

携帯端末による電子メール交換に基づく Web 検索

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

赤星 祐平[◇] 小山 聡[◇]
角谷 和俊[◇] 田中 克己[◇]

Yuhei AKAHOSHI Satoshi OYAMA
Kazutoshi SUMIYA Katsumi TANAKA

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 Web 情報検索利用の現状は、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やコンテンツが制限されている。そこで、本論文では、携帯端末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。

Mobile terminals with functions to access the Internet, such as PDA and cellular phones, become very popular. However, we rarely use web search functions because of its restrictions on display size and input devices. In this paper, we propose a mechanism to search and provide related information on Web by analyzing e-mail dialog by mobile terminals.

1. はじめに

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広く普及してきており、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、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、携帯端末上でも有用な情報を容易に入手し活用したいという潜在的要望は高いものとなっている。そこで、本研究では、携帯端末で Web 上の情報を活用するためのきっかけとして、携帯端末上でもよく利用される電子メール（以後メールと表記）を利用することを考える。ユーザー間で交換されるメールの対話の内容に関連する情報を自動的に推測し、その情報を Web から検索しユーザーに提示するシステムを提案する。このようなシステムにより、例えばメールで食事の約束をする場合

に、メールを交換しているユーザーは知らないが条件（料理の種類や場所など）を満たすような食事の店の情報が Web 上から検索され提示されることで、メールによるユーザー間の対話の補助・促進が可能になると考えられる。

2. 関連研究

本研究は、携帯端末上でのメールの交換内容から、関連情報を検索しユーザーに提示するが、田中らの研究[1]では、チャットを用いて、チャットの内容からデータベースへの問い合わせを適合フィードバックを用いて生成する手法が提案されている。対話を情報検索に利用する点では同じ発想であるが、情報検索のための条件、質問の生成については異なる手法を取っている。

また本研究では、一連のメール対話から話題を抽出し関連情報の検索を行うが、田島らの研究[2]では、複数のメールやネットニュースの内容を解析し、話題を抽出してクラスタリングする手法が提案されている。しかし本研究では、「携帯端末上でやり取りされるメール」という極めて限定されたものについて話題抽出などを行うため、異なる手法を取っている。

さらに本研究では、検索する情報の種類によって対話の分類を行っているが、そのうちのプロパティ検索型対話は、立石らの研究[3]における評判情報にヒントを得たものである。立石らの研究では、通常の Web 検索とは異なる、Web 上に存在する評判などの情報の検索手法について提案を行い、精度などの検証を行っている。

3. 本論文のアプローチ

携帯端末で Google などの Web 検索がほとんど利用されない理由は次のような点にあると考えている。

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

そこで、本研究ではこのような操作の煩雑さの軽減・解消の手がかりとして、メールの交換を利用する。

本論文において提案するシステムの概要は図 1 のようなものである。ユーザー間でメールを交換する際に、そのメールの内容をサーバーで解析し、関連情報を推測して Web 検索を行い、最終的に得られた情報をメール交換中のユーザーに提示するものである。



図1 システムの概要

Fig.1 Overview of the System

本システムで、メールが送信されてから関連情報を検索し

[◇] 学生会員 京都大学大学院情報学研究科修士課程

akahoshi@dl.kuis.kyoto-u.ac.jp

[◇] 正会員 京都大学大学院情報学研究科

{oyama, sumiya, tanaka}@dl.kuis.kyoto-u.ac.jp

ユーザーに提示するまでの流れは次のようになる。

1. ユーザーがメールを送信する。
 2. 送信されたメールと、これまでのメールによる対話履歴から、メールの内容とメール交換の話題を抽出する。
 3. 送信されたメールを解析し、対話の種類を判別する。
 4. 抽出されたメールの内容と対話の話題から検索質問を生成する。
 5. 生成された質問を用いて Web 検索を行う。
 6. 検索結果を対話の種類に基づきユーザーに提示する
- 以後、4章でメールからの内容抽出について、5章で対話の種類とその判別方法について、6章ではメールから得られた情報を基に関連情報を検索しユーザーに提示する手法について述べる。そして、7章では、本論文で提案した手法について検証するために行った予備実験について述べ、8章でまとめと今後の課題を述べる。

4. メールの内容抽出

携帯端末上で交換されるメールからその内容を抽出しようとする場合、次のような点が問題点として挙げられる。

- 1通のメールの長さが短い。
 - 単語の省略や指示語が多用される。
- しかし、一方では次のような特徴も見られる。
- 返信メールを送る場合には、その内容は返信元のメールの内容を受けたものになっている。

これらを基にして、関連情報をユーザーに提供するためのメールの内容抽出について述べる。

本論文で提案する方法では、交換された一連のメールから次の3つの要素を抽出する。

- 内容語
- 削除対象語
- 話題語

本章ではこれらの要素について、抽出手法も含めて述べる。

4.1 内容語

内容語とは、あるメールの関連情報を検索する場合に、検索の条件として必要であると考えられる語の集合である。内容語として抽出される語が、Web 検索を行う場合に検索質問を構成する主要なキーワードとなる。

内容語の抽出は次のような方法によって行う。

1. 関連情報の検索対象のメール中に疑問文が含まれる場合にはその疑問文を、含まれない場合にはメールに含まれる文すべてを形態素解析する。
2. 形態素解析した結果、名詞、形容詞、形容動詞と判断されたものを内容語として抽出する。
3. 代名詞などの指示語が含まれる場合には、その語が指す内容を、返信元のメールの内容語を参照して、その語を内容語として抽出する。

4.2 削除対象語

削除対象語とは、内容語とは逆に、関連情報を検索する場合に条件としては不適当な語の集合である。この削除対象語は、メール中に比較や否定の表現が含まれる場合に抽出を行う。

削除対象語の抽出は次のような方法によって行う。

1. メール中の文章に比較や否定などの削除対象語を含むと考えられる表現が含まれるかを判定する。
2. 削除対象語を含む表現があると判定されると、さらに、文中のどの部分に含まれるかを典型的表現と照合す

ることで推測する。

3. 削除対象語が含まれると推測される部分を形態素解析する。
4. 形態素解析の結果、名詞、形容詞、形容動詞と判断されたものを削除対象語として抽出する。
5. 代名詞などの指示語が含まれる場合は、その語が指す内容を、返信元のメールの内容語を参照して、その語を削除対象語として抽出する。

4.3 話題語

携帯端末でのメールを対象とする場合、メールの長さが短いと内容語が十分に抽出できないことがある。そうすると、関連情報の検索の精度が落ちてしまうことが考えられるため、それを防ぐために、本研究では話題語という要素をメールから抽出する。

話題語の抽出は次のようにして行う。

1. 数通の連続したメールについて、それぞれのメールの内容語と削除対象語を抽出する。
2. 抽出された内容語について、複数のメール中で内容語として抽出されているものを話題語として抽出する。
3. 2で抽出した話題語から、削除対象語に含まれている語を話題語から除く。

5. 対話の種類分類

メールの関連情報を検索しようとする場合、同じような検索のキーワードを利用した場合でも、メールの対話の流れによってユーザーの欲しい情報の種類は異なると考えられる。そこで、与える情報の種類によって、対話の種類を次の2つに分類する。

- オブジェクト検索型対話
- プロパティ検索型対話

本章では、対話の種類とその分類手法について述べる。

5.1 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、

- 「京都のラーメン屋ならどこがおいしい？」
- 「春に観光に行くならどこがお勧め？」

といった発言がメール中に含まれると、オブジェクト検索型対話として判別される。前者の発言では「京都」「おいしい」というラーメン屋の場所、感想を示す条件を挙げることで、答えとしてこれらの条件を満たすラーメン屋の名前という具体的な情報を得ようとしている。一方後者の発言では「春」「観光」「お勧め」という条件を挙げることで、答えとしてこれらを満たす観光地の地名という具体的な情報を得ようとしている。

5.2 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、その評判や感想など（プロパティ）を答えに求めようとする対話型である。オブジェクト検索型対話と対を成す対話型と考えることもできる。例えば、

- 「〇〇って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？」

といった発言がメール中に含まれると、プロパティ検索型対話として判別される。前者の発言では「〇〇」という具体的なラーメン店の名前を条件として挙げることで、答えとして

この店では出されるラーメンの味といった評判, 感想などの情報を得ようとしている. 一方後者の発言では, 「冬」「京都」という条件を挙げることで, 答えとして冬の京都の評判, 印象などの情報を得ようとしている.

5.3 対話の判別

オブジェクト検索型対話とプロパティ検索型対話を判別する方法を述べる.

オブジェクト検索型対話と判別されるべき文とプロパティ検索型対話と判別されるべき文を比べてみると次のような特徴が見られる.

- オブジェクト検索型対話と判別されるべき文章には一般名詞や形容詞などの修飾語といった具体性が低い語が多く出現する傾向にある.
- プロパティ検索型対話と判別されるべき文章には固有名詞のような具体性の高い語が多く出現する傾向にある.

そのため, つぎのようにして判別を行っている.

1. メールから抽出された内容語の品詞分類に基づいて, その文章の具体度の値を計算する.
2. 計算された具体度の値とある閾値を比較し, 値が閾値を超えない場合にはオブジェクト検索型対話と判別し, 超える場合にはプロパティ検索型対話と判別する.

また, 1で利用している具体度の計算方法は次のとおりである.

x_1, x_2, \dots, x_n をメールから抽出された内容語, $f(x)$ を内容語 x の具体度とする. この $f(x)$ の値は, x が固有名詞や地名を表す名詞であると大きくなり, 一般名詞や形容詞, 形容動詞になると小さいものとする. そして, 閾値 C として

$$\frac{\sum_{i=1}^n f(x_i)}{n} \begin{cases} < C : \text{オブジェクト検索型対話} \\ \geq C : \text{プロパティ検索型対話} \end{cases}$$

のようにして計算された値と閾値と比較することによって, オブジェクト検索型対話かプロパティ検索型対話を判別している.

6. 関連情報検索と提示

5章までで, メールの関連情報を提示するために必要となるメールの内容などの抽出について述べた. 本章では, メールの対話から抽出, 判断された内容語, 削除対象語, 話題語, および対話型を利用して, 関連情報を検索しユーザーに提示する手法について述べる.

6.1 情報の検索

ここでは, メールが送信されてから, 関連情報を検索するまでの流れを述べる (図2).

関連情報を検索しようとするメールを図2中の①のメールとすると,

1. メール①から内容語と削除対象語を抽出する.
2. メール①より前に送信された数通の一連のメール (図中では②~④) から, 個々のメールの内容語と抽出した上で, 話題語を抽出する.
3. 抽出された関連情報検索対象のメールの内容語と, その前に送信された一連のメールの話題語のうちで, 削除対象語に含まれていない語を抽出する.
4. 3で抽出された語すべてを含むことを条件として Web 検索を行う.

5. ユーザーに結果を提示する.

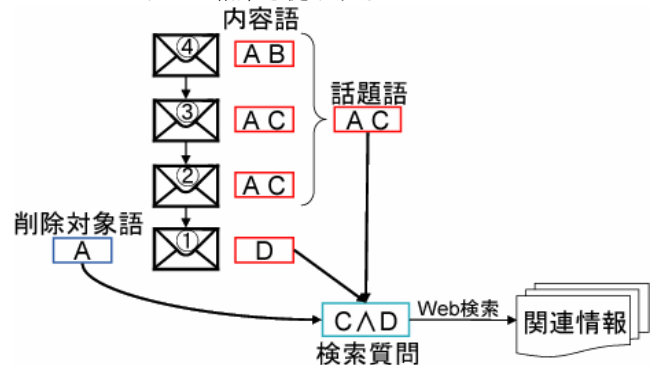


図2 検索質問生成の流れ
Fig.2 Query Generation

6.2 情報の提示の手法

関連情報として検索された結果は, Google などといった既存の検索エンジンを用いた場合には特に, ユーザーの欲しい情報と検索エンジンの結果の順位などが大きくかけ離れる場合が多くある. そこで本論文では, 5章で述べた対話型に基づいて, 得られた検索結果の提示の優先度の変化をさせる.

オブジェクト検索型対話と判別されている場合, 関連情報としてユーザーが要求していると考えられる情報に多く含まれるべきものとしては, ある店名などを示すような固有名詞が挙げられる. そこで, オブジェクト検索型対話の場合では検索結果の優先度を, 検索結果として得られた Web ページを個々に解析して, 固有名詞の出現頻度が高い順に優先的にユーザーに提示する.

プロパティ検索型対話と判別されている場合, 関連情報としてユーザーが要求していると考えられる情報には, 性質・状態を表すような形容詞, 形容動詞が多く出現すると考えられる. そこで, プロパティ検索型対話の場合では検索結果の優先度を, 検索結果として得られた Web ページ中の形容詞, 形容動詞の出現頻度が高い順に優先的にユーザーに提示する.

7. 予備実験

前章までに述べたことの有効性を検証するために, PC 上で簡易的な予備実験を行った. ただし, ここで用いたシステムでは, 6.2 節で述べた対話型による検索結果の優先度付けについては実装していない.

- 入力1: 京都でおいしいラーメン屋ってどこか知らない?
- 入力2: それなら「東龍」でしょう。
- 入力3: 「東龍」よりは「高安」がいいって聞いたけど...

- ・ラーメン屋「高安」のラーメンの情報
- ・京都のラーメン屋のリスト



図3 実行例
Fig.3 Example

この予備実験では, 形態素解析に茶筌[4]を, Web 検索のために Google Web API[5]を利用し, 携帯端末上で送信する

と想定されるメールの文章を順次入力することで、最大3文までさかのぼって文章の内容を解析し検索キーワードを抽出し、それを利用して関連する情報を Web から検索しユーザーに得られた Web ページへのリンクを提示するシステムを作製した。実行例を図3で示す。図3中の Web ページは入力1から順次入力して、入力3についての関連情報を検索したときに結果として得られたものである。

今回の予備実験を通して判明したことは次のような点である。

- 現在の手法の場合、メールの長さが長すぎると内容語が多く抽出されるために、検索結果が得られない場合がある。また逆に短すぎる場合、この場合は検索結果が多く得られるため、この場合も有用な検索結果が得られない。
- メール交換の回数が多くなると、徐々に省略される語句が増えてくるため、そうなると話題語が抽出されなくなってしまい、関連情報検索の精度の低下を招く。
- 指示語の内容を補完するために、返信元のメールを参照するという手法はうまく機能していた。
- 対話型の判別では、固有名詞、地名、一般名詞、形容詞・形容動詞の4種類に分けて具体度の値を設定して行ったが、妥当な判別結果を得られたことが多かった。しかし、内容語に抽出されない要素（助詞などの語や話の流れ）によって対話型が変わってしまうような文章について、判別がうまくいかないことがあった。例えば「京都のラーメンってどこがおいしい？」という文については、「おいしいラーメンを出す京都の店」を知りたいのか、「京都で一般的に食べられるラーメンの特徴」を知りたいのかで対話型が異なるが、現時点では判別不能である。

今回は PC 上で簡易的な予備実験を行ったため、今後は携帯端末上で同様に実験を行い、実用性などについて検証する必要がある。しかし、今回の実験で得られた感触としては、特に、情報のユーザーへの提示の方法を改善することでかなり使えるシステムになるとと思われる。

8. まとめと今後の課題

本研究では、携帯端末からあまり利用されていない Web 検索について、携帯端末上でも比較利用される電子メールの交換に着目し、電子メールの交換内容に関連した情報を自動的に推測、Web 上から検索しユーザーに提示するシステムを提案した。このようなシステムを利用することで、携帯端末上でも容易に Web 上に存在する有用な情報を利用することが可能になると考えている。

これからの課題としていくつかの点が挙げられる。まず、「情報をユーザーに効果的に提示する手法の検討」である。現段階では、検索された Web ページをそのままの状態でも、しかも、ユーザーが情報を必要としているかどうかにかかわらず一方的に提示するようになっている。しかし、携帯端末という制約のもとではこれは現実的でない。そこで、ユーザーの必要な情報だけをうまく携帯端末上で見せる手法を検討する必要がある。また、「語句の省略などの対応する手法の検討」も挙げられる。現時点では、代名詞などの指示語の補完については行っているものの、語句の省略などについては検討がなされていない。関連情報の検索精度を向上させるためにも、語句の省略に対応する仕組みが必要である。そし

て、「携帯端末で使用した場合の有用性などの実証」を行う必要がある。今後はこれらの課題について考察すると同時に、携帯端末においてメールの交換を情報収集だけでなく、例えば、メールでの対話内容を Web アノテーションとして利用するような、携帯端末からの情報発信の手法についても検討する予定である。

【謝辞】

本研究の一部は、平成14～16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」(14年度課題番号:14208036, 代表:田中克己), 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), および, 21世紀 COE プログラム「知識社会基盤構築のための情報学拠点形成」(代表:上林弥彦)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ィール抽出と動的情報探索」, 情報処理学会研究報告, Vol.2002, No.67, 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", Proceeding of 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, Vol.2001, No.69, 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>
- [5] Google Web API, <http://www.google.com/apis/>

赤星 祐平 Yuhei AKAHOSHI

京都大学大学院情報学研究科修士課程在学中。2003 京都大学工学部情報学科卒業。日本データベース学会学生会員。

小山 聡 Satoshi OYAMA

京都大学大学院情報学研究科助手。2002 京都大学大学院情報学研究科修士課程修了。博士(情報学)。情報検索, データマイニングなどの研究に従事。日本データベース学会, 電子情報通信学会, 人工知能学会, ACM, AAAI 各会員。

角谷 和俊 Kazutoshi SUMIYA

京都大学大学院情報学研究科助教授。1998 神戸大学大学院自然科学研究科修士課程修了, 博士(工学)。マルチメディアデータベース, データ放送に関する研究に従事。情報処理学会, 日本データベース学会, ACM, IEEE Computer Society, 映像情報メディア学会各会員。

田中 克己 Katsumi TANAKA

京都大学大学院情報学研究科教授。1976 京都大学大学院修士課程修了。博士(工学)。主にデータベース, マルチメディアコンテンツ処理の研究に従事。情報処理学会, 日本データベース学会, 人工知能学会, ACM, IEEE Computer Society 等各会員。