電子メールとスケジュールの関連性を考慮した情報閲覧システム

高橋　悟史† 　宮前　雅一†† 　寺田　努† 　西尾章治郎†

† 大阪大学大学院情報科学研究科　〒 565–0871 大阪府吹田市山田丘 1–5
†† 株式会社国際電気通信基礎技術研究所知識科学研究所　〒 619–0288 京都府相楽郡精華町光台 2–2–2
E-mail: †{takahashi.satoshi, tsutomu, nishio}@ist.osaka-u.ac.jp, ††miyamae@atr.jp

あらまし　近年，計算機の普及に伴い，電子メールやスケジュール等の個人情報を計算機で管理することが日常的になっている．一般的に，スケジュールに関係する事柄をメールで交換するなど，メーラが管理する情報とスケジューラが管理する情報には深い関連がある．しかし，既存のアプリケーションではこれらの情報間の連携が取られていないため，スケジュールの詳細を知りたいときにメーラ上で関係のあるメールを手動で検索するなど，情報の把握に煩雑な操作が必要である．そこで本研究では，メールとスケジュールの関連性を考慮した新しい情報提示手法を提案する．提案手法では，メールやスケジュールの表示時，関連情報を同時に提示し，さらに関連情報を選択することで情報の切り替えをシームレスに行えるようにする．これにより，関連情報の閲覧が容易になる．
キーワード　メール，スケジュール，関連情報，情報抽出，情報提示

An Information Browsing System Considering the Relationship between E-mails and Schedules

Satoshi TAKAHASHI†, Masakazu MIYAMAE††, Tsutomu TERADA†, and Shojiro NISHIO†

† Graduate School of Information Science and Technology, Osaka University
Yamadaoka 1–5, Suita-shi, Osaka, 565–0871 Japan
†† Knowledge Science Laboratories, Advanced Telecommunications Research Institute International
Hikaridai 2–2–2, Seika-cho, Soraku-gun, Kyoto, 619–0288 Japan
E-mail: †{takahashi.satoshi, tsutomu, nishio}@ist.osaka-u.ac.jp, ††miyamae@atr.jp

Abstract　In recent years, due to the popularization of personal computers, it becomes popular to manage personal information such as E-mails and the schedules. Generally, there is a close relationship between the information in E-mails and the information in schedules, such that the information related to the schedule is exchanged by E-mails. However, existing applications do not consider the cooperation between multiple types of information, therefore, complex operations are necessary to reach the intended information such that it is necessary to retrieve E-mails manually to get the detail of the schedule. Therefore, we propose a system to browse information with referential information. Our method not only displays the referential information in displaying E-mails and schedules but also enables users to switch information seamlessly. Using our system, users can browse the referential information easily.
Key words　E-mail, Schedule, Referential Information, Information Extraction, Information Browsing

1. はじめに

近年，計算機の普及に伴い，電子メールやスケジュール等の個人情報を計算機で管理することが日常的になってきている．例えば，電子メールは携帯端末の普及も貢献し，多くのユーザにコミュニケーションの手段として利用されており，ユーザはメーラを用いて交換された多量のメールを管理している．また，スケジュールも追加や修正が頻繁に行われ，管理が煩雑であることから，計算機を用いて管理することが有効である．特に近年は，メールやスケジュール，アドレス帳などを一括して管理できる PIM(Personal Information Manager) ソフトウェアやグループワークの支援を目的とし，ユーザ同士でスケジュールや ToDo リストなどを共有できるグループウェアの普及に伴い，高度な機能をもつスケジューラを利用することが一般的になりつつある．

しかし，既存のメーラやスケジューラを用いた管理方法には

詳細情報の把握が困難であるという問題点がある．これは一般的に，スケジュールに関係する事柄はメールで交換されることが多く，ユーザはスケジューラに最小限の情報のみを入力し，詳細情報をメーラで保管しておくことが多いためである．したがって，メーラが管理する情報とスケジューラが管理する情報には深い関連があるが，既存のアプリケーションではこれらの情報間の連携が取られていないため，スケジュールの詳細を知りたいときにはメーラ上で関係のあるメールを手動で検索しなければならない．同様に，メールでスケジュールの変更が通知された場合，該当するスケジュールをスケジューラの中から検索する必要があるが，大量のスケジュールの中から必要な情報のみを検索するのには非常に手間がかかる．このように，既存のメーラやスケジューラを個別に用いた場合，情報の把握に煩雑な操作が必要である．また，スケジュールを入力するのに手間がかかるという問題点もある．前述のように，スケジュールに関係する事柄はメールで交換されることが多いが，スケジュールを通知するメールを受信した場合，スケジュール情報を取り出して定型化し，スケジュールを入力するという作業を行う必要がある．しかし，これらの作業はユーザが手動で行わなければならない．

このように，既存のメーラやスケジューラには関連情報の把握や情報入力に煩雑な手間が必要であるという問題がある．これらの問題は，メールとスケジュールを管理する際，その内容から情報間の関連性や情報内に存在するスケジュールの記述を抽出できれば解決する．例えばメールを受信したとき，システムがメールの内容から関連しているスケジュールを自動的に検索して表示できれば，ユーザは容易に関連情報を把握できる．同時に，メール内にあるスケジュールの記述を抽出すれば，その情報を元に自動的にスケジューラにスケジュールを入力することができる．

そこで本研究では，メールとスケジュールの関連性を考慮した新しい情報閲覧システムである RI ブラウザ (Referential Information Browser) を提案する．提案システムでは，メールやスケジュールに関連する情報を自動的に検索し，情報提示の際に関連情報を同時に提示する．このとき，関連情報を表示するだけでなく関連情報を選択することでメールからスケジュール，スケジュールからメールへの情報の切り替えをシームレスに行えるようにする．また，メールにスケジュール記述がある場合は，スケジュール情報を抽出することでスケジュールの入力支援を行う．提案システムを用いることで，情報を効率的に探索，閲覧できるようになる．

以下，2. 章で関連研究について述べ，3. 章でシステムの設計について説明する．4. 章では実装について述べ，5. 章で評価と考察を行い，最後に 6. 章で本研究のまとめを行う．

2. 関 連 研 究

これまでに，メールやスケジュールなどの個人情報の管理を支援するための研究は多く行われている．特にメールからの情報抽出に関して様々な研究が行われており，メールマガジンサービスから配信されたニュース情報から注目ニュースを抽出する手法がある [7]．この手法では，ニュースの見出しの類似性に基づいて配信記事群をグループ化し，各グループ内の記事数などから話題性の高い注目ニュースを抽出している．しかし，この手法はメールニュースのような定型化されている文書で送られてくるメールにおいては有効であるが，知人間のメールのように文字の誤りや略語など表記のゆれが大きい文書への適用は困難である．

メールと PIM ソフトウェアとの連携を考慮した情報抽出手法としては，メールから日時・場所などのスケジュール情報を自動抽出する手法が提案されている [3]．この手法では，言語特徴から求めた表現パターンとのパターンマッチングを用い，メール内のスケジュールや ToDo を抽出している．また，電子メールに付加されている署名から送信者の氏名や電話番号などの住所録情報を自動抽出する手法なども提案されている [1]．実際のアプリケーションとしては，feedpath 社の feedpath Zebra [2] がある．feedpath Zebra はメールクライアント，スケジューラ，アドレス帳，Wiki などの機能を統合した Web サービスだが，メール内にある日付情報を選択することで，その日のスケジュールを表示したり，その日のスケジュールを新規作成できる機能をもつ．このように，メール単体における情報抽出および抽出した情報を単純に用いて他の情報と連携させる取り組みは行われているが，情報間の類似性や関連性を考慮した情報管理は行われていない．

また，情報間の類似性や関連性を考慮している研究として，過去のメールとの類似性からグループ分けを行い，返信文の下書きを自動的に作成する手法が提案されている [4] ．この手法では，メールの特徴的な単語を利用することで過去の履歴データとの類似度を計算し，類似しているメールをグループ分けしている．一方，個人スケジュールを共有し，ワークグループに所属しているユーザ同士のコミュニケーションを支援するグループウェアカレンダーシステムである Augur [6] では，共通するスケジュールをスケジュールから抽出した特徴的な単語を用いて検索している．これらの研究でもメール同士やスケジュール同士の類似性をみているが，メールとスケジュール間の類似性や関連性を考慮した管理や定時はされていない．

そこで本研究では，メールとスケジュールに対してそれぞれ特徴抽出を行い，メールとスケジュール間の関連性を考慮する手法を提案し，提案手法を用いた情報管理システムを構築する．

3. システムの設計

1. 章で述べたように，既存のメーラやスケジューラを用いた管理には以下のような問題点がある．

- 詳細情報の把握が困難

一般的に，メーラが管理する情報とスケジューラが管理する情報には深い関連があるが，既存のアプリケーションではこれらの情報間の連携が取られていない．そのため，スケジュールの詳細を知りたいときにはメーラ上で関係のあるメールを検索しなければならないが，大量のスケジュールの中から必要な情報のみを検索するのは困難な作業である．

- スケジュールの入力に手間が必要

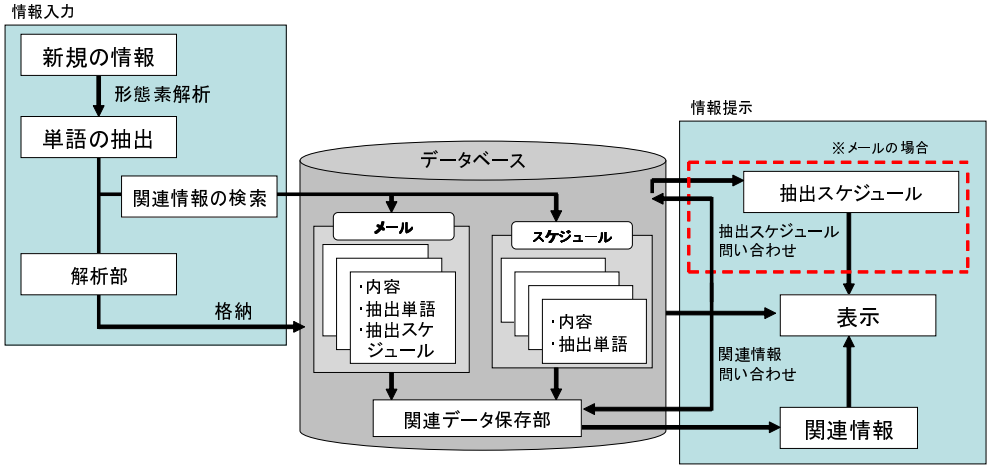

図 1　Referential Information Browser の概要

スケジュールに関係する事柄はメールで交換されることが多いが，スケジュールを通知するメールを受信した場合，スケジュール情報を取り出して定型化し，スケジュールを入力するという作業をユーザが手動で行わなければならない．

以上のような問題を解決するために，本研究では新たな情報閲覧システムである RI ブラウザ (Referential Information Browser) を提案する．RI ブラウザは下記の機能をもつ．

- 関連情報の自動検索

関連情報を検索する手間を軽減するために，システムが自動的に関連情報を抽出する．そのために，メールやスケジュールが入力された際，その内容を解析することで関連情報を検索する．また，あるメールに関して返信されたメールは，返信元のメールと共通の内容であることが予測されるため，返信元のメールと関連しているスケジュールは，返信されたメールでも関連している考えられる．このように，メールとスケジュールの内容だけでなく，返信関係なども考慮に入れて関連情報を検索する．

- 関連情報の提示

メールやスケジュールを閲覧する際に，情報の詳細や関連情報を容易に得られるようにするため，メールやスケジュールに関係する情報を同時に表示する．また，メールでスケジュールの時間変更が通知された場合など，変更先に既にスケジュールが存在するかを確認しなければならないときには，メールとスケジュール一覧を比較する必要がある．そこで，関連する情報を同時に表示するだけではなく，ユーザがその情報を選択することで情報の切り替えをシームレスにおこなえるインタフェースを提供する．

- スケジュールの入力支援

ユーザのスケジュール入力を支援するため，メールに記述されたスケジュールを自動的に検出し，それを元にスケジュールの入力を行えるようにする．具体的には，メールの内容を解析し，メールにスケジュール記述がある場合にスケジュール情報を抽出する．ユーザがメールを閲覧する際に，抽出された情報を提示し，スケジューラへの入力支援を行う．

以上で述べた機能をもつシステムを構築するために，RI ブラウザでは，メールやスケジュールの内容を解析して得られた結果や，返信関係の情報を用いて関連する情報を検索し，関連情報の表示およびシームレスな情報の切り替えを実現する．また，メール内に記述されたスケジュール情報を抽出することにより，スケジュールの入力支援を行う．

RI ブラウザの概要を図 1 に示す．RI ブラウザでは，受信したメールやユーザにより入力されたスケジュールを解析して，その情報の特徴的な単語を抽出する．関連情報の検索では，抽出した特徴的な単語と各情報がもつその情報の特徴的な単語との比較や返信情報から検索を行い，検索結果を関連データ保存部に保存する．また，解析部ではメールの内容を解析することでメール内のスケジュール記述の抽出などを行う．RI ブラウザでは各情報をデータベースに格納する際，各情報の抽出単語およびメールであれば抽出したスケジュール記述を付加し，データベースに格納する．情報を表示する際は，データベースの関連データ保存部からその情報の関連情報を取得し同時に表示する．また関連情報を表示するだけでなく，選択することで簡単にその情報を参照できるようにしている．

3.1　関連情報の検索

本節ではメールとスケジュールの関連情報の検索方法について詳細に述べる．

3.1.1　メールのもつ関連情報の検索

メールの関連スケジュールは以下のスケジュールとする．

（1） 文書内の単語が類似しているスケジュール

メールと関連があるスケジュールは，その内容が似ているため出現単語も類似していると考えられる．例えば，ミーティングやソフトボール大会の通知をするメールでは，「ミーティング」や「ソフトボール大会」という単語がそのメールを特徴的に表している．同様に，「ミーティング」や「ソフトボール大会」といった単語を特徴的な単語としてもつスケジュールやメールは，同じくミーティングやソフトボール大会に関する話題についての情報だと予測できるので，何かしらの関連があると考えられる．そこで，メールから抽出した特徴的な単語と，各スケジュールの特徴的な単語を比較し，一致する単語があれば関連性があるとする．提案手法では，特徴的な単語を抽出するため

に，まず形態素解析を行うことでメールおよびスケジュールの文書を単語に分解し，一般名詞，固有名詞，サ変接続名詞を抽出する．一般名詞やサ変接続名詞の単語は「ミーティング」や「忘年会」,「説明」や「受付」といったスケジュールを表す単語が多く，固有名詞は，スケジュールに関係する地名や人名を表すことが多い．このように，一般名詞や固有名詞，サ変接続名詞はその話題を特徴付ける単語になりやすいため，これらの単語を特徴的な単語の候補として抽出している．一方，接続詞や助詞などは特徴的な単語にはなりにくいため，今回は利用しない．抽出した単語からの特徴的な単語の決定には，キーワード抽出において一般的に用いられる *tf*idf* 法を用いる．*tf*idf* 法では，対象となる文書においてその単語がどれくらい特徴的かを *tf*idf* 値により表す．*tf*idf* 値は以下の式により算出する．

$$tf(t,d) * idf(t) = tf(t,d) * idf(t)$$

$tf(t,d)$(*normalized within-document frequency*) は，単語 t がメール d 内に現れる頻度を示す．$idf(t)$(*inverse document frequency*) は，データベース内のメールとスケジュールの全情報において単語 t が現れる情報数を $df(t)$ とし，次式で定義される．

$$idf(t) = \log \frac{df(t)}{N} + 1$$

得られた *tf*idf* 値は，$tf(t,d)$ 値が高く，$df(t)$ 値が低い単語が特徴的と判別される．即ち，文書中に出現する回数が多く，全情報における出現頻度が少ない単語が特徴的な単語として選出される．提案手法では，10 個の特徴単語を抽出し，そのうち 2 個以上の単語が一致する単語をもつスケジュールはそのメールの関連情報となる．この 2 個という閾値はユーザにより変更可能になっており，必要な情報が取得できないようであれば閾値を低く，余分な情報が多いときには閾値を高くすることで関連情報の量を調整できる．また，特徴的な単語として選出される単語には，一般名詞や固有名詞，サ変接続名詞以外にもユーザ特有の略語や名称を含む単語が存在することが考えられる．そこで，システムでは判断できない各ユーザにとっての特徴的な単語を追加登録や削除ができるようにしている．

（2） 返信関係にあるメールの関連スケジュール

一般に 返信関係にあるメールは同じ話題に関する内容であることから，返信関係にあるメールの関連スケジュールとの関連性も高いと考えられる．そこで，提案手法では返信関係にあるメールの関連スケジュールも関連情報とする．これにより，あるスケジュールの変更通知に対する返答のみで返信されたメールがあった場合，出現単語数が少なく単語の類似性だけでは関連性がないと判断されても，返信関係にあるメールが関連スケジュールとしてもっていれば，関連のある情報として認識できる．返信関係にあるメールは，メールヘッダの Message-Id: フィールドおよび，References:フィールドを用いて特定する．

（3） 文書内の単語が類似しているメールの関連スケジュール

返信関係にあるメール以外にも，特徴単語が類似しているメールは関連性が高いと考えられる．そこで，提案手法では単語類似度の高いメールの関連スケジュールも関連情報とする．(1) と同様の手法を用いて類似メールを抽出するが，メールはスケジュールと異なり本文に含まれる単語が多いことから，初期設定では特徴単語 10 個のうち，4 つ以上の単語が一致する場合に関連性があるメールと判断する．この閾値もユーザにより変更可能にしている．このようにして得られたメールの関連スケジュールをそのメールの関連スケジュールとする．

以上のようにして得られたメールとスケジュール間の単語の一致数やメールの返信関係，メール間の単語の一致数などの情報間の関連性に関するデータは関連データ保存部に保存される．情報提示を行う際は，関連データ保存部にあるデータに基づき関連情報を確認し表示を行う．

3.1.2 スケジュールのもつ関連情報の検索

スケジュールの関連メールとは，以下のメールである．

（1） 文書内の単語が類似しているメール

スケジュールと関連するメールは，その内容が似ているため，出現する単語も類似していると考えられる．提案手法ではメールから関連スケジュールを導出した場合と同様に *tf*idf* 法を用いて関連メールを抽出する．

（2） スケジュールの作成元になったメール

前述のように提案システムでは，ユーザによるスケジュール入力作業を支援するため，メールからスケジュール記述が抽出された場合，得られたスケジュール情報をスケジューラに追加する．このとき，抽出元となったメールはそのスケジュールに関する内容であることから，提案手法では，そのスケジュールの作成元となったメールを強い関連をもつメールとする．

（3） スケジュールの作成元になったメールと返信関係にあるメール

一般的に返信関係にあるメールは，同じ話題に関する内容であることから，前述のようなスケジュールの抽出元となったメールと返信関係にあるメールとの関連性も高いと考えられる．そこで，提案手法ではスケジュールの作成元になったメールと返信関係にあるメールも関連情報とする．

以上のようにして得られた情報間の関連性に関するデータはメールの場合と同様に関連データ保存部に保存され，情報提示の際に利用される．

3.2 情報の提示方法

RI ブラウザでは，ある情報が表示されているとき，その情報の関連情報を同時に表示することで詳細情報の確認の手間を軽減する．ここで関連情報はメールの受信時間やスケジュールの時間が近いものほど強い関連性をもつと考えられる．例えば，ミーティングのように定期的に繰り返されるスケジュールは，何度もメールでスケジュールのやりとりやされているが，実際に関連性がある情報は，期間的に近いものに限られる．そのため，関連情報を表示する際には，その情報と関連情報との時間が近いものを上位に表示する．

また，時間以外に情報にはそれぞれ注目すべき項目があると考えられることから各情報の注目度を定義し，注目度に応じて表示順を決定している．

メールの場合は，以下の式から情報ごとの注目度を求める．

```
< スケジュール記述 >
  pattern ::=  {(“日程”|“日時”)(“は”|“:”)?)}
                    < 時間表現 > {(“から”|“～”|“より”|“スタート”|“-”)?}
                      {< 時間表現 > (“まで”|“に”|“かけて”)}?
```

図 2　スケジュール記述抽出の正規表現パターン例

```
< 年表現 >
  pattern ::= ({[0-9]+“年”|“平成”[0-9]+“年”})?

< 日付表現 >
  pattern ::= ([0-9]+“月”[0-9]+“日”)|([0-9]+“/”[0-9]+(“日”)?)

< 日付相対表現 >
  pattern1 ::= (“今月”| “来月”)
  pattern2 ::= {(“今日”|“きょう”|“本日”)|(“明日”| “翌日”| “あした”)|
                  (“明後日”| “あさって”)|(“明々後日”| “しあさって”)}

< 時刻表現 >
  pattern ::=(“午前”|“午後”)?{([0-9]+“時” [0-9]+“分” )|([0-9]+“:” [0-9]+ )}
```

図 3　時間表現抽出の正規表現パターン例

メールの注目度 = カテゴリ+キーワードの存在
+緊急性+スケジュール記述の有無

まず，カテゴリとはその内容がどういった種類のものか，例えば spam や仕事関係のメールなどの分類をベイジアンフィルタにより行い，予め決めておいたカテゴリに応じてスコアを増減させる．次に，返信してくださいや変更になりましたといった，注目すべきキーワードの存在があった場合にスコアを高くする．また,「至急」や「明日まで」のような期限が含まれているとき，その残り時間に応じてスコアを高くする．最後に，メール内にスケジュールの記述があれば，新たなスケジュールの通知やスケジュール情報の変更を含むメールであることが考えられるのでスコアを高くする．キーワードや期限表現の抽出には，後述するスケジュール記述の抽出と同様のパターンパッチングにより行っている．これらの要素から各メールの注目度を導出し，注目度が高い情報をより上位に表示している．

一方，スケジュールの場合は，以下の式から情報ごとの注目度を求める．

スケジュールの注目度 = キーワードの存在 + 緊急性

まず，キーワードについてはメール同様，スケジュール内の注目すべきキーワードの存在からスコアを高くする．次に緊急性は，スケジュールの開始時間と現在時間から残り時間を計算し，その残り時間に応じてスコアを高くする．これらの要素から各スケジュールの注目度を導出し，注目度の高い情報を上位に表示している．

このように，関連情報を同時に表示するだけでなく，より注目すべきと思われる情報を上位に表示することでユーザの情報認識を助ける．

3.3　スケジュール入力支援

スケジュールの入力支援をするため，メール内に「日程は 3 月 7 日から 3 月 10 日です」や「日時 : 2 月 24 日 13 時 30 分～」といった記述がある場合，スケジュールになり得る情報として抽出する．抽出には，長谷川らの研究 [3] をもとに図 2 に示す

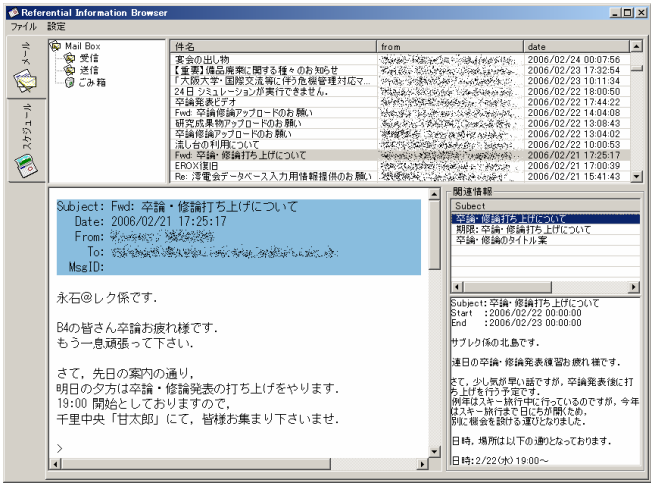

図 4　メール中心の表示画面例

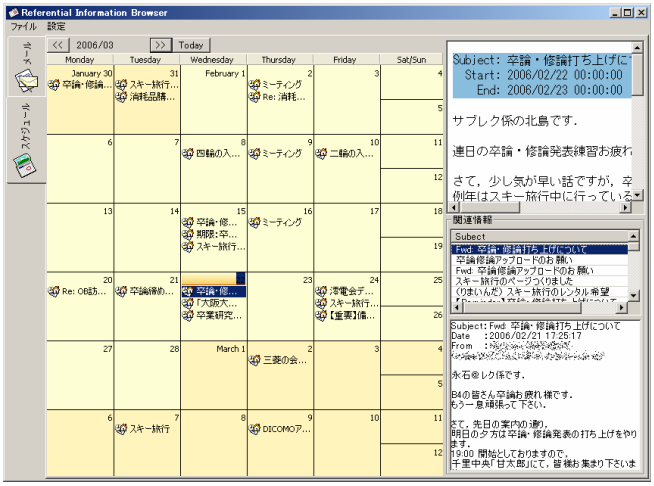

図 5　スケジュール中心の表示画面例

ような正規表現を用いたパターンマッチングを用いる．次に，抽出された時間表現に対し，図 3 に示すような正規表現パターンを用いて時間を抽出することでスケジュールの開始時間と終了時間を抽出する．提案システムでは，ユーザがメールを閲覧している際に，スケジュール記述が存在しているメールであれば，ユーザによるスケジュール入力作業を支援するため，ここで得られたスケジュール情報をスケジューラに追加するかどうかを尋ねるメッセージを出す．

4.　実　　装

本研究では提案システムのプロトタイプを実装した．実装には，Microsoft 社の Visual C++ .NET を用いて行った．メールおよびスケジュールの形態素解析には，京都大学情報学研究科-日本電信電話株式会社コミュニケーション科学基礎研究所共同研究ユニットプロジェクトによって開発されたオープンソース形態素解析エンジン MeCab0.93 [5] を用いた．実装したプロトタイプの情報提示例を図 4，図 5 に示す．

図 4 はメール中心の表示画面例であり，右上部にメールの一覧があり，一覧の中から情報を選択すると，その情報が左下のメール本文表示部に表示され，右下の関連情報表示部にそのメールの関連情報が表示される．このとき，表示されている関

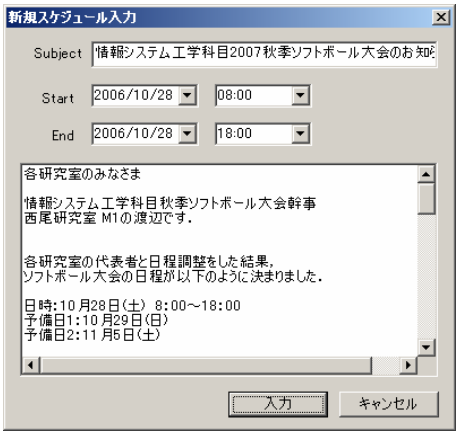

図 6　スケジュール入力画面例

連スケジュールを選択することで，図 5 に示すスケジュール中心の表示に切り替わり，その情報が表示される．関連メールが選択された場合は，そのメールの内容がメール本文表示部に表示される．

図 5 に示すスケジュール中心の表示画面例では，左部のスケジュールの一覧があり，一覧の中から情報を選択すると，右上のスケジュール本文表示部にそのスケジュールが表示され，右下の関連情報表示部には，そのスケジュールの関連メールが表示される．ここでも，メール中心の画面同様，表示したい関連メールを選択することでメール中心の画面に切り替わりそのメールの詳細を確認できる．

また，メール内にスケジュール記述があった場合，図 6 のように，ポップアップによりスケジュール情報を提示する．ユーザはそのスケジュールをスケジューラに入力したい場合，下部の入力ボタンを押すことで簡単にスケジュールの追加が行える．また，ユーザにより必要に応じて修正も行うことができるので，抽出ミスが生じた際にも簡単に修正ができる．

4.1　利　用　例

本システムの利用例として，下記のシナリオを考える．

- A さんはスキー旅行に行くことになっておりスキー旅行の日程を確認中に，近々スキー旅行の代金を支払わなければならないことを思い出した．そこで，いつまでにいくら払わなければならないかを確認したい．

まず，スキー旅行のスケジュール表示は図 7 のようになっている．ここで，右下の関連情報表示部に表示されている関連情報を見ると,「スキー旅行代の集金」というサブジェクトのメールが確認できる．その情報を選択すると，図 7 のようにその情報の一部を見ることが出来る．しかし，代金や支払い期限までは確認できなかったので，表示されている関連情報上をクリックすると図 8 の「スキー旅行代の集金」のメールが表示されている画面に即座に切り替わる．これにより，払うべき代金が 30,000 円であることと，2 月 24 日 17 時までであることが確認できた．

このように，RI ブラウザを使うことで検索機能を利用せず，簡単に関連情報を確認することができ，様々な情報を効率よく得ることができる．

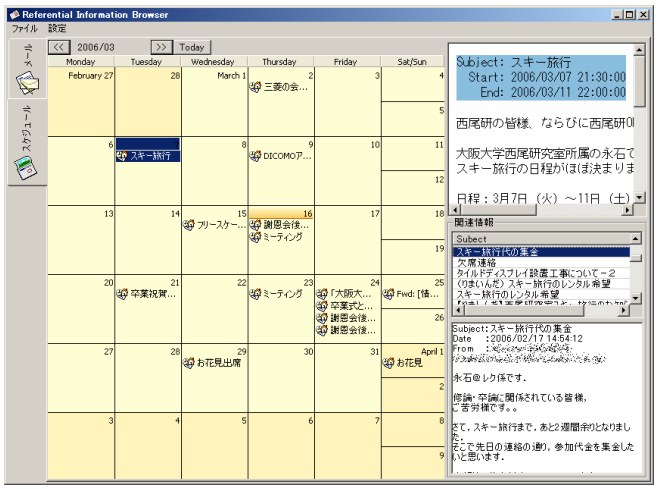

図 7　スキー旅行のスケジュール表示

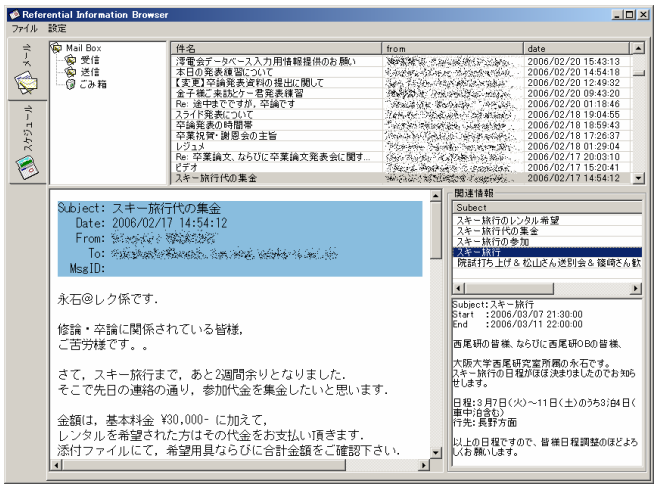

図 8　スキー旅行代の集金のメール表示

4.2　処 理 時 間

本システムでは，メールやスケジュールが登録される際に，内容解析および関連情報の検索を行うため，通常のメーラやスケジューラと比べてデータ登録処理に時間がかかる．また，システムにすでに登録されている情報が多いほど処理時間も多くなるため，スケーラビリティを高めるための工夫が必要になる．そこでプロトタイプシステムでは，単語とその単語が含まれる情報の対を逆引き単語辞書として記録しており，関連情報検索を高速化している．これにより，関連情報検索はメールおよびスケジュールがそれぞれ 1934 通，953 個存在する環境でも 1 つの処理に平均して 0.4 秒程度でおこなえるが，メールやスケジュールに含まれる単語数が多い場合には，5～7 秒程度かかってしまう場合もあり，多くのメールが一度に到着した場合などには体感できる程の処理遅延が発生する．そこで，プロトタイプシステムでは登録時の内容解析や関連情報検索の処理をバックグラウンドで実行することで，利用者は解析処理を意識することなくシステムが利用できる．

見た目上の処理時間をさらに削減する方式に関しては今後の課題であるが，スケジュールの周期性を考慮して，より関連のありそうな時期の情報を先にチェックして表示するといった方法が考えられる．

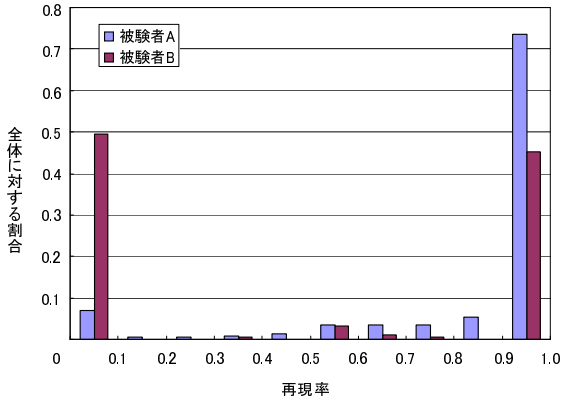

図 9　メールのもつ関連情報の再現率分布

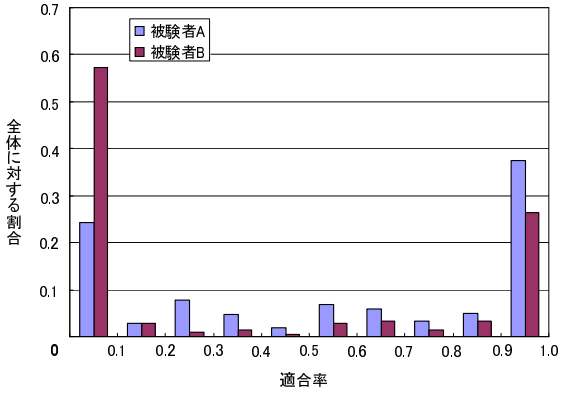

図 10　メールのもつ関連情報の適合率分布

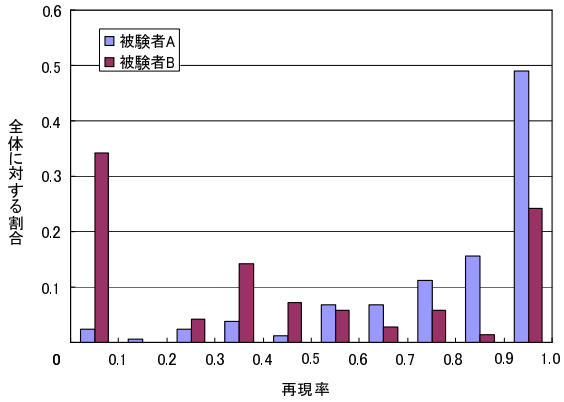

図 11　スケジュールのもつ関連情報の再現率分布

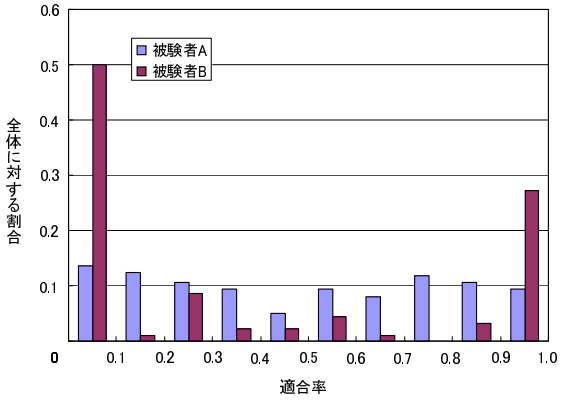

図 12　スケジュールのもつ関連情報の適合率分布

5. 評　　価

実際にプロトタイプを用いて，被験者 2 名の 6ヶ月間のメールとスケジュールをもとに関連情報の取得精度を評価した．評価に利用したデータ数は，被験者 A はメール 915 通とスケジュール 161 個，被験者 B はメール 934 通スケジュール 953 個である．評価は，被験者が主観的に判断した関連性と提案システムにより判定された関連性との再現率と適合率を求めることで行った．再現率は，被験者が判断した関連性のうち実際にシステムが関連があると判断した割合を表し，この値が低いときは本来抽出されるべき情報が不足していることを表す．一方，適合率はシステムが判断した関連性のうち被験者が関連があると判断した割合を表し，この値が低いとき不適切な情報をシステムが多く抽出していることを表す．これらの再現率・適合率を全てのメールとスケジュールに関して，システムが判断した関連情報と被験者により判断した関連情報から導出した．ただし，スケジュールは自動入力によるものを含まないとする．

5.1　評 価 結 果

評価結果として，メールのもつ関連情報の再現率および適合率の分布を図 9，図 10 に示す．また，スケジュールのもつ関連情報の再現率及び適合率の分布を図 11，図 12 に示す．

まず，図 9 のグラフより，2 人の被験者共に，メールにおいて必要な関連情報を 90%以上取得できているメールと，10%未満しか取得できていないメールがほとんどであることがわかる．これは，多くのメールにおいて関連情報を精度よく検索できる一方で，メールや関連するスケジュールで，特徴単語がうまく抽出できていなかった場合は必要な関連情報をほとんど検索できないことを示している．特に人名などの固有名詞が多く含まれている場合，それらの単語が特徴的な単語として扱われてしまい，必要な単語が除外されてしまうことがある．次に，図 10 のグラフより，再現率と同様に 2 人の被験者共に適合率が 90%以上のメールと，10%未満のメールが多いことがわかる．適合率が 90%以上の約 3 割のメールに関しては，関連していない不要な情報が少なく，的確に関連情報を取得できている一方で，適合率が 10%未満のスケジュールに関しては，余分な情報が多いことを示している．しかし，これらの余分な情報に関しては，ユーザによって単語類似の閾値設定を変更し，関連情報の量を調整できることから，ユーザインターフェース上で動的に閾値を調整して余分な情報を少なくできる．また，前述のように関係のある話題のやりとりは時間的に近いことが多いため，本システムでは時間的に近い情報を上位に表示している．その結果，余分な情報は下位に表示され，実際に利用する際には余分な情報は気にならなかった．

次に，図 11 のグラフより，スケジュールもメールとほぼ同じように，必要な関連情報を 90%以上取得できているスケジュールと，10%未満しか取得できていないスケジュールが多いことがわかる．メールに比べて必要な関連情報を 90%以上取得できているスケジュールの割合が少ないのは，スケジュールはサブ

ジェクトのみで記述されていることが多く，関連性を見るための単語が少ないので，必要な関連情報の取得がメールに比べて困難であるためである．特に被験者 B のスケジュールに関しては，必要最低限の情報しか記述されていないスケジュールが多かったため，再現率が全体的に低くなっていた．しかし，詳細情報が書かれているスケジュールに関しては，多くの情報で必要な関連情報を取得できている．一方，図 12 のグラフより，適合率が 90%以上のスケジュールはあまりなく，適合率が低くなってしまっているスケジュールが多い．これは，前述のようにスケジュールには関連性を見るための単語が少ないため，あまり特徴的でない単語が特徴単語として抽出されてしまい，実際には関連のない情報まで関連情報として取得してしまうためである．しかし，この問題に関しては，メール同様，閾値の設定による情報量の調整や，時間的に近い情報が上位に表示されることから利用の際にはあまり気にならない．

5.2 考　　察

メール，スケジュール共に再現率より適合率が低くなったのは，余分な情報が多く含んでしまうことを意味するが，これは情報の内容を端的に表す特徴的な単語を正しく抽出できなかったため，関連情報の検索に失敗し関連情報を過分に取得してしまうためであると考えられる．例えば，メールには文書の初めに送信者の名前が書かれていることが多くあるが，同じくメールの下部に署名の形で名前が書かれることがある．そのため，同じ文書内に名前が複数回出現するため *tf* 値が上がり，結果として *tf*idf* 値が大きくなるので特徴的な単語として選出されてしまう．このような文書が複数あった場合,「○○さんとの打ち合わせ」のように名前がスケジュールのタイトルとして利用されているようなスケジュールがあると，関係のない話題でも全て関連性があるとされてしまう．逆に，複数人の名前が記述されているメールでは，名前のみが特徴的な単語として抽出されてしまい，話題を特徴付ける単語が特徴的な単語として選出されず，関連情報を取得できないということもあった．

これは，提案システムでは情報間の単語の類似性およびメールの返信関係のみを用いることで関連情報を検索しているためであり，この問題を解決するためにはメールの送信者や送信対象，情報に関連付けられた日時などさまざまな要素を利用して情報間の類似性を計算することが必要であると考えられる．ただし，多様な情報を用いれば関連情報を広く取得でき再現率の向上が期待できる一方，不適切な情報も多数含まれて適合率がさらに低下する恐れがある．適切な情報間の類似性検出アルゴリズムの提案は今後の課題であるが，関連度という新たなパラメータを定義することでより精度の高い関連情報抽出が行えるのではないかと考えている．

また，関連情報の取得に失敗している場合に，ユーザの操作による学習を行い関連情報取得精度を向上させることも考えられる．提案手法では情報から抽出した特徴的な単語を用いて情報間の関連性を計っているが，ユーザが不適切と判断した情報や不足していると判断した情報に含まれる単語を記録しておき，関連性の抽出の際にそれらの単語の *tf*idf* 値を減少・増加させることで精度の向上が期待できる．

6. おわりに

本研究では，電子メールとスケジュールの関連性を考慮した新しい情報提示手法を提案した．提案手法では，メールやスケジュール情報に関連する情報を自動的に検索し，関連情報としてユーザに提示することで，関連情報の検索にかかる手間を軽減する．また，関連情報を表示するだけでなく，関連情報を選択することで，任意の情報に移動できるようにすることで，情報の切り替えをシームレスに行えるようにする．これにより，関連情報の検索，閲覧が容易になる．さらに，スケジュールの入力支援としてメール内に書かれているスケジュールの記述から自動的にスケジュールを入力可能にすることで，スケジュールの入力支援を行っている．

提案手法による関連情報の取得精度の評価結果より，提案システムが多くの情報に対する関連情報を十分に取得できることを確認した．

今後は，関連情報を取得する際にメールの送信者や送信時間，スケジュールの時間といった様々な要素を考慮することで関連情報取得精度を向上させる予定である．また，今後は関連情報の提示順序に関する評価や複数の被験者による提案システムの運用実験を行い，求める情報へと到達する早さ等により提案システムの実用性の評価を行うことを考えている．

謝　　辞

本研究の一部は，文部科学省 21 世紀 COE プログラム「ネットワーク共生環境を築く情報技術の創出」の研究助成によるものである．ここに記して謝意を表す．

文　　献

[1] 浅野久子, 加藤恒昭, 高木伸一郎, “Signature の局所的パターンマッチによる電子メールからの送信元住所録情報の抽出とそれを用いた住所録管理システム,” 情報処理学会論文誌, vol.39, no.7, pp.2196–2206, July 1998.

[2] feedpath Zebra, http://www.feedpath.co.jp/zebra/.

[3] 長谷川隆明, 高木伸一郎, “文書構造の認識と言語の特徴の利用に基づく電子メールからのスケジュールと ToDo の抽出,” 情報処理学会論文誌, vol.40, no.10, pp.3694–3705, Oct. 1999.

[4] 樋口英司, 荒木健治, “類似度情報を用いたグループ化による電子メール返信文下書き自動生成手法について,” 第 19 回人工知能学会全国大会論文集, 1E2–02, June 2005.

[5] MeCab: Yet Another Part-of-Speech and Morphological Analyzer, http://mecab.sourceforge.jp/.

[6] J. Tullio, J. Goecks, E.D. Mynatt, and D. H. Nguyen, “Augmenting shared personal calendars,” in Proc. the 15th annual ACM symposium on User interface software and technology, pp.11–20, Oct. 2002.

[7] 柳瀬隆史, 仲尾由雄, “メールマガジンを利用した注目ニュースの自動抽出,” 情報処理学会研究報告 (自然言語処理研究会 1999-NL-136), vol.2000, no.29, pp.151–158, Mar. 2000.