) + (place_pat)
 + object_pat + inform_act_pat,
 object_pat + {start_pat + (end_pat) + (place_pat)
 |(reference_pat))} + inform_act_pat,
 start_date&time + (place_pat) + の
 + ?event_name + (の件) + {について|(です) + が
 |の + {お知らせ|(ご) + 案内} + (メール) + です},
 start_date&time + は + ?event_name + (の日)
 + {です|{と|に} + なって},
 reference_pat + ?event_name + {について + (の)
 |の} + {お知らせ|(ご) + {案内|連絡}} + {(メール)
 + です|{を} + {致|いた} + します}}

inform_act_pat ≡ {開催|({致|いた}) + {します
 |したい}|する}, 催し|{たい|ます}, お知らせします,
 あります, 開きます}

request_act ≡ {{{(ご|御)}} + {回答|連絡|報告}
 + (を)|{お知らせ|教えて|メール|リプライ}}
 + {{{(して) + {下さ|くださ|頂|いた|だ}}|{願|い
 |ねが|い} + (ます)}}

start_pat ≡ {?start_date&time
 + ({から|~|より|スタート|~})}

end_pat ≡
 {?end_date&time} + ({まで|に|にかけて})}

place_pat ≡ {place_pat1|place_pat2}
 place_pat1 ≡
 {?place + {におきまして|において|にて|で}}
 place_pat2 ≡
 {{場所|会場} + {は|については} + ?place
 + {で|を|と|に}}

object_pat ≡ {?event_name + {を|について|が}}

time_limit_pat ≡
 {time_limit_pat1|time_limit_pat2}
 time_limit_pat1 ≡ {?time_limit + {{まで|迄|中}
 + (に)|を + {締|〆} + (め) + {切|き}}}
 time_limit_pat2 ≡
 {{締め切り|〆切|〆きり|期限} + {は|が|を}
 + ?time_limit + (まで) + {と|です|に}}

reference_pat ≡ {下記の|以下の} + {通り|と|おり
 |{要領|日程|次第} + {で|にて}}

?start_date&time ≡ {?date_pat + (?time_pat)}

?end_date&time ≡
 {{{?date_pat + (?time_pat)|?time_pat}}

?time_limit ≡ {{{?date_pat + (?time_pat)
 |(?date_pat) + ?time_pat}}

図 2: パターン例の一覧

ただし、?date_pat, ?time_pat は、日時を表現するための、記号や数字および特定の文字からなるパターンとして記述することにより、年月日や時刻を格納することができる。?place と ?event_name は任意の文字列を格納するが、文字列中に読点「、」等を含む場合は、それ以前を削除した文字列を格納する。図 2 に示すパターンを用いたパターンマッチを行うことにより、イベントやコミットメントのスケジュール情報を抽出する。ただし、() は省略可能で、{a|b} は括弧内の a, b のうちいずれかを使うものとする。+ は文字列同士の結合を示す。

3.2.3 抽出アルゴリズム

電子メールのテキストからスケジュール情報を抽出するアルゴリズムについて、以下に示す。

1. 電子メールのテキストをラベル付きのアイテムイゼーションがなされている部分とそうでない通常の文の部分に分ける。
2. アイテムイゼーションの部分についてはラベルの出現位置や空行の有無を用いてアイテムイゼーションの部分と通常の文の部分については句点等を基にして一文単位に区切る。
3. アイテムイゼーションがなされているラベルに日時、場所や期限を表すキーワードがあれば、そのキーワードを基にして開催日時、開催場所、イベント名、期限を抽出する。
4. 通常の文の部分においてパターン event_pat にマッチするならば、開催日時、開催場所、イベント名を抽出する。
5. 文章中においてパターン request_pat や time_limit_pat にマッチするならば、コミットメントの有無とその期限を抽出する。

パターンマッチング処理の原則は、下記の通りである。

- 文単位でマッチングを行い、パターンが複数あるときは、前の方からパターンマッチをそれぞれ試みる
- 文に一つのパターンがマッチした場合、残っているパターンでパターンマッチを行わない
- パターンマッチは place_pat を除いて最長マッチングとする
- 読点を含む文字列にマッチングした場合、読点より前の文字列は削除する

3.2.4 具体例

上述したアルゴリズムを用いて、図1の電子メールを例に取り、スケジュール情報の抽出を行う具体例を示す。

- ラベル付けされたアイテムイゼーションがないため、文章中から一文ずつ区切った文からパターンマッチにより抽出する。
- start_pat* が日時表現と「～」に、*end_pat* が日時表現と「に」に、*place_pat* が「で」に、主張型言語行為 *inform_act* が「あります」に、*object_pat* が「が」にそれぞれマッチするので、*event_pat* の最初のパターンに一文「12月10日(火) 15:00～16:30に104C会議室でインターネット講習会があります。」がマッチする。
- start_pat* に対応する「12月10日(火) 15:00」を開始日時として、*end_pat* に対応する「16:30」を終了日時として抽出する。
- place_pat* に任意の文字列と「で」がマッチするので、*place* として「104C会議室」を開催場所として抽出する。
- object_pat* が任意の文字列と「が」にそれぞれマッチするので、*event_name* にイベント名として「インターネット講習会」を抽出する。
- 要求型言語行為 *request_act* が「回答ください」に、*time_limit_pat* が日時表現と「中に」にマッチするので、*request_act* と *time_limit_pat* が一文「各位受講実績と参加の可否(否の場合は理由を添えて)を小原まで11/18日中にご回答ください。」がマッチする。
- コミットメントを有りにして、*time_limit* に対応する「11/18日」を期限として抽出する。

図1に示した電子メールからの抽出結果を表7に示す。ただし、後述する情報の補完を行っている。

3.3 情報の補完

電子メール本文のみでは完全な日時の情報を得られない場合も少なからずある。スケジュールの管理を考えた場合、完全な日時情報を得られないことは都合が悪い。このような場合は、電子メールのヘッダ情報や我々の持っている常識から情報を補完することが有効であると考えられる。

表 7: 抽出結果例

抽出項目	抽出内容
イベント名:	インターネット講習会
開始日:	12/10/1997
開始時間:	15:00
終了日:	12/10/1997
終了時間:	16:30
場所:	104C会議室
行為種別:	返信
期限:	11/18/1997

表 8: 日付変換テーブル

相対表現	絶対表現
本日, 今日	送信日付
明日	送信日付+1日
今週(の)W曜日	送信日付の週のW曜日の日付
来週(の)W曜日	送信日付の次週のW曜日の日付

電子メールを書く人や書く状況によっては、イベントの開始日が完全でないことがある。例えば、同じ月の中でのイベントを通知する場合は、年や月は省略されることがある。このような場合には、ヘッダ情報の送信日時を表す Date: ヘッダを参照することによって、年や月を補完する。また図1のようにその日1日で終了するイベントの通知には、通常終了日は明示的に書かれていない。このような場合は、終了日を開始日と同一日で補完することが必要である。

また、絶対的な日時の表現だけでなく、相対的な日時の表現が用いられる場合もある。例えば、「本日」「明日」「今週」などの表現がそうである。このような相対的な日時の表現にマッチングした場合も、表8に示す日付の置換の一例に基づいて、ヘッダ情報から情報を補完して絶対的な表現に置換する。

4 評価

本稿で提案したアルゴリズムを計算機上に実装した。OSはSunOS4.1.4、使用言語はperl5.003である。電子メールからのスケジュール情報抽出の実験結果を評価するに当たって、以下の2つの基準に分けて評価を行った。

- スケジュール情報のフィルタリングの判定精度
- 各項目毎の抽出情報の抽出精度

1番目の精度についてはスケジュール情報を含む電子メールの検出の再現率と適合率で精度を表す。電子

表 9: スケジュール情報抽出の精度

項目	既知該当数	正解数 (再現率)	抽出数 (適合率)	未知該当数	正解数 (再現率)	抽出数 (適合率)
イベント	60	55(91.7%)	55(100%)	53	46(86.8%)	46(100%)
開始日	60	57(95.0%)	57(100%)	53	47(88.7%)	47(100%)
終了日	9	8(88.9%)	9(88.9%)	10	5(50.0%)	9(55.6%)
開始時刻	53	49(92.5%)	50(98.0%)	43	35(81.4%)	36(97.2%)
終了時刻	29	26(89.7%)	27(96.3%)	27	22(81.5%)	25(88.0%)
開催場所	60	54(90.0%)	55(98.2%)	53	46(86.8%)	48(95.8%)
イベント名	60	53(88.3%)	58(91.4%)	53	41(77.4%)	50(82.0%)
コミットメント	46	46(100%)	46(100%)	121	115(95.0%)	117(98.3%)
期限	25	23(92.0%)	23(100%)	44	34(77.3%)	37(91.9%)

メールがスケジュール情報を含むか否かの判断は、電子メール中に何らかのイベントの開催日時と開催場所が含まれる、もしくは受信者に対する返信や提出といったコミットメントが含まれるか否かでを行った。開催日時は開始日、開始時刻、終了日、終了時刻から構成され、少なくとも開始日を含むとしている。2番目の精度については、スケジュール情報を含むと判定された電子メールから抽出された情報の正解率で表すことにする。情報抽出アルゴリズムを規定するために調べた既知の電子メール 80 通と無作為に抽出した未知の電子メール 319 通について、スケジュール情報抽出の実験を行った。イベントとコミットメントのフィルタリングと各項目の抽出情報についての再現率と適合率の結果を表 9 に示す。開始日、期限の日時表現についてはヘッダ情報から補完したり、相対表現から絶対表現へ置換した情報を用いて正解かどうかの判断を行った。

5 考察

5.1 議論

スケジュール情報を電子メールのフィルタリングやスケジューラへの入力として用いることを想定すれば、本稿で主張した手法による電子メールからのスケジュール情報抽出は高い精度を示しており、十分実用に耐えるものである。パターンマッチングを用いているので、電子メールの文章に見られる様々な表現を数多くパターンに取り入れていけば、精度は現状よりもさらに向上すると考えられる。

5.1.1 未解決な問題

イベントの方がコミットメントに比べると再現率が低い、これは開催日時と開催場所を両方抽出できないとスケジュールとは見なさないため、開催場所を抽出するのに失敗しているからである。開催

日時に比べると開催場所の抽出は数字や記号による表層のパターンマッチングが使えないため数段難しい。それに加えて、開催場所の情報が言語の特徴を付随せずにそれ単独で現れる例や、日時のラベルの中に場所の情報が併記されている例が数例見られ、これが再現率の低下に繋がったと考えられる。また、終了日の正解率が低い原因として、複数のイベントが併記されていることや繰り返しのイベントが記述されていることが挙げられる。連続した期間の表現は抽出できるが、独立して記述された期日の表現は抽出できなかった。これについては、パターンの表現方法やアルゴリズムを見直すことによって対応可能であると考えている。

5.1.2 今後の課題

本稿では、スケジュール情報のあるイベントの開催日時と開催場所を含むか、あるいは受信者が返信や提出にコミットする行為を含む情報と定義した。しかし、実際に多くの電子メールを見ると、これでは不十分であることがわかった。暗黙の了解から省略されたか電子メールのやりとりの過程で既に伝達されていたかと考えられる、開催場所のないイベントの通知や、開催場所は未定であるが、時間だけは確保するように依頼する通知もあった。これらの開催場所が書かれていない電子メールに対処するには、抽出したイベント名から判定するイベントのカテゴリを使うことが有効であると考えている。開催場所がないイベントについては、カテゴリを利用することによってスケジュール情報か否かを判別していきたい。また本稿ではコミットメントについては、返信や提出以外の行為は抽出しないことにしたが、これはネットワーク上での電子メールのやりとりで情報伝達が完結することを仮定しているからである。コミットメントについても今後さらに多くの種別を扱っていきたい。

5.1.3 方式の有効性

ネットワークを介してイベントやコミットメントのスケジュール情報を伝達する方法として、組織内部のイントラネットにおいては、グループウェアのためのソフトウェアを使う方法もある。この場合、スケジュール情報を特定の形式で送信し、受信側では特定の形式の情報をスケジュール情報として表示することになる。しかし、この方法は送信側と受信側が同じグループウェアのソフトウェアを使用していることが前提になる。また、電子メールの内部構造に制限を与えたり、タグを付けたりする方法は、受信側にとっては便利ではあるが、送信側にとっては自由に記述することが制限されるため、自然言語を用いたメッセージの伝達というせつかくの利点を生かすことができない。本稿で提案した手法は、テキストファイルを入力するので、上述の制限を受けることがなく、組織の内部外部ともに隔たり無くイベントやコミットメントのスケジュール情報を伝達することが可能である。

5.2 関連研究

情報抽出の研究開発では、従来より新聞記事を対象にした国際会議 MUC[3]がある。しかしながら、計算機やネットワーク環境が発達するにつれて、新聞記事等の文語的な言葉で記述されている文章の他に、電子メールやネットニュース、ホームページ等のより口語的な言葉で記述され得る文章に接する機会が増えてきており、今後はこれらを対象とした情報抽出の研究開発が益々重要になると考える。

形態素解析を行わずにパターンマッチングにより情報を抽出するという、我々が採用した手法を用いた研究に、ネットニュース記事のダイジェスト自動生成を行う研究がある[4]。会議の告知や論文募集に関する記事を対象にしており、これらの記事の中心的な情報を抽出する際に、パターンマッチングのみを用いている。これらの記事の特徴は、かなり意図的に定型化されている文章であったり、目的がはっきりしている文章であることにある。本稿で対象としているテキストは、利用者が日常受信するあらゆる内容や意図を有する電子メールである。利用者へ通知されるイベント情報や利用者へコミットされる行為といった特定の情報に着目し、それらの情報を含む電子メールを抽出することは、一種のフィルタとしても機能していることに等しい。

電子メールを対象にした研究では、電子メールに付与された署名情報を抽出する研究がある[1]。電子メールに署名があるかどうかを文末表現のパターンマッチングを行うことによって検出すること、署

名情報からレイアウトを考慮したパターンマッチングを行うこととによって、送信者の名前、所属、住所、電話番号等を抽出するものである。ヘッダ情報を利用することは共通しているが、電子メールの文書内容までは対象にはしていない。またスケジュール調整に関する研究[2]では、利用者の都合を考慮してネットワーク上で自動的にスケジュールの調整を行うシステムについて提案されている。調整のためのメッセージは電子メールで行われるが、すべての情報にタグが付けられており、自然言語が用いられたメッセージの交換ではない。

6 おわりに

本稿では、イベントやコミットメントのスケジュール情報を伝達している電子メールの言語の特徴に着目し、それに基づいたパターンを用いたパターンマッチングによって、任意の電子メールからイベントやコミットメントのスケジュール情報を含む電子メールだけを選別し、細部の情報である開催日時や開催場所、期限等を高い精度で抽出する方法を提案した。既知と未知の電子メールを用いて行った実験の結果から、利用者が受信する多量の電子メールの中から、利用者が行うコミュニケーションにおいて重要なスケジュール情報を含む電子メールのみを高い精度でフィルタリングし、抽出された詳細情報はスケジュールへの入力として用いることに対して実用に耐え得る程の高い正解率を有することを実証した。

今後は、個人情報を用いた知的な行動選択の支援や、過去の履歴を考慮した連続的なコミュニケーションの支援に取り組んでいく。

参考文献

- [1] 浅野 久子, 大山 芳史: 電子メールからのパーソナル情報抽出方法の検討, 第52回情報処理学会全国大会 4J-4, 1996.
- [2] Thomas Haynes, Sandip Sen, Neeraj Arora and Rajani Nadella: An Automated Meeting Scheduling System that Utilizes User Preferences, In *proceedings of the First International Conference on Autonomous Agents (Agents'97)*, pp.308-315, 1997.
- [3] *Proceedings Sixth Message Understanding Conference (MUC-6)*: Advanced Research Projects Agency Software and Intelligent Systems Technology Office, Morgan Kaufmann Publishers, Inc., 1995.
- [4] 佐藤 円, 佐藤 理史, 篠田 陽一: 電子ニュースのダイジェスト自動生成, 情報処理学会論文誌 Vol.36, No.10, pp.2371-2379, 1995.
- [5] Terry Winograd and Fernando Flores: コンピュータと認知を理解する. 産業図書, 1989.