

携帯端末による電子メール交換に基づく Web 検索

赤星 祐平[†] 小山 聡^{††} 角谷 和俊^{††} 田中 克己^{††}

[†] 京都大学工学部 〒 606-8501 京都市左京区吉田本町

^{††} 京都大学大学院情報学研究科 〒 606-8501 京都市左京区吉田本町

E-mail: [†]{akahoshi,oyama,sumiya,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし 現在、インターネットの利用可能な PDA や携帯電話といった携帯端末が広く普及している。しかし、携帯端末からのインターネット利用の現状は、端末の操作性などから、あまり利用されておらず、利用する場合でも用途が非常に限られており、PC などでもよく利用される Web 検索についてはほとんど利用されていない。そこで、本研究では、携帯端末からでもよく利用されている電子メールを利用して、電子メールの交換内容から関連情報を推測し、その情報を含む Web ページを検索し、ユーザーに提示するようなシステムを提案する。

キーワード Web とインターネット、情報検索、電子メール、携帯端末

A Web Search Method Based on E-mail Dialog by Mobile Terminals

Yuhei AKAHOSHI[†], Satoshi OYAMA^{††}, Kazutoshi SUMIYA^{††}, and Katsumi TANAKA^{††}

[†] Faculty of Engineering, Kyoto University Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501 Japan

^{††} Graduate School of Informatics, Kyoto University Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501 Japan

E-mail: [†]{akahoshi,oyama,sumiya,tanaka}@dl.kuis.kyoto-u.ac.jp

Abstract Mobile terminals with functions to access the Internet, such as PDA, cellular phone, become very popular. However, we rarely use these functions because of its troublesome handling. If we use, these are used for limited purposes. We rarely use Web search that is often used on PC. In this paper, we propose a mechanism to search and provide related information on Web by analyzing e-mail dialog by mobile terminals.

Key words Web and Internet, Information Retrieval, E-mail, Mobile Terminal

1. 序 論

近年、携帯電話や PDA といったようなインターネットの利用可能な携帯端末が普及してきており、いつでもどこでもインターネットを利用できる環境が整いつつある。しかし、現状では携帯端末からのインターネット利用は

- 電子メールのやり取り
- 携帯端末向けの特定の Web ページの閲覧

といった限られた用途でのみ行われている。PC などでもよく利用される検索エンジンを用いた Web 検索は携帯端末からはほとんど利用されておらず、携帯端末から閲覧する Web ページの種類は非常に限られている場合が多く、Web 上にある有用な情報を携帯端末上ではほとんど利用できていない。

しかし、携帯端末上では操作は面倒であると言うものの、携帯端末からの電子メール（以後、メールと表記する）の利用の状況を考えると、その利用頻度はかなり高く、作成されるメールは、通常のメールほどではないものの、それなりの長さのあるものであることが多い。また、Web 上の情報を携帯端末上でも利用したいと言う潜在的な要望は高いものと考えられる。

これらの状況を踏まえて、本研究では、携帯端末からの Web 利用の機会を増やすことを目指して、携帯端末上でも比較的に利用されるメールのやり取りを利用し、その内容に関連するような情報を持つ Web ページをインターネット上から検索し提示するようなシステムを提案する。メールの交換内容から話題などを抽出し、それらを基にメール交換の内容に合うような Web の情報を検索し提示を行う。

このようなシステムを利用すると、例えばメールを交換しながら食事の約束をする場合に、ユーザーの食事場所などに関する知識に加えて Web 上からメール中の条件などに合うような食事場所の候補などを含む情報を提示されることで、選択する候補が増えるなどの対話の補助、促進が可能になると考えられる。さらに、この一連の流れを応用することで、メールの対話内容を Web アノテーションとして付加し活用することが可能になると考えられる。

2. 関連研究

田中らの研究[1]では、チャットの対話を用いた情報探索の手法が提案されている。あらかじめ用意されたデータベースに

対して、ユーザー間の対話から意図を抽出し、その意図に合った、データベースに対する問い合わせを適合フィードバックを用いて生成し情報探索を行う手法が述べられている。本研究では、ユーザー間の対話から、あらかじめ用意されたデータベースからではなく対話の関連情報を持つ Web ページを検索し情報を提示する手法を提案する。

田島らの研究 [2] では、ハイパーテキストとみなすことが可能な WWW やネットニュース、メールの集まりについて、話題を抽出してクラスタリングを行う手法について述べられている。ネットニュースやメールについては、その返信関係から生成される木構造を基にクラスタリングを行っている。本研究では、携帯端末から送受信されるメールから話題抽出を行うために、メール同士の類似度計算などが非常に困難となることから、この研究で行われている話題抽出の手法とは異なる手法で話題抽出などを行っている。

また、本研究では、対話の内容によって関連情報として検索すべき情報の種類は異なると考えているため、関連情報を検索するための検索質問の生成方法を変えることを行っている。検索すべき情報の種類として、ある店の評判やうわさなどといった属性的な情報を与えようとする対話の種類があるが、立石らの研究 [3] では、通常検索エンジンを利用するだけではうまく取得できないインターネット上に存在する評判情報の検索の手法について提案し、その精度などの評価を行っている。

3. 本研究のアプローチ

携帯端末上で Web 検索が行われない理由としては

- 情報検索のためのキーワード入力
- 検索結果から「本当に欲しい情報」の発見

が面倒であるという点が主に挙げられると考えている。そこで、本研究では、この 2 点について、それらで生じる面倒臭さの軽減・解消の手がかりとして、携帯端末上からでもよく利用されるメールを利用するものである。

本研究において、メールの交換内容から関連 Web ページを推測し、ユーザーに情報を提示するまでのおよその流れは次のようになる。

- (1) 携帯端末からメールを送信する
- (2) 送信されたメールから、メールの内容を表す語（内容語）を抽出する
- (3) メール対話の内容を抽出し、検索質問生成のためのキーワードを抽出する
- (4) メール対話の種類を解析する
- (5) メール関連情報を検索する
- (6) 得られた検索結果の Web ページを、メール対話の種類を基に解析し、適当な情報を選択する
- (7) ユーザーに情報を提示する

前述の Web 検索を携帯端末上で行わない理由と対応付けすると、(2) と (3) を情報検索のためのキーワード入力の省略のために、(4) と (6) を、本当に欲しい情報の発見の面倒臭さの軽減のために行おうとするものである。本研究のイメージは図 1 に示している。

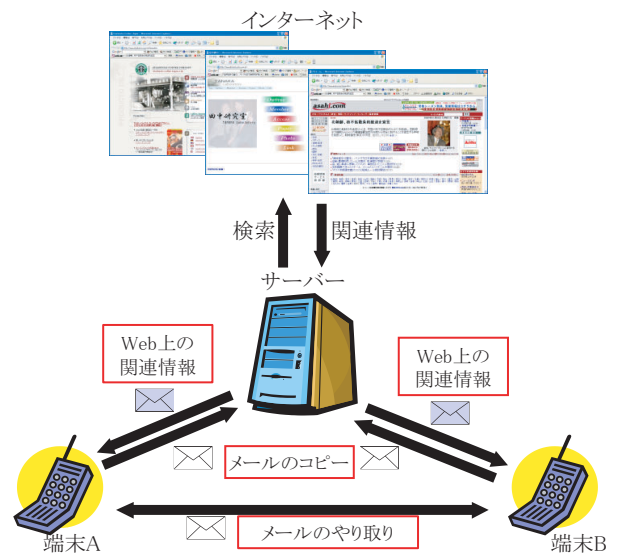


図 1 本研究の概要図

4. メールの内容の抽出手法

メールの交換の内容から関連情報を含む Web ページを検索し提示しようとする場合、まず、個々のメールの内容やメールの対話の大まかな流れをメールから抽出する必要がある。そこで本研究では、メールから抽出する要素について、メールの「内容語」「話題語」「削除対象語」という 3 つを定義し、それらをメールから抽出することで関連情報の検索を行うための質問生成に利用している。本章では、まず携帯端末でやり取りされるメールの特徴を考察し、内容語、話題語、削除対象語の定義を説明し、それらの抽出手法について述べる。

4.1 携帯端末から送受信されるメールの特徴

携帯端末でやり取りされるメールについて、その特徴を考えると次のような点が挙げられる。

- 1 通のメールの長さが通常のメールの長さ比べて短い
- 単語の省略や指示語が多用される傾向にある

つまり、携帯端末でメールを交換しようとする、端末の操作の面倒臭さなどの端末に起因する問題点によって 1 通のメールの長さが短くなる傾向にある。また、単語の省略や指示語が多用される。しかし、携帯端末でのメールのやり取りがたいいの場合に互いに面識のあるような相手と行われることから起因される暗黙の了解（ユーザー間の共通認識）が存在することが多いために、短いメールであっても互いのメール内容の理解に支障は生じないのである。

しかし一方では、このような特徴もある。

- 返信メールを送る場合は、その内容は最新のメールの内容を受けていることが多い

通常、PC などで返信メールを送る場合は、返信元のメールの内容を引用して返信メールを作成することがよく行われる。しかし、携帯端末の場合、表示画面が小さいなどの要因から、多くの場合で返信元のメールの内容を引用して返信メールを作成することを行わない。その代わりに、最近に送られたメールに対して返信することが多く、つまりは返信されるメールは、そ

の直前に送られたメールの内容を受けたものであることが多い。よって、メール上で多用される省略や指示語の内容は、直前に送られてきたメールを参照することで把握することが可能であると考えられる。

これらの点を考慮しながら、本研究ではメールの内容の抽出などを行っている。

4.2 メールから抽出する内容

本研究では、メールの関連情報を検索するために次のものをメールから抽出する。

- 「内容語」
- 「話題語」
- 「削除対象語」

ここで出ている「内容語」「話題語」「削除対象語」とはそれぞれ次のような語のことである。

内容語

内容語とは、ある単一のメールについて、そのメールの内容を把握するために必要とする語のことである。本研究では、メールの内容に関連する Web ページを検索するための質問を生成する際には、主に内容語の中から適当なものを選択し質問生成に用いるので、内容語は関連情報の検索に大きな影響を与えるものである。

話題語

話題語とは、ある連続する数通のメールについて、そのメールの対話内容を把握するために必要とする語のことである。「複数のメールを対象にした内容語」のように言い換えることも可能であると考えられる。この話題語は、関連情報を検索する際の検索質問を生成する際に、その補助をするために主に利用する。

削除対象語

対話を行っているうちに、ユーザーから「はいらない/不要」などといった発言が行われることがあるが、そのような発言で出現する語を内容語として抽出し、その語を用いて Web 検索を行うと、ユーザーにとって必要の無い情報を含むページが結果として得られることは明らかである。よって、関連情報を検索する場合に検索質問として含まれてはならない語を、本研究では「削除対象語」として登録し利用をする。この語を抽出する際には、メールに含まれる文章の種類を解析し、そのような語が含まれていると考えられる文章の部分から抽出をする。

次節以降ではメールから抽出するそれぞれの項目について、その抽出手法について述べる。

4.3 内容語の抽出

メールの内容語は、前述の通り、関連情報を Web から検索する際に、その結果に大きな影響を与える要素となる。また、本研究では、内容語として抽出した語をそのまま利用して、検索質問を生成する。

内容語として、本研究では次のような語をメールから抽出している。

- 名詞
- 代名詞で指されている内容
- 形容詞

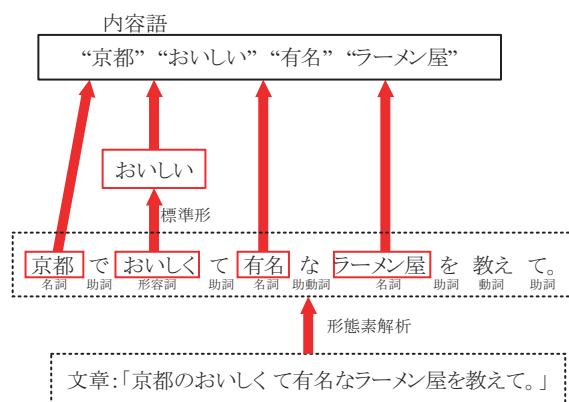


図2 内容語抽出の例

• 形容動詞

このような語を抽出する理由としては、現在の Web 検索では基本的に単語を並べることで条件を指定し検索する形を取ることが多いため、検索質問に利用する頻度が高いと考えられるこれらの語を抽出している。

これらの語の抽出の方法は次の通りである。

(1) 内容語の解析対象の文章の形態素解析を行う。形態素解析では、語の「品詞分類」、形容詞など活用をする語の場合にはその語の基本形（終止形）を決定する

(2) 形態素解析の結果を解析する

(a) 代名詞が検出された場合には、指示代名詞（「ここ」、「そこ」など）であれば、直前のメールの内容語を参照し、その中から「名詞」や「代名詞でさされた内容」となる語を探し、その内容を指示代名詞の内容とし、その語を内容語に登録する。人称代名詞（「私」、「あなた」）と指示代名詞で不定称（「どこ」、「だれ」など）の場合は無視する

(b) 名詞（形容動詞語幹も含む）が検出された場合、その語が人名やストップワード（「こと」、「もの」といった形式名詞など）であった場合には無視し、そうでない場合にはその内容を内容語として登録する

(c) 形容詞が検出された場合は、その語の基本形を内容語として登録する

(3) 以上で検出された語をその文章の内容語として保存をする

ここで述べた抽出の手法について、「京都でおいしくて有名なラーメン屋を教えて」という文章に適用した場合の抽出までの流れを図2で示している。

4.4 話題語の抽出

話題語とは、前述の通り、数通のメールを通じての内容語に相当するものである。したがって、話題語の抽出については、通常であれば複数あるメールの内容を語の出現頻度などから解析してクラスタリングを行い、同じ内容であると判断されるメールを対象に話題を抽出すべきである。しかし、携帯端末からのメールを想定する場合、

- メールが長さが短い
- 省略、指示代名詞の利用が多い

という前述の特徴から、田島らの研究[2]などで提案されてい

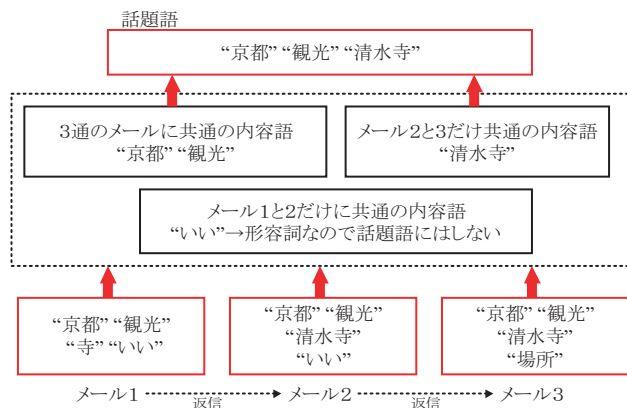


図3 話題語抽出の例

る個々のメールの内容を単語の出現頻度などを基に解析しクラスタリングの手法を取ることは困難であると考えられる。よって、本研究では、一般的にはその数を可変とするべきであると考えられるが、最新のメールから一定数のメールまで遡って、そのメール群を対象に話題語の抽出を行う。

本研究で、メールの交換の話題語として抽出する語は次のような語である。

- 名詞（形容動詞語幹である名詞は除く）
- 代名詞で指されている内容

話題語の候補としては、内容語では候補としていた「形容詞」や「形容動詞」などといった、いわゆる「修飾語句」は利用していない。その理由としては、対話の話題として考えた場合に、形容詞などの修飾語句を話題語に含めると話の内容が非常に狭まってしまう恐れがあるために、話題にある程度の幅を持たせようとするためである。

また、話題語は内容語として抽出された語の中から選択し抽出する。その理由としては、メールのそれぞれの内容が累積されることでメール交換の話題が形成されるので、メール交換の話題語として抽出される語は、メールの内容語として抽出されていた語の中から選択し抽出することが妥当であると考えられるためである。

話題語を抽出するまでの流れは次の通りである。

- (1) 話題語を抽出する対象のメールについて、個々のメールの内容語を抽出する
- (2) 抽出されたメールの内容語を比較する
 - (a) 抽出対象の全メールに共通に出現している内容語のうち、形容詞を除くものを話題語として登録する
 - (b) 抽出対象のメールのうち前後の2通のメール（あるメールとそれに対する返信メール）を比較し、共通に出現する内容語のうち、形容詞を除くものを話題語として登録する
- (3) 以上で登録されたものをメール交換における話題語とする

図3で話題語抽出の例を示している。3通のメール（メール1、メール1の返信メールに当たるメール2、メール2の返信メールにあたるメール3）についてそれぞれ内容語が抽出された後、3通のメールで共通に出ている内容語、メール1とメール2に

共通に出ている内容語、メール2とメール3に共通に出ている内容語をそれぞれ探し、見つかったものから形容詞などを除いて得られたものを話題語としている。

4.5 削除対象語の抽出

メールの内容語、主題語の抽出手法は述べたが、抽出された内容語や主題語を並べるだけでは関連情報の検索はうまくいかない。たとえば、「大阪より京都に行きたい」という発言があった場合、単にこの発言から内容語として名詞を取り出すと「大阪」「京都」という2つの単語が出てくるが、関連情報を検索しようとする場合「大阪」「京都」の2つの語を同時にキーワードにすることは適当でない。それは、「大阪より京都に行きたい」という文章から、この発言を行ったユーザーは「大阪の情報よりも京都の情報が欲しい」という意図を持っていると考えることができるからである。そこで本節では、メールに含まれる文章の種類を判断を行い、削除対象語を抽出する方法を述べる。

本研究において、関連情報の検索のための質問の生成の際に重要であると考えている文章の種類には次のようなものがある。

- 比較（「より××がいい/悪い」など）
- 部分否定（「はいいらない」など）
- 部分肯定（「は必要」など）

比較の文章が含まれる場合は、比較対象となっている語について、文章の内容から検索質問生成の際にその語を質問のキーワードに含めるかどうかを判断することが必要になると考えられる。部分肯定の場合はその文章中で示された語以外の要素について、検索質問に含めるかどうかを判断する必要があると考えられる。さらに部分否定の場合は、文章中で指定された語を、検索質問からは必ず必要があると考えられる。

本研究では、これらの文章の種類を「典型的な表現」と照合することで判断を行う。

a) 比較

比較表現の場合は、「AよりBが～」というような表現が見られた場合に比較の文章であると判断を行う。AとBの比較の内容については「～」の部分に含まれる表現から判断を行う。「～」の内容が肯定的内容である場合は、Bを内容語として登録し、Aの方は「削除対象語」として登録を行い、話題語や検索質問の生成の際に、削除対象として登録された語がそれらに含まれないようにする。「～」が否定的内容であれば、Aの方を内容語として、Bの方を「削除対象語」として登録を行う。

b) 部分否定

部分否定表現は、「Aは××（否定的表現）」というような表現が見られた場合に部分否定の文章であると判断する。そして、Aを「削除対象語」として登録を行い、この語が話題語や検索質問に含まれないようにする。また、全面否定の表現は部分否定の表現の一部として考えるものとし、否定されるものが可能性のある全てのものとし、その場合は、削除対象語に、文章中で否定されていると見られる全ての語を登録する。

c) 部分肯定

部分肯定表現は、「Aは（肯定的表現）」というような表現が見られた場合に部分肯定の文章であると判断する。そして、Aは内容語として登録を行う。全面肯定の表現は部分肯定に表

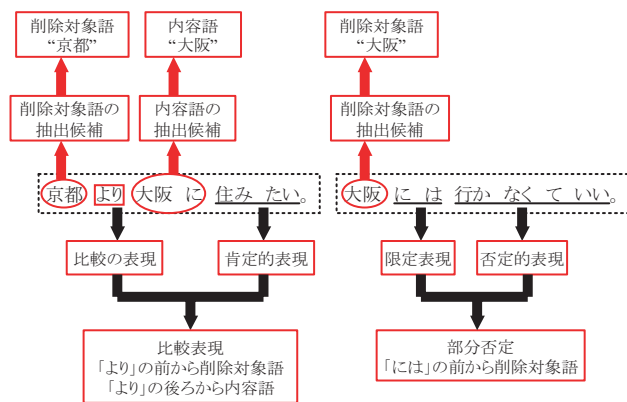


図4 削除対象語の抽出例

現の一部と考え、肯定されるものが可能性のある全てのものとし、その場合は、肯定していると考えられる全てのものを内容語として登録する。

それぞれ、文章の種類の判断を行い、削除対象語を抽出する際には、削除対象語が存在する文章中のある部分に対して、前述の内容語を抽出する手法を利用し、抽出された語を削除対象語として登録する。図4では、「京都より大阪に住みたい」(比較)と「大阪には行きたくない」(部分否定)という削除対象語を抽出することになる2つの例文について、文章の種類の判断から削除対象語の抽出までの流れを示している。それぞれ、比較と部分否定の典型的な表現を発見した場合に、文章中のどの部分から内容語、削除対象語のいずれを抽出するか判断し、それに基づきそれぞれ抽出を行っている。

5. 対話の種類とその判断手法

4. では内容語、話題語、削除対象語などの抽出手法について述べたが、それだけでは関連情報をうまくユーザーには提示できない。対話の種類によって関連情報として提示すべき情報の種類に違いがあると考えられるからである。そこで本章では、関連情報として提示すべき情報の種類によって対話型を分類し、それぞれの分類について、その性質や対話型の判断手法について述べる。

5.1 対話の種類分類

本研究では、対話の種類分類を、「あるメールの返事としてどのような情報が求められているか?」という観点から行った。そして2つの分類を行っている。その種類が次の通りである。

- オブジェクト検索型対話
- プロパティ検索型対話

以下で、それぞれの対話の分類について、その内容と判断手法を述べる。

5.2 オブジェクト検索型対話

オブジェクト検索型対話とは、いくつかの属性(プロパティ)の条件を挙げることで、その属性を満たすもの(オブジェクト)を見つけようとする対話である。例えば

- 「京都のラーメン屋でどこがおいしい?」
- 「冬の観光に行くならどこがいい?」

このような質問がメール中に含まれるとオブジェクト検索型対

話と判断される。前者では「京都」「ラーメン屋」「おいしい」という条件を挙げることで、回答として「京都でおいしいラーメンを出しているお店の名前」を求めている。後者では「冬」と「観光」という条件を与えることで、冬に観光に行くのにお勧めの「地名」を求めている。

つまり、比較的一般性の高い単語を挙げることで、それらを満たす、より一般性の低い答えを求めようとするのがこの対話型である。現在一般に行われる単語キーワードを用いたWeb検索と同じ発想であるとも考えることができる。

5.3 プロパティ検索型対話

プロパティ検索型対話とは、ある特定のもの(オブジェクト)を条件として挙げることで、その評判など(プロパティ)を回答として求めようとする対話型である。

- 「ってというラーメン屋さんっておいしいの?」
- 「冬の京都ってどんな感じなの?」

このような質問がメール中に含まれる場合このプロパティ検索型対話と判断される。前者では「」という特定の店を挙げることで、その店で出されるラーメンの評判を聞こうとしている。後者では「冬の京都」という条件を挙げて、漠然とではあるが冬の京都の印象などを聞こうとしている。

つまり、一般性の低い単語をあげることで、それを満たす、より一般性の高い答えを求めようとするのがこの対話型である。見方を変えると、前述のオブジェクト検索型対話の条件と回答が入れ替わった対話がこの対話型になると言える。これは、一般的なWeb検索とは違い、立石らの研究[3]で行われている評判情報検索の一種と見ることもできる。

5.4 対話分類の判断方法

以上で挙げた2つの対話型について、ここではその判断手法を述べる。

まず、およその流れを示すと次のようになる。

- (1) 送られてきたメール中に疑問文が含まれるときには、疑問文中の内容語から疑問文の一般性の度合いを計算する
- (2) 送られてきたメール中に疑問文が含まれない場合には、メール中の内容語からメールの一般性の度合いを計算する
- (3) 計算された一般性の度合いが一定以下であればオブジェクト検索型対話、一定より大きい場合はプロパティ検索型対話と判断する

1で用いたオブジェクト検索型対話とプロパティ検索型対話の判断のための「一般性の度合い」の計算手法は次の通りである。

x_1, x_2, \dots, x_n を疑問文に含まれる内容語、 $g(x)$ を内容語 x の一般性の値とする。ただしここで、一般性の値は「一般性が低いほど値が大きい」ものとする。つまり、 x が固有名詞になると値は大きくなり、一般名詞であると値は低くなる。そして閾値 C として、

$$\frac{\sum_{i=1}^n g(x_i)}{n} \begin{cases} \leq C & (\text{オブジェクト検索型対話の場合}) \\ > C & (\text{プロパティ検索型対話}) \end{cases}$$

つまりは、疑問文中に含まれる内容語の一般性の値の平均をとり、その値と閾値の比較によって対話型を判断する。

オブジェクト検索型対話とプロパティ検索型対話の区別のために以上のような手段を取るのには、オブジェクト検索型対話になる場合の質問文に含まれる語と、プロパティ検索型対話になる場合の質問文に含まれる語を比べた場合、5.2 や 5.3 で述べた通り、

- オブジェクト検索型対話と判断される質問文では、比較的一般性の高い語が並ぶことが多い

- プロパティ検索型対話と判断される質問文では、比較的一般性の低い語が並ぶことが多い

という特徴を見ることができると考えるからである。つまり、このことから、質問文中の内容語の一般性の平均を取ることによって判別可能であると考えられる。

6. 関連情報の検索と提示

ここまでで、メールの関連情報を検索するために、メールから対話の内容などキーワードの抽出を行った。本節では抽出されたキーワードから検索質問を生成し、その質問で検索を行い得られた結果のユーザーの提示の方法について述べる。

6.1 検索質問の生成

ここでは、メール中から抽出してきた内容語、話題語、削除対象語の3つの要素から Web 検索を行うための検索質問の生成について述べる。

検索質問の生成についてはつまり、 $C = \{c_1, c_2, \dots, c_n\}$ を内容語、 $T = \{t_1, t_2, \dots, t_m\}$ を話題語、 $R = \{r_1, r_2, \dots, r_l\}$ を削除対象語とすると、検索キーワード集合 Q は、

$$Q = (C \cup T) - R$$

で表される（ただし、 $A - B$ は集合 A と B の差集合を表す）。そして、 Q の要素の積を検索質問として Web 検索を行う。つまり、検索質問は $Q = \{q_1, q_2, \dots, q_k\}$ とすると

$$q_1 \wedge q_2 \wedge \dots \wedge q_k$$

となる。

6.2 検索結果のユーザーへの提示

前節で生成された検索質問を用いて Web 検索を行うが、5. で述べたように、対話の内容によって、同じ検索質問が生成されてもユーザーの欲しい情報と言うものは異なると考えられる。そこで、本節では、5. で判別した対話の種類を基に、検索結果をユーザーに提示する方法について述べる。

6.2.1 オブジェクト検索型対話

対話がオブジェクト検索型対話と判断された場合には、検索結果として、例えば店名といった具体的な名称などがユーザーに求められていると考えられる。そこで、この対話の場合には検索結果の Web ページに固有名詞が多く含まれるページが、よりユーザーの要望に沿った情報を含むページと考えることができる。そこで、オブジェクト検索型対話の場合は Web ページに出現する全ての単語の中で固有名詞が出現する頻度が高い順にユーザーに提示するといいたいと考えられる。

つまり、 $f(P)$ を Web ページ P 中の固有名詞の総数、 $g(P)$ を Web ページ P 中に出現する単語の数として、

$$F(P) = \frac{f(P)}{g(P)}$$

なる $F(P)$ の値を、検索結果として得られた全ての Web ページについて計算し、その値が大きくなる Web ページを優先的にユーザーに提示するようにする。

6.2.2 プロパティ検索型対話

対話がプロパティ検索型対話と判断された場合には、検索結果として、例えばある店の評価といった様な情報がユーザーに求められていると考えられる。一般に、評価などの情報を表現する際には、

- 数値などによるランク付けによる表現

- 「はいいい」といった形容詞などによる表現

が多く用いられる。しかし、一方では、プロパティ検索型対話はオブジェクト検索型対話の対となる対話であることから、固有名詞の出現頻度が低い Web ページには評価などの情報が多く含まれていると考えることもできる。そこで、本研究では、前者の考え方を採用し、Web ページ中の形容詞、形容動詞の出現頻度の高い順に優先してユーザーに情報を提示するようにする。

つまり、 $h(P)$ を Web ページ P 中の形容詞、形容動詞の総数、 $g(P)$ を Web ページ P 中に出現する単語の数として

$$G(P) = \frac{h(P)}{g(P)}$$

なる $G(P)$ を検索結果として得られた全ての Web ページについて計算し、その値が大きくなる Web ページの情報を優先的にユーザーに提示するようにする。

7. システムの評価

これまで示してきた方式について、本節では PC 上でプロトタイプシステムを作成し、一部に限られているが行った評価について述べる。

7.1 プロトタイプシステムの概要

本研究で作成した PC 上でのプロトタイプシステムでは、Microsoft Visual C#.NET, Google Web API [5], 形態素解析ツール「茶筌」[4] を利用して、順次入力された携帯端末上を想定したメール（例：図 6）を解析し、その内容を解析して内容語、話題語、削除対象語を抽出し、また、対話の種類を判断して、生成された検索質問を利用して検索を行っている。処理の流れは図 5 のようになっている。

7.2 評価

現段階では、

- 内容語、話題語、削除対象語の抽出
- 指示語の内容補完
- 対話の種類判別

について、作成したプロトタイプで評価を行っている。

7.2.1 内容語、話題語、削除対象語の抽出について

4. で提案した手法によって、図 6 のようなメールから内容語などを抽出を行った場合、次のような点が分かった。

- メールが長くなると内容語が多く抽出される傾向にあり、そうなる関連情報検索がうまくいかなくなる

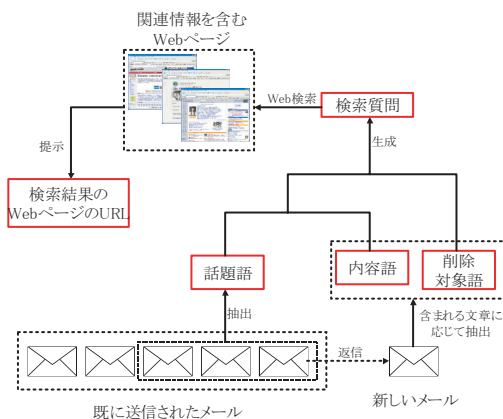


図5 プロトタイプ処理の流れ

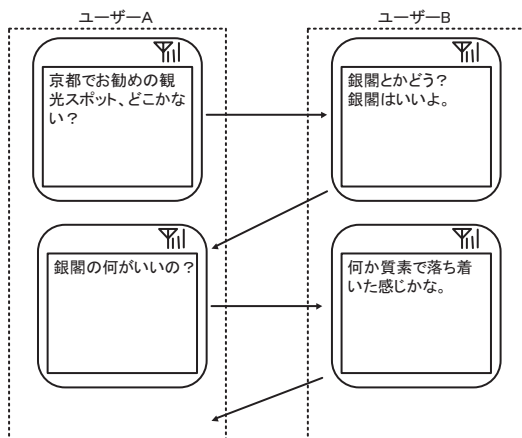


図6 メールによる対話の例

• 対話が長く続き、省略される語が増えると話題語の抽出がうまくいなくなるため、関連情報の検索の精度が落ちるつまり、携帯端末からのメールであると、言いたいことをコンパクトにまとめてメールにする傾向があるために、メールから内容語などを抽出することが比較的うまくいくが、メールが長くなると、本文中のどの部分がか最もメールの送信者が言いたいことなのかを判断するのが難しくなる。また、メールを短くするために不必要、もしくは了解済みの事項に関しては本文中から省略されるため、省略された内容を以下に補完するかのことが重要になる。

7.2.2 指示語の内容補完

指示語の内容補完については、メール中のある指示語の内容は、そのメールの直前に送信されたメールの内容語を、指示語の指す内容として利用しているが、その方法はうまくいっているように思われる。

7.2.3 対話の種類判別

今回は、対話の種類を判別するに当たって、次のような値を設定して実験を行った。

固有名詞	$g(x) = 1.0$
名詞(地名)	$g(x) = 0.7$
名詞(一般名詞)	$g(x) = 0.4$
形容詞, 名詞(形容動詞語幹)	$g(x) = 0.1$

そして、閾値 C を 0.4~0.5 の近辺で調整しながら行ったが、

判別の間違いを生じる場合は、奥の場合プロパティ検索型対話をオブジェクト検索型対話と誤判別することが多かった。また、今回は、助詞などに注目をせずに対話の種類判別を行おうとしたため、微妙なニュアンスで対話の種類が異なると思われる場合には判別が不可能になる。例えば「京都のラーメンのどこがおいしい?」と「京都のラーメンではどこがおいしい?」では、前者は一般で京都のラーメン屋で出されるラーメンのどのような点がおいしいのか、つまりプロパティ検索型の例であり、後者は、京都のラーメン屋さんではどこの店のラーメンがおいしいのか、つまりオブジェクト検索型の例になる。しかし、本研究で現在取っている手法では、助詞の違いのみで対話の種類が異なる場合の判断ができないため、双方は同じ対話の種類(オブジェクト検索型)とみなされてしまう。

7.3 今後の課題

現段階では、正確な評価を行うことができておらず主観的な評価が中心であるため、より客観的な基準に基づく評価の方法を検討し、それに基づいた評価を行う必要がある。また、まだ実装できていない機能についても早急に実装し評価を行う必要がある。

システムについては、メールの本文が長くなったり、対話が長く続くと関連情報検索の精度の低下が見られたため、長くなった場合に的確に適切な情報のみを抽出する方法を検討する必要がある。また、語の省略に対応する方法を検討することが、大きな課題であり、検索する情報の精度を上げる大きな要因になると考えられる。

そして最終的には、実際に携帯端末上で利用できるシステムを構築し、携帯端末で利用した場合の長所、短所などの検討が必要である。

8. 結論

本研究では、現状ではあまり利用されていない携帯端末からの Web の閲覧、情報検索について、比較的利用される電子メールの交換に着目し、その電子メールの交換内容から関連情報を検索してユーザーに提示するシステムの提案を行った。これから、現段階ではあまり考慮していない、ユーザーへの情報の提示について考えていく必要がある。携帯端末でのメールのやり取りの関連情報を、必ずしも携帯端末上で見せる必要はないかと考えている。例えば、

- PC 上で情報を提示
- 公共物(オーロラビジョンなど)を介して提示

などが情報提示に利用する機器として考えられる。PC で情報提示する場合は、携帯端末を情報収集のためのものと割り切って、携帯端末に蓄積した情報を PC に送ることで、PC で解析して関連情報を検索・提示するようにする手法が考えられる。公共物を介して提示するというのは、例えば個人的なメッセージを時間を決めて街角のビジョンに映すサービスなどがあるように、必ずしもある個人が欲しい情報を携帯端末上で提示してその人のみで利用するのではなく、公共のオーロラビジョンといったようなものに提示することで情報を広く共有できるようにするものである。また、提示する情報も Web ページをそのまま利用

するのではなく、必要な情報（例えば、ラーメン屋の話中なら、ラーメン屋の名前、営業時間、そのラーメン屋のメニューなど）だけ取り出して提示の方が効率よく情報を活用できるのではないかと考えている。このように、提示した情報をうまく利用できるような工夫について考えたい。

本研究は、現状では Web 上のテキスト情報についてのみ関連情報を検索し取得するような形になっている。しかし、将来的には携帯端末であっても第 3 世代以降の携帯電話や公衆無線 LAN の普及によって通信速度が高速化していくことや、様々な端末にカメラが装備されることによって静止画像、動画像のやり取りが携帯端末でも頻繁に行われるようになるのではないかと考えられる。そこで携帯端末上で文字情報のみでなく、静止画像や動画像も含めたマルチメディア検索の可能性などについても考えていきたい。

また、本研究の一連の仕組みの応用の一例として、メールを利用した Web アノテーションがあると考えられる。これは、メールの対話の内容そのもの、もしくはその要約を、関連情報の含まれる Web ページにアノテーション（注釈）として付加し利用しようとするものである。例えば、次のようなメールの対話について考える。

- A: 先日の　　という研究会で面白い発表があったよ。
B: どんな研究？
A: A 大学の B さんの　　っていうタイトルの研究なんだけど。
B: その研究の何が面白いの？
A: 今までのやり方とまったく違った斬新なアイデアだったよ。
B: それなら × × 研究会の　　の方がすごかったと思うけど...
A: ...

このような対話があったときに、　　研究会の Web ページや A 大学の B さんの作成している Web ページにある　　という論文の箇所に、この対話の内容をアノテーションとして付加するという流れである。本論文のアプローチ（Web 上の関連情報をメールの対話の適当な部分に付加）とは逆の発想であるが、蓄積されたアノテーションを利用することで、メールの対話内容に関連する情報を検索する場合のみならず、一般の Web 検索においても検索の精度向上などが可能であると考えられる。このアプローチについても考えていきたい。

謝　　辞

本研究の一部は、平成 14 年度文部科学省科学研究費基盤研究 (A)(2)「モバイル環境におけるコンテンツのマルチモーダル検索・提示と放送コンテンツ生成」(課題番号：14208036、代表：田中克己)および平成 14 年度 NEC 共同研究「クロスメディア情報流通システムにおける情報メディアの活性化の研究」(代表：田中克己)による。ここに記して謝意を表します。

文　　献

- [1] 田中貴志, 小山聡, 角谷和俊, 田中克己: ユーザ間の対話からのユーザプロファイル抽出と動的情報探索, 情報処理学会研究報告, 2002-DBS-128-6, pp.39-46, (2002)
- [2] K. Tajima, Y. Mizuuchi, M. Kitagawa, and K. Tanaka: Cut as a Querying Unit for WWW, Netnews, and E-mails, Proc. ACM Hypertext '98, pp.235-244, (1998)
- [3] 立石健二, 石黒義英, 福島俊一: インターネットからの評判情報

- 検索, 情報処理学会研究報告, 2001-NL-144-11, pp.75-82, (2001)
[4] 形態素解析システム 茶筌, <http://chasen.aist-nara.ac.jp/index.html>.ja
[5] Google Web API, <http://www.google.com/apis/>