

大規模テキスト知識ベースに基づく自動質問応答 —ダイアログナビ—

清田 陽司[†] 黒橋 禎夫^{†,‡} 木戸 冬子^{†,‡}

本論文では、大規模テキスト知識ベースに基づく対話的自動質問応答システム「ダイアログナビ」について述べる。本システムは、2002年4月からWWW上で一般公開し、パーソナルコンピュータの利用者を対象としてサービスを行っている。実世界で用いられる質問応答システムにおいては、ユーザ質問の不明確さや曖昧性が大きな問題となる。本システムは、「エラーが発生した」のような漠然とした質問について、対話的に聞き返しを行うことによってユーザが求める答えにナビゲートする。聞き返しの方法としては、頻繁になされる漠然とした質問に対する聞き返しの手順を記述した対話カードを用いる手法と、自動的に聞き返しの選択肢を編集して提示する手法を組み合わせて用いている。また、適切なテキストを正確に検索するために、ユーザ質問のタイプ、同義表現辞書や、日本語の文の係り受け関係などを利用している。

キーワード: 対話システム, 情報検索, 質問応答, テキスト知識ベース

Dialog Navigator : A Question Answering System based on Large Text Knowledge Base

YOJI KIYOTA[†], SADA O KUROHASHI^{†,‡} and FUYUKO KIDO^{†,‡}

This paper describes a dialog based QA system, Dialog Navigator, which can answer questions based on large text knowledge base. This system is targeted at users of personal computers. We released the system on the WWW in April 2002. In real world QA systems, vagueness of questions is a big problem. Our system can navigate users to the desired answers using the following methods: asking users back with dialog cards, and description extraction of each retrieved text. Another feature of the system is that it retrieves relevant texts precisely, using question types, synonymous expression dictionary, and modifier-head relations in Japanese sentences.

KeyWords: *Dialogue System, Information Retrieval, Question Answering, Text Knowledge Base*

1 はじめに

何かを調べたいとき、一番よい方法はよく知っている人(その分野の専門家)に直接聞くことである。多くの場合、自分の調べたいこととその答えの間には、具体性のズレ、表現のズレ、

[†] 東京大学大学院情報理工学系研究科, Graduate School of Information Science and Technology, The University of Tokyo

[‡] 科学技術振興事業団 さきがけ研究 21, PRESTO, JST

^{†,‡} マイクロソフト株式会社, Microsoft Co., Ltd.

背景の認識の不足などがあるが、専門家は質問者との対話を通してそのようなギャップをうめてくれるのである。

現在、WWWなどに大規模な電子化テキスト集合が存在するようになり、潜在的にはどのような質問に対してもどこかに答えがあるという状況が生まれつつある。しかし、今のところWWWを調べても専門家に聞くような便利さはない。その最大の原因は、上記のようなギャップを埋めてくれる対話的な能力が計算機にないためである。例えば、ユーザがWWWのサーチエンジンに漠然とした検索語を入力すると多くのテキストがヒットしてしまい、ユーザは多大な労力を費して適切なテキストを探さなければならない。

このような問題は、ドメインを限定し、ユーザが比較的明確な目的を持って検索を行う場合でも同様である。我々は予備調査として、マイクロソフトが提供している自然言語テキスト検索システム「話し言葉検索」¹の検索ログを分析した。その結果、全体の約3割の質問はその意図が不明確であることがわかった。このような曖昧な質問に対しては多くのテキストがマッチしてしまうので、ユーザが検索結果に満足しているとはいえない。この問題を解決するためには、「曖昧な質問への聞き返し」を行うことが必要となる。

すでに実現されている情報検索システムには、大きく分けてテキスト検索システムと質問応答システムの2つのタイプがある。前者は質問キーワードに対して適合するテキスト(のリスト)を返し、後者は質問文に対してその答えを直接返す。しかし、曖昧な質問を行ったユーザを具体的なテキストまたは答えに導く必要性は両者に共通する。以下では、「曖昧な質問への聞き返し」に焦点をあてて、過去の研究を俯瞰する(表1)。

テキスト検索システムにおいて、質問とテキストの具体性のギャップを埋めるために聞き返しを行う方法としては、以下の手法が提案されてきた。

- テキストによる聞き返し

検索結果から適合テキストをユーザに判定させ、それを検索式の修正に反映させる手法は、SMARTシステムなどで実験が行われている(Rocchio 1971)²。Google³などのWWWサーチエンジンでは、検索結果からテキストを1個選んで、その関連テキストを表示させることができるが、この方法もユーザによる適合テキストの判定とみなすことができる。

- 関連キーワードによる聞き返し

検索結果から、ユーザが入力したキーワードに関連するキーワードを抽出し、選択肢として提示するシステムとしては、RCAAU(西村, 河野, 長谷川 1996)、DualNAVI

1 <http://www.microsoft.com/japan/enable/nlsearch/>

2 このようにユーザが適合テキストを選ぶ方法は、「適合性フィードバック」とよばれている。しかし、ユーザに聞き返しを行って何らかの情報をえること全体が、広い意味での適合性フィードバックであるので、ここではその用語は用いていない。

3 <http://www.google.com/>

表 1 情報検索の種々のタイプ

手法 / システム	ユーザ質問	出力	聞き返しの媒体	規模
一般的なテキスト検索システム	キーワードのリスト	テキストのリスト	×	
テキストによる聞き返し (SMART, WWW サーチエンジン)	キーワードのリスト	テキストのリスト	テキスト	
関連キーワードによる聞き返し (RCAUU, DualNAVI, Excite)	キーワードのリスト	テキストのリスト	キーワード	
テキストと関連キーワードによる聞き返し (THOMAS)	キーワードのリスト	テキスト	テキストと キーワード	
クラスタリング (Scatter/Gather, WebSOM)	キーワードのリスト	テキストのリスト	クラスタ (キーワード or テキストで表現)	
人工言語による知識体系の利用 (UC)	自然言語	自然言語 (答え)	自然言語	×
FAQ テキストの利用 (FAQ Finder)	自然言語	自然言語 (答え)	×	
ドメイン独立テキストの利用 (TREC QA / NTCIR QAC)	自然言語	自然言語 (答え)	×	
京都大学ヘルプシステム	自然言語	自然言語 (答え)	自然言語	
ダイアログナビ	自然言語	自然言語 (状況説明文)	自然言語	

(Takano, Niwa, Nishioka, Iwayama, Hisamitsu, Imaichi, and Sakurai 2000), Excite⁴などがある。

- テキストと関連キーワードを組み合わせた聞き返し

THOMAS (Oddy 1977) は、ユーザの情報要求を「イメージ」とよばれるキーワード集合として保持し、テキスト1個と関連キーワードを併せて提示してそれらの適合性をユーザに判定させるプロセスを繰り返すことで、「イメージ」を徐々に具体化させようとするシステムである。ただし、1970年代に提案されたシステムであり、小規模なテキスト集合にしか適用できない。

- クラスタリング

検索されたテキストをクラスタリングし、クラスタを選択肢として提示するシステムとしては、Scatter/Gather (Hearst and Pedersen 1996), WEBSOM (Lagus 2000) などがある。これらのシステムでは、各クラスタは、それに属するテキストのリストや、代表的なキーワードのリストとして表現されている。

これらのシステムの聞き返しの媒体は、いずれもキーワードまたはテキストのレベルである。しかし、キーワードは抽象化されすぎており表現力がとぼしく、逆にテキストは具体的すぎるため、聞き返しの媒体としては必ずしも適切ではない。

一方、質問応答システムとしては、1980年代に UC (Wilensky, Arens, and Chin 1984) など

⁴ <http://www.excite.com/>

のシステムが研究された。これらのシステムは、ユーザの意図が曖昧な場合に自然言語による聞き返しを行う能力を備えていたが、そのためには人工言語で記述された、システムに特化した知識ベースが必要であった。しかし、十分な能力をもつ人工言語の設計の困難さ、知識ベース作成のコストなどの問題から、このような方法には明らかにスケーラビリティがない。

1990年代になって、電子化された大量の自然言語テキストが利用可能になったことから、自然言語テキストを知識ベースとして用いる質問応答システムの研究が盛んになってきた。インターネットのニュースグループのFAQファイルを利用するシステムとしては、FAQ Finder (Hammond, Burke, Martin, and Lytinen 1995)がある。また最近は、構造化されていないドメイン独立のテキスト(新聞記事やWWWテキスト)を用いた質問応答システムの研究が、TREC QA Track (NIST and DARPA 2001)やNTCIR QAC (Fukumoto, Kato, and Masui 2002)において盛んに行われている(Harabagiu, Moldovan, Pasca, Mihalcea, Surdeanu, Bunescu, Girju, Rus, and Morărescu 2001; Ferret, Grau, Plantet, Illouz, Jacquemin, Masson, and Lecuyer 2001; Murata, Utiyama, and Isahara 2002; Kawahara, Kaji, and Kurohashi 2002)。しかし、これらのシステムはユーザの質問が具体的であることを前提にして、1回の質問に対して答えを1回返すだけであり、曖昧な質問に対して聞き返しを行う能力は備えていない。

京都大学総合情報メディアセンターのヘルプシステム(Kurohashi and Higasa 2000)は、自然言語で記述された知識ベースとユーザ質問の柔軟なマッチングに基づいて、曖昧な質問に対して自然言語による聞き返しを行うことができるシステムである。しかしここでは、記述の粒度をそろえ、表現に若干の制限を加えた知識ベースをシステム用に構築しており、「曖昧な質問への聞き返し」のプロトタイプシステムという位置づけが適当である。

これに対して、本論文では、既存の大規模なテキスト知識ベースをもとにして、自然言語による「曖昧な質問への聞き返し」を行い、ユーザを適切なテキストに導くための方法を提案する。具体的には、パーソナルコンピュータのWindows環境の利用者を対象とした自動質問応答システム「ダイアログナビ」を構築した(図1)。本システムの主な特徴は以下の通りである。

- 大規模テキスト知識ベースの利用
マイクロソフトがすでに保有している膨大なテキスト知識ベースをそのままの形で利用する。
- 正確なテキスト検索
ユーザの質問に適合するテキストを正確に検索する。そのために、質問タイプの同定、同義表現辞書による表現のずれの吸収、係り受け関係への重みづけなどを行っている。
- ユーザのナビゲート
ユーザが曖昧な質問をしたとき、対話的に聞き返しを行うことによってユーザを具体的な答えにナビゲートする。聞き返しの方法としては、対話カードと状況説明文の抽出の2つの方法を組み合わせて用いる。どちらの方法が用いられても、システムは具体的な



図 1 ダイアログナビのユーザインタフェース

フレーズを聞き返しの選択肢として提示する。

図 1 の例では、「エラーが発生する」という漠然とした質問に対して 2 回の聞き返しを行ってユーザの質問を対話的に明確化させた後、知識ベースを検索してその結果を提示している。その際、ユーザの質問をより具体化させるような部分を検索されたテキストから抽出して提示している。

本論文では、このような対話的質問応答を可能とするためのシステムを提案する。まず 2 節において、システムの構成を示す。つづいて、3 節では正確なテキストの検索を行うための手法を、4 節ではユーザのナビゲートを実現するための手法を、具体的に提案する。さらに 5 節において、提案手法を実装したシステム「ダイアログナビ」を公開運用して得られた対話データベースの分析結果を、提案手法の評価として示す。最後に 6 節で本論文のまとめを述べる。

表 2 ダイアログナビで用いるテキスト知識ベース

知識ベース	件数	文字数	マッチング対象
用語集	4,707	700,000	見出し (1 文)
ヘルプ集	11,306	6,000,000	タイトル (1 文)
サポート技術情報	23,323	22,000,000	文書全体 (複数文)

音声認識ソフトウェアがインストールされた環境でページ違反が発生する

最終更新日: 1999/08/18
 文書番号: J049655

この資料は以下の製品について記述したものです。

- Microsoft(R) Internet Explorer Version 5 (以下 Internet Explorer 5)
- Microsoft(R) Windows 98 (以下 Windows 98)

概要
 この資料は、Windows 98 上に Internet Explorer 5 がインストールされた環境で、音声認識ソフトウェアが起動されていると、Internet Explorer 5 を起動した際に、ページ違反が発生する現象について説明したものです。

内容
 以下の条件を満たすときに Internet Explorer 5 を起動すると、ユーザー補助プログラムの OLEACC.DLL が不正なメモリ領域を参照することにより、ページ違反が発生する場合があります。

- Windows 98 にユーザー補助プログラムがインストールされている
- 音声認識ソフトウェアが起動している

回避方法
 Windows 98 システムアップデートモジュールをインストールします。システムアップデートモジュールには、新しい OLEACC.DLL が含まれており、この不具合が修正されていることを確認しております。これは Windows 98 Service Pack 1 に含まれるモジュールとなっており、Windows Update からダウンロードすることができます。

入手方法

- (1) [スタート]メニューから [Windows Update] をクリックします。
- (2) 画面の指示に従い "Windows Update へようこそ" が表示されたら、"製品の更新" をクリックします。
- (3) "ソフトウェアの選択" 画面にて、"Windows 98 System Update" にチェックをつけ、"ダウンロード" ボタンを押します。
- (4) 画面の指示に従い、モジュールをインストールします。

図 2 マイクロソフト・サポート技術情報の例

2 ダイアログナビの構成

ダイアログナビにおいて使用するリソースを以下に示す。

- **知識ベース**
 マイクロソフトがすでに一般に公開しているテキスト知識ベースをそのまま用いる。その種類と規模を表 2 に示す。また、知識ベースのうちサポート技術情報に含まれるテキストの例を図 2 に示す。
- **同義表現辞書 (3.2.1項, 図 3)**
 ユーザ質問文と知識ベースの間の表現のずれを吸収するために、同義語や同義フレーズをグループ化した辞書を用いる。現在、ダイアログナビの同義表現辞書には、919 グループの同義表現が存在し、3512 語・217 フレーズが登録されている。
- **上位・下位語辞書 (3.2.2項, 図 4)**
 上位・下位の関係にある語 (「ブラウザ」と「Internet Explorer」など) を関係づけた辞書を用いる。現在、200 語が登録されている。

[発生する]	発生する，起きる，おきる，起こる，おこる
[読む]	読む，よむ，読める，よめる，読み込む，よみこむ，読み込める，よみこめる
[メール]	メール，メイル，電子メール，電子メイル，Mail，E-Mail
[メールを読む]	メールを読む，メールを受信する，メールを見る，メールを受ける，メッセージを受信する，メッセージを受ける
[パソコンを起動する]	パソコンを起動する，Windows を起動する，電源を入れる，電源をオンする，ブートする，パソコンを立ち上げる，スイッチを入れる

図 3 同義表現辞書の例

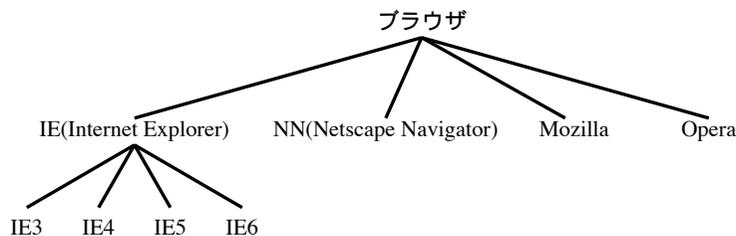


図 4 上位・下位語辞書の例

```

<CARD>
<ID> エラー
<UQ> エラーが発生する
<REPLY> エラーはいつ発生しますか？
<SEL action=CARD card_id="エラー/Windows 起動中"> Windows 起動中
<SEL action=CARD card_id="エラー/ログイン時"> ログイン時
<SEL action=CARD card_id="エラー/印刷時"> 印刷中
...
</CARD>
    
```

```

<CARD>
<ID> エラー/Windows 起動中
<UQ> Windows の起動中にエラーが発生する
<REPLY> あなたがお使いの Windows を選んでください。
<SEL action=RET phrase="Windows 95 の起動中にエラーが発生する"> Windows 95
<SEL action=RET phrase="Windows 98 の起動中にエラーが発生する"> Windows 98
...
<SEL action=RET phrase="Windows XP の起動中にエラーが発生する"> Windows XP
</CARD>
    
```

図 5 対話カードの例

- 対話カード (4.2節，図 5)

曖昧なユーザ質問文のうち典型的なものに対して，どのような聞き返しを行うかを記述したカードを利用する。

ダイアログナビの内部の処理と、ユーザとの対話の関係を図 6 に示す。基本的な流れは、対話カードに基づくユーザとの対話によってユーザの質問が具体化され (図 6 の左側のループ)、具体化された質問によって知識ベースが検索され (右側の処理へ移行)、検索結果が自動編集され選択肢の形でユーザに提示される。ユーザの最初の質問が具体的な場合は、対話カードとはマッチせず右側の処理へ移行し、はじめから知識ベースの検索結果が提示される。

図 6 中の各モジュールの働きは以下の通りである (詳細は次節以降に示す)。

- 入力解析モジュール
質問文を 3 種類の質問タイプ (Symptom 型, How 型, What 型) に分類し、質問文の内容表現を抽出する。さらに、構文解析、キーワードと同義表現の抽出などを行う。
- テキスト検索モジュール
対話カードおよび知識ベース (以下、これらを総称してテキストという) とユーザ質問文のマッチングを行い、スコアの高いテキストを返す。マッチングの際には、同義表現辞書、上位・下位語辞書を用いて表現のずれを吸収する。
- 状況説明文抽出モジュール
知識ベース中のユーザ質問文とマッチした文の、マッチした部分の周囲を抽出することによって、ユーザにとって簡潔でわかりやすい選択肢を提示する。

3 テキストの検索

質問応答システムにおいてまず重要なことは、質問の答えを含むと思われるテキストを十分な精度で検索できることである。そのために、質問タイプとプロダクト名による知識ベースの絞り込みを行う。また、表現のずれを吸収するために同義表現辞書 (図 3) を利用したマッチングを行う。さらに、スコア計算において、「ファイル 開く」のような係り受け関係に加点することによって、検索の精度を向上させる (Zhai, Tong, Milic-Frayling, Evans 1996)。

3.1 マッチングの前処理

ユーザ質問文とテキスト内の文 (以下、テキスト文という) は、それぞれ構文解析を行って文節単位の係り受け構造に変換した上でマッチングを行う。この節では、マッチングを行うまでの前処理についてまとめる。

3.1.1 構文解析とキーワード抽出

ユーザ質問文とテキスト文の両者について、JUMAN (黒橋, 長尾 1999)、KNP (Kurohashi and Nagao 1994) によって構文解析を行い、各文節に含まれるキーワードを抽出する。JUMAN において、普通名詞・固有名詞・人名・地名・組織名・数詞・動詞・形容詞・形容動詞・副詞・

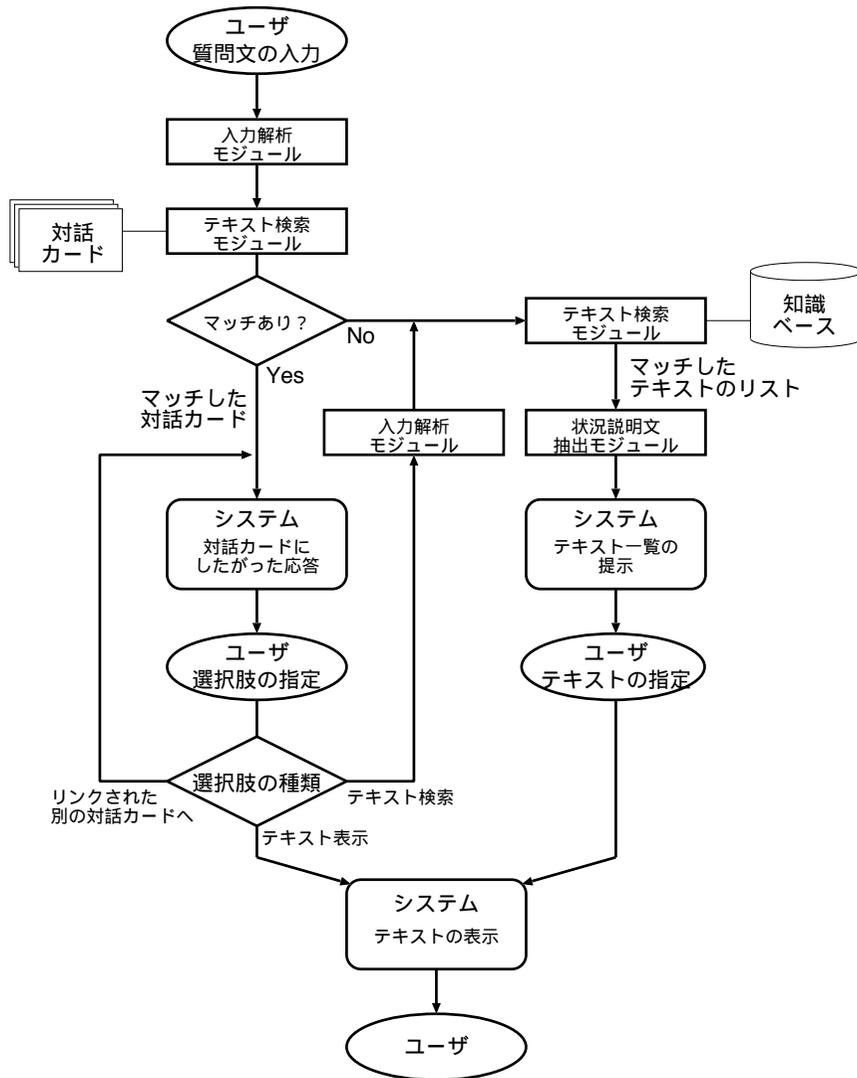


図 6 ダイアログナビのフローチャート

カタカナ・アルファベットと解析された語の原形をキーワードとみなす。ただし、一般的な語彙「する」「ある」「行う」「おこなう」「行く」「いく」「なる」「下さる」「くださる」「ごさいます」「できる」「出来る」は、キーワードとしない。

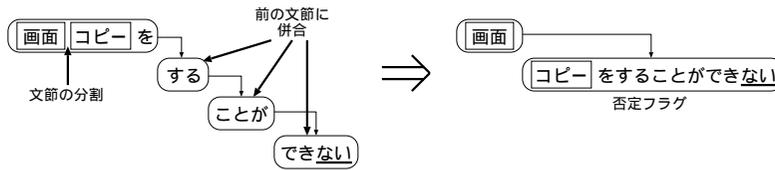


図 7 文節の分割・併合処理と否定フラグの付与

表 3 「話し言葉検索」の質問文タイプ

質問文タイプ	説明	質問文の例 (文末表現パターン)	割合
What 型	用語の意味や定義などをたずねる質問	～って何ですか、～の説明をして、～の意味を教えて	約 10%
How 型	操作の方法などをたずねる質問	～方法を教えて、～にはどうしたらいいの、～の使い方	約 35%
Symptom 型	遭遇している問題や症状を述べ、その解決策をたずねる質問	～してしまう、～が使えません、～ができない	約 50%
その他			約 5%

3.1.2 文節の分割・併合処理

マッチングのスコアを計算する際、KNP が出力した文節をそのまま用いることには問題がある。例えば、「画面をコピーできない」は 2 文節、「画面コピーをすることができない」は 4 文節と解析されるが、両者は同じことを表現している。これを適切に扱うためには、両者の単位をそろえる必要がある。

本システムは、下記のルールに従って文節を分割・併合する (図 7)。

- (1) 複数のキーワードを含む文節は、1 キーワード毎に分割する。分割された隣り合う文節同士は、係り受けの関係にあるものとする。ただし、カタカナ語・アルファベット・数詞が隣接している箇所では分割しない。このような語同士が隣接する場合は、「ウィンドウズ 98SE」のようにプロダクト名などを表していることが多いからである。
- (2) 「(～) について」「(～) こと」などの複合辞・形式名詞・副詞的名詞からなる文節、キーワードを含まない文節は、直前の文節に併合する。

3.1.3 否定フラグの付与

ユーザ質問文とテキスト文のマッチングの際に否定表現のバリエーションを吸収するために、文節にフラグを付与する。具体的には、形容詞「ない」、助動詞「ぬ」、または形容動詞「不可能だ」を含む場合に否定フラグを付与する (図 7 右)。

3.1.4 ユーザ質問文のタイプ推定と文末表現の削除

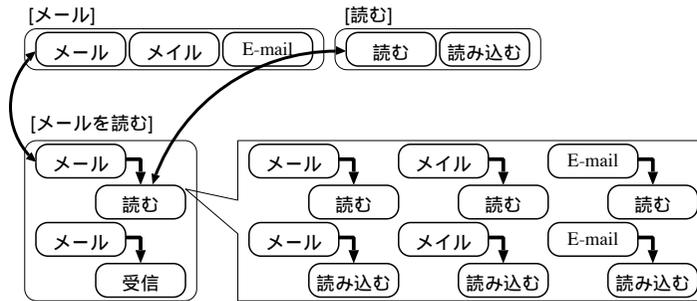


図 8 同義表現辞書の再帰的展開

「話し言葉検索」の検索ログを分析した結果，表 3 に示すようにユーザの質問には主に 3 つのタイプが存在することがわかった．本システムでは，表 3 の文末表現パターンを用いて，ユーザ質問文の質問タイプ (What 型，How 型，Symptom 型，タイプなしのいずれか) を推定する．また，文末表現パターンのうち，「～って何ですか」「～方法を教えて」のようにテキスト検索においてノイズとなるものについては，ユーザ質問文から削除する．

3.2 表現のずれの吸収

適切なテキストを検索するためには，ユーザ質問文とテキストの間の表現のずれが大きな問題となる．本システムでは，同義表現辞書と上位・下位語辞書を用いることによってこの問題に対処する．

3.2.1 同義表現辞書

表現のずれは語のレベルだけでなく，「パソコンを起動する」「Windows を起動する」「電源を入れる」のように，2 文節以上のフレーズレベルにおいても多数存在する．そこで，同義語だけでなくフレーズレベルのものも含んだ同義表現をグループ化した同義表現辞書を作成し，これを用いて同義表現のマッチングを行う．

同義表現辞書の例は図 3 に示した．本辞書の作成は，「話し言葉検索」のログを解析し，頻出する同義表現をグループ化することによって行った．また，和語動詞（「戻る」など）の可能形（「戻れる」）や読み（「もどる」「もどれる」）も同義表現として登録した．

なお，同義表現辞書には再帰的な関係が含まれているため，これをあらかじめ展開しておく．図 8 においては，「メールを読む」には 2 つのキーワード「メール」「読む」が含まれるが，「メール」には同義語「メイル」「E-mail」が存在し，「読む」には同義語「読み込む」が存在する．この場合，「メールを読む」というフレーズを $3 \times 2 = 6$ 通りに展開する．

マッチングの際には，ユーザ質問文とテキストの両者について，同義表現辞書を調べて，そ

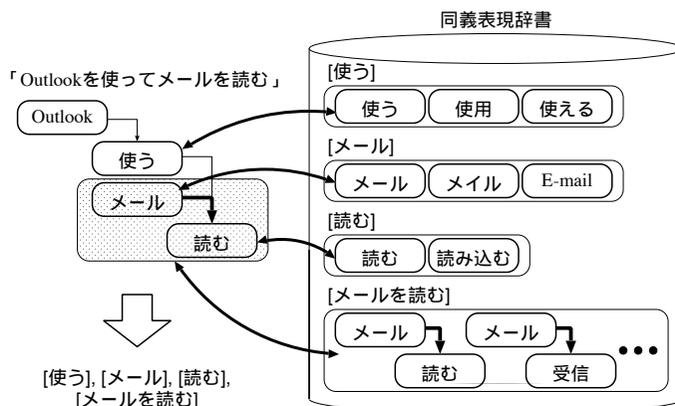


図 9 ユーザ質問文と同義表現辞書の照合

ここに含まれる同義表現グループを抽出し、同一グループのものがあればマッチするとみなす。ただし、3.3節で述べるように、テキストについてはあらかじめ同義表現グループを抽出しておく。

図 9 に、ユーザ質問文と同義表現辞書の照合の例を示す。この例では、4 つの同義表現グループ [使う]、[メール]、[読む]、[メールを読む] が抽出される。

3.2.2 上位・下位語辞書

同義表現辞書ではうまく扱えない表現のずれも存在する。例えば、「ブラウザ」 \iff 「IE6」、 \iff 「IE5」といった表現のずれに対して、「ブラウザ」「IE5」「IE6」をすべて同義語として扱うことは問題である。なぜなら、「IE5」に関する質問に対して、「IE6」に関するテキストを示すことは適切でないからである。

そこで、図 4 に示すような上位・下位語辞書を作成し、テキストに現れるキーワードの上位語・下位語を、キーワードと同様に扱うことによってこの問題に対処する。例えば、「IE6」がテキストに現れる場合はその上位語「IE」「ブラウザ」もキーワードとして扱い、「IE」がテキストに現れる場合はその上位語「ブラウザ」と下位語「IE3」「IE4」「IE5」「IE6」もキーワードとして扱う。ユーザ質問文についてはこの扱いを行わないことによって、「IE5」と「IE6」がマッチすることが避けられる。

3.3 転置インデックスの利用

テキストを高速に検索するために、前もってテキストからキーワード・同義表現グループの抽出と、キーワードの上位・下位語の展開を行い、転置インデックスを作成しておく。

本システムは、ユーザ質問文から抽出されたキーワードと同義表現グループについて転置イ

表 4 質問タイプによるテキスト集合の絞り込み

テキスト集合	質問タイプ			
	What 型	How 型	Symptom 型	タイプなし
用語集 (What 型)	○			○
ヘルプ集 (How 型)	○	○		○
サポート技術情報 (Symptom 型)	○		○	○
	○	○		○
	○	○	○	○

ンデックスを参照し，1 個以上のキーワードまたは同義表現が一致するテキストを，次節で述べる質問タイプ・プロダクト名による絞り込みの対象とする．

3.4 知識ベースの絞り込み

テキスト検索の精度を向上させるために，質問タイプとプロダクト名による知識ベースの絞り込みを行う．

3.4.1 質問タイプによる絞り込み

テキスト検索モジュールは，入力解析モジュールによって推定された質問パターンにもとづいて，表 4 に示すようにテキスト集合を絞り込む．原則として，用語集は What 型，ヘルプ集は How 型の質問に対応させる．サポート技術情報については Symptom 型・How 型を示すタグが付与されているので，これを利用する．

なお，What 型の質問については必ずしも用語集を用いて答えればよいとは限らない．例えば「コントロールパネルについて教えて」とような質問は What 型に分類されるが，用語の定義ではなく操作方法などについて聞いていると解釈することもできる．よって，全てのテキストを検索対象とした上で，複数の知識ベースのテキストがユーザ質問とマッチした場合には用語集のテキストを最初に提示する．

3.4.2 プロダクト名による絞り込み

ヘルプ集・サポート技術情報については，図 2 に示したようにすべてのテキストに対象プロダクト名が明示されているので，これを利用してテキストの絞り込みを行う．

質問文にプロダクト名 (Windows NT, Word, Excel など) が出現する場合は，そのプロダクトを対象とするテキストを検索対象とする．質問文に複数のプロダクト名が出現する場合 (「Excel で作った表が Word で読み込めない」など) は，いずれかのプロダクトを対象とするテキストを検索対象とする．

3.5 テキストのスコア計算

転置インデックスを参照して得られ、さらに質問タイプ・プロダクト名によって絞り込まれた各テキストを対象として、ユーザ質問文との間で係り受け関係まで考慮した類似度計算を行う。ただし、絞り込まれたテキスト数が1000個を超える場合は、転置インデックスにおいて一致したキーワード・同義表現グループの数の多い順に、上位1000個までを対象とする。

3.5.1 文類似度の計算

ユーザ質問文とテキスト文の2文の類似度の計算は、3.1節で述べた文節を単位として行う。2文の互いに対応する文節と係り受け関係の割合(被覆率)をそれぞれ計算し、その積を2文の類似度とする。

まず、2文間で、以下の条件によって文節・係り受け関係に対応づける。その際、対応する文節・係り受け関係に対応度(0以上、1以下の値)を付与する。

- (1) ユーザ質問文の文節 A に含まれるキーワードと、テキスト文の文節 A' に含まれるキーワード(あるいはその上位・下位語)のいずれかが一致する場合、A と A' を対応づける。対応度は、以下のように計算する。

- (a) A, A' に共通のキーワードが含まれる場合は、以下の計算式によって対応度を計算する。

$$(\text{対応度}) = \frac{(\text{共通して含まれるキーワード数})}{(\text{A, A' のうちの多い方のキーワード数})}$$

例えば、「Windows 98 SE」(3語)と「Windows 98」(2語)については、2語「Windows」「98」が共通して含まれるので、対応度は $2/3(\approx 0.67)$ となる。ただし、多くの場合、文節は1キーワードのみを含むので、対応度は1.0となる。

- (b) A のキーワードと、A' のキーワードの上位語または下位語が一致する場合は、対応度は0.9とする。

- (c) A と A' の否定フラグが一致しない場合は、対応度は一致する場合の0.6倍とする。

- (2) ユーザ質問文内の係り受け関係 $A \rightarrow B$ とテキスト文内の係り受け関係 $A' \rightarrow B'$ について、文節 A, A' と文節 B, B' がそれぞれ対応する場合、それらに対応づける。 $A \rightarrow B$ の対応度は A, A の対応度と B, B の対応度の積とする。

- (3) ユーザ質問文から抽出された同義表現グループとテキスト文から抽出された同義表現グループが一致する場合、それらが抽出された文節・係り受け関係に対応づける(図10)。対応度は1.0とする。

以上の処理の結果、両者の文節・係り受け関係に対応度が付与される。複数の対応を持つ文節・係り受け関係については、いずれか大きな対応度をその対応度とする。

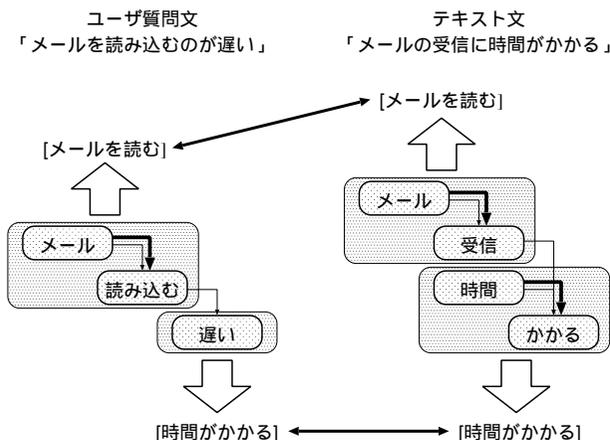


図 10 同義表現の対応づけ

ユーザ質問文, テキスト文の被覆率は, それぞれ以下の式によって計算する.

$$(\text{被覆率}) = \frac{(\text{文節の対応度の和}) + (\text{係り受け関係の対応度の和}) \times 2}{(\text{文節の総数}) + (\text{係り受け関係の総数}) \times 2}$$

ユーザ質問文, テキスト文の両者の被覆率の積を, 両者の類似度とする.

図 11においては, ユーザ質問文, テキスト文ともに 3 つの文節と 2 つの係り受け関係が対応を持っており, 対応度はすべて 1.0 である. 両者の被覆率はそれぞれ 1.0, 0.54 であるので, 類似度は 0.54 となる.

3.5.2 テキストのスコアと代表文

各テキスト中でもっとも類似度の大きな文をテキストの代表文とし, その類似度をテキストのスコアとする.

3.5.3 サポート技術情報の扱い

サポート技術情報は, 表 2に示したようにテキスト全体の複数文がマッチングの対象となるため, 特別な扱いをしている.

- テキスト文の長さが一様ではないので, テキスト文の被覆率を考慮しない. すなわち, ユーザ質問文とテキスト文の類似度は, ユーザ質問文の被覆率とする.
- 一つの事象を複数文で説明している場合が多いので, 前後の文とのマッチングを考慮する. ユーザ質問文とテキスト文 S_n の間で類似度を計算する場合は, ユーザ質問文の文節・係り受け関係と, S_n の前後の文 (S_{n-1} , S_{n+1}) の文節・係り受け関係の対応にも, 対応度 0.5 を与える.

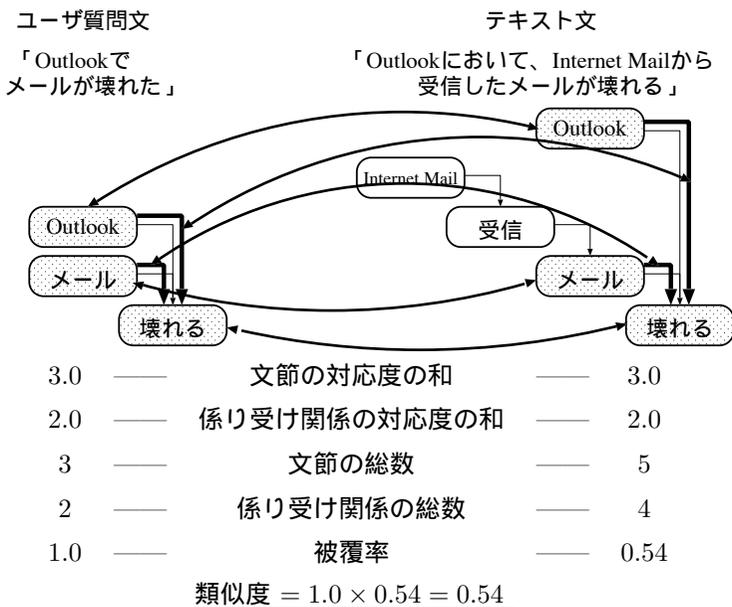


図 11 ユーザ質問文とテキスト文の対応づけと類似度の計算

- サポート技術情報のテキストには、図 2に示したように、セクションが存在する。これらのセクションのうち「タイトル」「概要」「現象」「症状」セクションには、ユーザが頻繁に質問することがら書かれていることが多い。そこで、文の存在するセクションに応じて、類似度に下記の係数を掛け合わせる。

- タイトル・概要 1.0 倍
- 現象・症状 0.8 倍
- 上記以外 0.6 倍

3.6 選択肢の絞り込み

テキスト検索モジュールは、3つのテキスト集合(用語集・ヘルプ集・サポート技術情報)ごとに、テキストのスコアに基づいてユーザに提示する選択肢を絞り込む。

テキストをスコアの大きい順に整列し、上位 n 個までをユーザに提示する選択肢とする。ただし、スコアが閾値 t を下回るものは対象外とする。また、同じスコアの複数のテキストが n 位前後で並ぶ場合は、それらをすべて含める。 n, t の値は、表 5に示すようにテキスト集合ごとに定めた。

複数のテキスト集合から選択肢が得られた場合は、用語集、ヘルプ集、サポート技術情報の

表 5 選択肢の絞り込みのパラメータ

テキスト集合	最大選択肢数 n	スコア閾値 t
用語集	2	0.8
ヘルプ集	5	0.3
サポート技術情報	10	0.1
対話カードの <UQ>(4.2節)	1	0.8

順で提示する。

4 ユーザのナビゲート

ユーザが自分の知りたいことを普通に表現しても、それで一意に適切なテキストが決まることは少ない。例えば「Windows98 で起動時にエラーが発生した」という比較的具体的な質問であっても、いくつかの原因と対策があり、それぞれにテキストが存在する。ユーザの質問がさらに曖昧であったり抽象的であったりする場合には、より多くのテキストが候補として選ばれる。いずれにせよ、ユーザが、複数のテキスト候補の中から、自分の状況に一番適切なものを選択することが必要になる。

WWW のサーチエンジンは、テキスト中から検索語を含む部分を抽出してユーザに提示することによって、ユーザのテキスト選択を補助している。本システムでは、この考え方を一歩進め、ユーザの質問（遭遇している問題）をより具体化するような説明文をテキスト中から自動的に抽出し、それらを選択肢として提示するという形でユーザへの聞き返しを行う。

しかし、ユーザの質問が非常に曖昧な場合には上記の方法はうまく機能しない。そこで、頻繁に尋ねられる曖昧な質問に対して、それをどのように対話的に具体化するかを対話カードという形式で体系化した。例えば、図 12 に示すように、ユーザが「エラーが発生した」という質問をした場合、「エラーが発生したのはいつですか」「使っている Windows のバージョンは何ですか」などの聞き返しを行って、ユーザの問題を具体化する。

4.1 状況説明文の抽出

ユーザ質問とマッチした知識ベース中の文では、その中のマッチしなかった部分に、ユーザの問題をより具体化する状況説明が与えられていると考えられる（このような部分を状況説明文とよぶ）。たとえば、ユーザが「ページ違反が発生する」と質問し、これが「IE5 を起動した際にページ違反が発生する」という文にマッチした場合、マッチしていない「IE5 を起動した際に」という部分が状況説明文となる。ユーザの質問にマッチした複数の文からそれぞれ状況説明文を抽出し、ユーザに選択肢として提示すれば、ユーザは自分の状況に適合するものを容

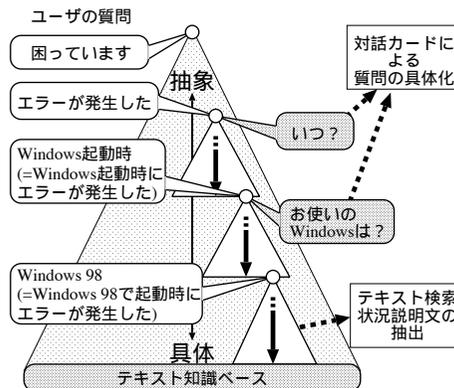


図 12 ユーザのナビゲート

易に選択することが可能となる。

状況説明文抽出のアルゴリズムを以下に示す。

- (1) 「この資料では、(～)」「以下の」「(～する) 問題について説明しています」など、頻出する冗長な表現をパターンマッチにより削除する。
- (2) 文を次の箇所 で分割する。分割された各部をセグメントと呼ぶ。
 - 連用修飾節
 - 「～とき」「～際」「～場合」「～最中」など
 - 読点を伴うデ格
- (3) セグメントのうち、すべての文節がユーザ質問文中の文節と対応するものを削除する (同義表現として対応する文節も含む)。
- (4) 末尾 (削除されたセグメントを除く) のセグメントを状況説明文の核とする。
- (5) 核のセグメントと、それに直接係るセグメントのみを、状況説明文として選択する。

アルゴリズムの適用例を図 13 に示す。まず、左の文は 2 つのセグメント A・B、右の文は 3 つのセグメント C・D・E に分割される。このうち、左の文のセグメント B と、右の文のセグメント C・E は、すべての文節がユーザ質問文と対応するため削除される。結果としてセグメント A と D が状況説明文の核となり、「IE5 を起動した際に」と「タスクスケジューラを使うと」が状況説明文として出力される。

サポート技術情報のテキストについては、各選択肢テキストの代表文から状況説明文を抽出する。用語集・ヘルプ集のテキストについては、各テキストの見出し語・タイトル自体が簡潔な説明文となっているので、この処理の対象とはしない。

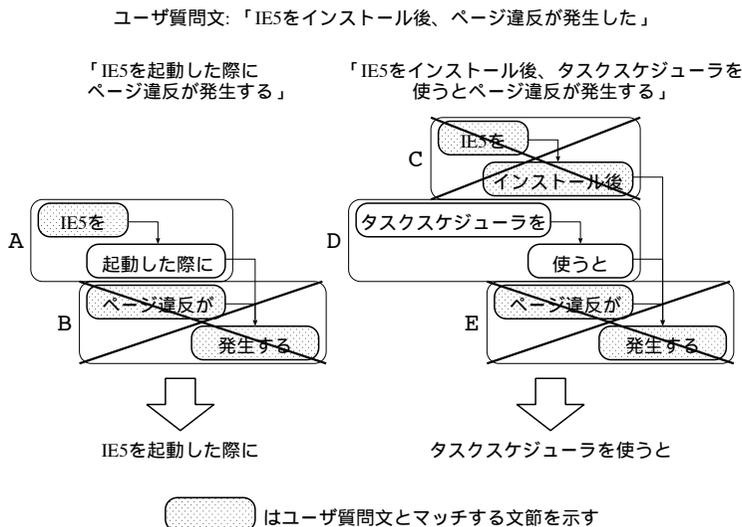


図 13 選択肢テキストからの状況説明文の抽出

4.2 対話カードを用いた聞き返し

ユーザの質問が非常に曖昧な場合には、テキスト検索の精度が低くなり、多くの不適切なテキストがマッチしてしまう。このような場合に状況説明文の抽出を行っても、誤りを含んだ多くの選択肢が得られることになり、ユーザの助けとはならない。

そこで、頻繁に尋ねられる曖昧な質問に対して、それを対話的に具体化する手順を対話カードという形式で体系化した。1枚の対話カードは、あるユーザ質問に対して、どのような聞き返しをすればよいかを記述したもので、以下の要素から構成されている(図5)。

<ID>: 対話カードのID。

<UQ>: ユーザ質問文。この部分がユーザの質問文とマッチすればこの対話カードが利用される。

<REPLY>: システムからユーザへの聞き返し発話。

<SEL action=CARD/SHOW/RET ... >: 聞き返しの際、ユーザに提示する選択肢。それぞれの選択肢にはユーザがそれを選んだ場合のシステムの動作が記述されている。action=CARDの場合には card_idで示された対話カードに移る。action=SHOWの場合には urlで示されたwebページ(マイクロソフトのサイトの種々のドキュメント)または text_idで示された知識ベースのテキストを表示する。action=RETの場合には phraseで示された質問文によって知識ベースを検索する。

対話カードの利用例を図1によって説明する。まずユーザが「エラーが発生した」という質問をすると、質問文と各対話カードの<UQ>の部分とのマッチングを3節で述べたアルゴリズムに

よって行う。この結果、図 5 上段の対話カードが選ばれる。システムはこのカードに従って、「エラーはいつ発生しますか?」という聞き返しを、選択肢を示して行う。ユーザが「Windows 起動中」を選ぶと、システムは図 5 下段の [エラー/Windows 起動中] の対話カードに移って、「あなたがお使いの Windows を選んでください」という聞き返しを行う。ここでユーザが「Windows 98」を選ぶと、「Windows 98 の起動中にエラーが発生する」を質問文として知識ベースのテキストの検索を行う。

対話カードはこのように階層的に構成されており、そのすべてのカードの <UQ> が検索対象となっている。すなわち、図 12 で示したさまざまなレベルの曖昧性・抽象度の質問を全体的にカバーするように設計されている。たとえば、ユーザが「Windows を起動中にエラーが発生する」と質問した場合には、はじめから図 5 下段のカードを用いた対話が行われることになる。

また、対話カードの枠組みは「U:こんにちは S:こんにちは」「U:このシステム使いやすいですね S:ありがとうございます」のようなドメインとは関係のない例外的な対応を行う場合にも利用している(この場合は <SEL> のないカードとなる)。このような対応ができなければ、通常の検索、すなわち知識ベースに対して「このシステム使いやすいですね」で検索を行ってしまい、「システム」や「使う」を含む知識ベースを提示するということが起こってしまう。ユーザの例外的な発話に対する不適切な動作を防ぎ、正常な対話を維持するという意味で、対話カードによる例外処理は重要である(このような例外的な対話は次節では「範囲外」の対話と扱っている)。

5 評価

ダイアログナビは、<http://www.microsoft.com/japan/navigator/> において、2002 年 4 月から公開サービスを行っている。ユーザとの質問応答のログは、対話データベースとして蓄積している。

本システムの評価としては、対話セッションの評価、ユーザとシステムのふるまいの分析、状況説明文の抽出結果が妥当かどうかの評価の 3 種類を行った。

5.1 対話セッションの評価

対話データベース中の 2002 年 8 月 1 日～31 日の質問応答ログの中から無作為に 300 回のアクセスを選択した。それらを人手によってタスク指向対話としての意味的まとまりに分割した。このまとまりを対話セッションとよぶ。結果として、378 個の対話セッションが得られた。1 対話セッション当たりの発話ペア(ユーザのキーボード入力またはマウスでの選択と、それに対するシステムの応答)の回数は、平均 1.42 回であった。

378 対話セッションそれぞれに対して、評価者 1 名が次のいずれかの評価を与えた。

成功:知識あり システムが少なくとも1つの適切なテキストを選択肢として提示した。

成功:知識なし 知識ベースに適切なテキストがなく，かつシステムが「該当する情報を見つけることができませんでした」と回答した。

失敗:知識あり 知識ベースに適切なテキストがあるにもかかわらず，システムがそれらのうちどれも選択肢として提示することができなかった。

失敗:知識なし 知識ベースに適切なテキストがないにもかかわらず，システムが不適切なテキストを選択肢として提示した。

範囲外:適切 ユーザの質問は想定ドメインの範囲外であったが，システムは適切な応答をした。

範囲外:不適切 ユーザの質問が想定ドメインの範囲外であり，システムが不適切な応答をした。

本評価は，仮想的なテストセットによる評価とは異なり，実際にサービスを行った場面でのシステムのふるまいを正確にとらえている。しかし，不特定多数のユーザの真の意図に基づいて応答の適切さを判断することはできないという問題がある。そこで，対話セッションを以下の4つのタイプに分類し，それぞれのタイプについての評価のガイドラインを以下のように定めた上で評価を行った。評価の例を表6～表9に示す。なお，表において，“U:”はユーザの発話，“S:”はシステムの発話を示す。また，“ ”は評価者が「適切なテキスト」と判断したテキストを示す。

- タイプ A: ユーザの質問が具体的で，適切なテキストの特定に必要な情報がすべて指定されている対話セッション (表6)。

この場合は，ユーザが指定した情報がすべて含まれているテキストを，適切なテキストであるとする。

システムが提示した選択肢中に適切なテキストが存在する場合 (A-1, A-2) は「成功:知識あり」とする。その他の場合は，評価者がキーワード検索システム⁵を用いて知識ベースを網羅的に検索し，適切なテキストが見つければ「失敗:知識あり」(A-4, A-5)，見つからなければ「成功:知識なし」(A-3)または「失敗:知識なし」(A-6)とする。

なお，ユーザがセッションの一部で曖昧な質問をしていても (A-2, A-4)，セッション全体として必要な情報がすべて指定されているときは，このタイプとする。

- タイプ B: ユーザの質問が曖昧で，適切なテキストの特定に必要な情報が一部欠落している対話セッション (表7)。対話カードが使用されたセッションは除く。

この場合は，ユーザの状況に完全に合致するテキストはどれかを判断することはできないので，ユーザが与えた指定したすべての情報が含まれているテキストを，適切なテキストであるとみなす。

ユーザの質問が1単語のみである場合 (B-3, B-4) は，その単語が含まれるすべてのテキストを適切なテキストであるとみなす。

⁵ キーワード入力に対して，マッチするすべてのテキストを表示する評価用システム。

表 6 対話セッション評価の例 (タイプ A)

A-1	U: Excel で行を追加したい S: 以下の選択肢から選んでください。 1. <u>Excel で空白のセル, 行, 列を挿入する</u> 2. <u>EXCEL: 行の挿入</u> 3. ...
	評価 成功:知識あり
A-2	U: 起動ディスク S: 以下の選択肢から選んでください。 1. <u>W98: PC-9800 シリーズ対応用起動ディスクの起動メニュー</u> 2. <u>Windows XP インストール時にフロッピーでコンピューターを起動するためのディスクの入手方法</u> 3. ... U: Windows 2000 での起動ディスクの作成方法は? S: 以下の選択肢から選んでください。 1. <u>Windows NT ブートディスクを作成しブート障害に備える方法</u> 2. <u>Windows 2000 Server の新規インストールを実行する方法</u> <u>CD-ROM ドライブからコンピューターを起動できない場合, Windows 2000 ブートディスクを作成します.</u> 3. ...
	評価 成功:知識あり
A-3	U: USB1.1 から usb2.0 に交換できますか? S: 入力された質問から, 該当する情報を見つけることができませんでした.
	評価 成功:知識なし
A-4	U: マクロを使いたい S: 以下の選択肢から選んでください。 1. <u>Access でマクロを使ってプロパティを設定する</u> 2. <u>Access Visual Basic でマクロを使ったプロパティの設定</u> 3. ... U: Excel です S: 以下の選択肢から選んでください。 1. <u>Excel 2000 の起動時に特定のファイルを開く方法</u> 2. <u>Excel: Web 上に Excel データを保存する方法</u> 3. ...
	適切なテキスト <u>Excel のマクロについて</u> 評価 失敗:知識あり
A-5	U: ウィンドウズ XP を削除するにはどうしますか? S: 入力された質問から, 該当する情報を見つけることができませんでした.
	適切なテキスト <u>Windows XP をアンインストールする方法</u> 評価 失敗:知識あり
A-6	U: ラリースポーツ チャレンジで正しく画面が表示されない S: 以下の選択肢から選んでください。 1. <u>W98SE:初期設定の画面で [ESC] を連打すると表示が不正</u> 2. <u>[CHT2] キャラクターが正しく表示されない場合</u> 3. ...
	評価 失敗:知識なし

- タイプ C: 対話カードが利用された対話セッション (表 8) .

この場合は, 対話カードの最も下の階層までユーザが選択肢を指定し, かつ適切なテキストまたは選択肢が提示された対話セッションを「成功: 知識あり」と判断する (適切なテキストの判断基準はタイプ A に準ずる). 対話カードの作成の際には, 各々の選択肢に

表 7 対話セッション評価の例 (タイプ B)

B-1	U: 表を書きたい S: 入力された質問から, 該当する情報を見つけることができませんでした. U: 表を作りたい S: 以下の選択肢から選んでください. 1. <u>PowerPoint で表を作成する</u> 2. <u>Word で表を作成する</u> 3. ...	評価 成功:知識あり
	B-2	U: 強制終了します S: 以下の選択肢から選んでください. 1. <u>EXCEL: 強制終了およびハングアップの際のチェック項目</u> 2. <u>EXCEL: 強制終了によって正常に保存されなかったファイルを回復する方法</u> 3. ...
B-3	U: Linux S: 以下の選択肢から選んでください. 1. <u>Linux を削除し Windows 2000 または NT をインストールする方法</u> 2. <u>Linux RPC クライアントからサービス拒否攻撃を実行できる可能性がある</u> 3. ...	評価 成功:知識あり
	B-4	U: FreeBSD S: 入力された質問から, 該当する情報を見つけることができませんでした.
B-5	U: 自動校正を解除したい S: 入力された質問から, 該当する情報を見つけることができませんでした.	適切なテキスト [WD98] 文章校正に関する Q&A
	B-6	U: ホームページの文字が大きい S: 以下の選択肢から選んでください. 1. <u>[HOW TO] 大きな文字配列を初期化する方法</u> 2. <u>[SNA] 127 よりも大きい文字が NvRunCmd によって間違っマッ</u> 3. ...
B-7	U: CADソフトが動かない S: 以下の選択肢から選んでください. 1. <u>[OFFXP]Office のセットアップ後に DVD ソフトが動作しない問題について</u>	評価 失敗:知識なし

対応する質問文 (phrase) に対して適切なテキストが提示されるかどうかをチェックしているため, 適切なテキストが提示されないことはほとんどなかった.

- タイプ D: ユーザの質問が想定ドメインの範囲外である対話セッション (表 9).
この場合は, 対話カードを利用して応答したとき (D-1) と, テキストを検索した結果として該当する情報がないと応答したとき (D-2) は「範囲外: 適切」, 検索されたテキストを提示してしまったとき (D-3) は「範囲外: 不適切」とした.

表 10 の右側 (計の欄) に対話セッション評価の結果を示す. 成功の割合は, 「範囲外」を除い

表 8 対話セッション評価の例 (タイプ C)

C-1	U: エラーが発生しました .
	S: エラーはいつ発生しますか? 1. <u>Windows 起動中</u> 2. <u>ログオン (ログイン) 時</u> 3. ...
	U: Windows 起動中
	S: あなたがお使いの Windows を選んでください . 1. <u>Windows 95</u> 2. <u>Windows 98</u> 3. ...
	U: Windows 98
	S: 以下の選択肢から選んでください . 1. <u>W98:起動時のフォントドライバが読み込めないエラー</u> <u>「JIS フォントドライバがインストールされていません」等のフォントの読み込み</u> <u>に関するエラーが発生した</u> 2. <u>W98:起動時に「<ファイル名> が不正ありません」のエラーについて</u> 3. ...
	評価 成功:知識あり
C-2	U: XBOX について教えて
	S: Microsoft が開発した次世代家庭用ゲーム機 . 詳しくは以下のリンクを参照してください . 1. <u>XBOX</u> 2. <u>発売予定のゲームタイトル</u> 3. ...
	評価 成功:知識あり

表 9 対話セッション評価の例 (タイプ D)

D-1	U: あなたは誰ですか?
	S: Dialog Navigator です .
	評価 範囲外:適切
D-2	U: 夏休みの天気を知りたい
	S: 入力された質問から、該当する情報を見つけることができませんでした .
	評価 範囲外:適切
D-3	U: お名前は?
	S: 以下の選択肢から選んでください . 1. <u>用語の説明: 名前</u> 2. <u>Access のフィールド, コントロール, およびオブジェクトの名前について</u> 3. ...
	評価 範囲外:不適切

た 230 対話セッションのうち 75%であった .

対話セッション内において対話カードによって応答が行われたかどうかと、対話セッションの評価の関係を表 10左側に示す . 現在、対話カードの枚数は 216 枚 (深さは最大で 3 階層) である . 評価対象の対話セッション中、対話カードが利用された割合は、「範囲外」を除いて 17%(= 38/(38 + 192)) であり、対話カードが利用されたセッションの大部分は「成功」であっ

表 10 対話カード利用の有無と対話セッション評価

評価		セッション内における 対話カードによる応答		計
		あり	なし	
成功	知識あり	38 (100%)	111 (58%)	149 (65% / 39%)
	知識なし	0 (0%)	25 (13%)	25 (11% / 7%)
	計	38 (100%)	136 (71%)	174 (76% / 46%)
失敗	知識あり	0 (0%)	15 (8%)	15 (7% / 4%)
	知識なし	0 (0%)	41 (21%)	41 (18% / 11%)
	計	0 (0%)	56 (29%)	56 (24% / 15%)
小計 (範囲外を除く)		38 (100%)	192 (100%)	230 (100% / 61%)
範囲外	適切	57 (—)	0 (—)	57 (— / 15%)
	不適切	3 (—)	88 (—)	91 (— / 24%)
	計	60 (—)	88 (—)	148 (— / 39%)
合計		98 (—)	280 (—)	378 (— / 100%)

(単位: 対話セッション数)

た。また、範囲外の質問に対しても対話カードでカバーされている範囲ではほぼ適切に対応できており、全体として対話カードという枠組みは有効に機能していると考えられる。

対話セッションの失敗の最も大きな原因は、知識ベース、同義表現辞書の不足である。ユーザ質問文に対して適切なテキストが存在しない場合、A-3のように適切なテキストがないことを判断するのは難しく、A-6・B-7のように誤ることが多い。かりに、表5のスコア閾値 t を大きくすればこの失敗を減らすことはできるが、その代償として適切なテキストが存在する場合の「成功:知識あり」が減って、「失敗:知識あり」が増えてしまう。A-5・B-6のような「失敗:知識あり」を減らすには、同義表現辞書をより充実させ、適切なテキストを大きなスコアでマッチさせる必要がある。

また、A-4のように、対話のコンテキストを考慮していないために失敗した対話セッションもあった。この種の失敗を減らすには、コンテキストを考慮したテキストの検索を行う必要がある。

なお、同義表現辞書と、例外処理的な対話カードについては、対話データベースで顕著なものについて随時データの修正・作成を行っている。このことによって、公開当初の成功率は60%程度であったが、徐々に改善され、現在では表10で示したとおり70%を越える成功率となってきている。

5.2 ユーザとシステムのふるまいの分析

前節で述べた378対話セッション内において、ユーザがどのような行動をしたか、システムがそれに対してどのような応答を行ったかを調べた(図14)。

ユーザの質問文の入力(518回)のうち、対話カードによって応答されたものは19%(=

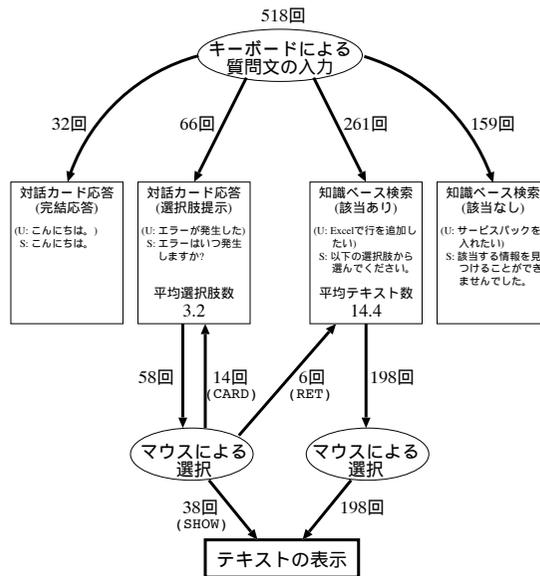


図 14 ユーザ行動とシステム応答の回数分布

表 11 ユーザ質問文の長さとはシステム応答の関係

質問文の長さ	対話カード応答		知識ベース検索		計
	完結応答	選択肢提示	該当あり	該当なし	
1 文節	29 (13%)	17 (8%)	115 (52%)	59 (27%)	220 (100%)
2 文節	3 (2%)	37 (28%)	46 (35%)	47 (35%)	133 (100%)
3 文節	()	10 (14%)	33 (45%)	30 (41%)	73 (100%)
4 文節	()	2 (6%)	22 (65%)	10 (29%)	34 (100%)
5 文節以上	()	()	45 (78%)	13 (22%)	58 (100%)
すべて	32 (6%)	66 (13%)	261 (50%)	159 (31%)	518 (100%)

(単位: 回)

(32 + 66)/518) であった。また、質問文の長さとはシステム応答の関係 (表 11) を調べたところ、対話カードは、主として短い質問文 (3 文節以下) に対応していることがわかった。一般的には、短い質問文ほど曖昧である。よって、図 12 のユーザ質問の hierarchy において、上の方の曖昧な質問文に対応するという対話カードの枠組みは、有効に機能していると考えられる。

また、ユーザ質問文の長さとは知識ベースの検索結果の関係 (表 12) も調べたところ、適切なテキストの割合は、質問文が長いほど少ないことがわかった。一般的には、長い質問文ほど専門的なものが多い。よって、知識ベースはそのような質問文を十分カバーしていないと考えられる。

一方、ユーザ質問文の長さとはテキスト数 (ユーザに提示した選択肢の数) の関係については、

表 12 ユーザ質問文の長さとは知識ベース検索結果の関係

質問文の長さ	平均テキスト数	適切なテキストの割合
1 文節 (115 回)	18.2 個	49 %
2 文節 (46 回)	9.1 個	28 %
3 文節 (33 回)	16.0 個	22 %
4 文節 (22 回)	10.5 個	10 %
5 文節以上(45 回)	10.6 個	11 %
すべて (261 回)	14.4 個	35 %

ユーザ質問文が 1 文節の場合のテキスト数が特に多かった。これは、B-3 の対話セッションのようにユーザが入力した 1 キーワードを含むテキストを多数提示してしまうことが多かったのが原因である。一方、質問文がある程度長い場合は、選択肢の絞り込みのパラメータ (表 5) によって、ユーザへの聞き返しとして適切な数に絞り込まれている。

5.3 状況説明文抽出の評価

2002 年 8 月 1 日～31 日の対話データベースから、5 つ以上の選択肢が返されたユーザ質問文をランダムに 100 個選んだ。さらに、選択肢中で上位 5 個中にランキングされているサポート技術情報の状況説明文を、評価者 1 名が「妥当」「不十分」「冗長」の 3 段階で評価した。上位 5 個中において、タイトルが代表文として選ばれている 152 個のテキストは、代表文がそのまま状況説明文となるため除外した。結果として、348(= 100 × 5 - 152) 個の状況説明文が評価の対象となった。

状況説明文の評価は、ユーザが選択肢を選ぶために必要十分な情報を、それぞれの選択肢が含んでいるかどうかという観点から行った。具体的には、まず質問文に対する選択肢 (5 個) どうしを比較し、どの情報が選択肢を選ぶ上で最も重要かを判断する (この情報を、最重要情報とよぶ)。さらに、各々の選択肢について、以下のいずれかの評価を与える。

- 妥当: 最重要情報が過不足なく含まれている。
- 不十分: 最重要情報が含まれていない。
- 冗長: 最重要情報以外の情報が著しく多く含まれている (目安としては、最重要情報以外の情報の文字数が、最重要情報の文字数の 1/2 を超えるとき)。

表 13 に状況説明文の評価結果を示す。抽出された状況説明文のうち、61 % は妥当なものであった。

また、状況説明文の平均文字数は 68.9 文字、状況説明文の抽出対象となった各テキストの代表文の平均文字数は 81.6 文字であった。したがって、提案手法による代表文の圧縮率 (= (1 - 状況説明文の平均文字数 / 代表文の平均文字数) × 100) は 15.6 % であった。

表 14 に状況説明文の評価の例を示す。この例においては、評価者は、「音が出ない具体的な

表 13 状況説明文抽出の評価結果

評価	選択肢数
妥当	213 (61%)
不十分	27 (8%)
冗長	108 (31%)
合計	348 (100%)

表 14 状況説明文抽出の評価の例

状況説明文	元の文	評価
U: 音が出ない S: 以下の選択肢から選んでください。		
1. [NT] Crystal Audio や SoundBlaster AWE32 利用時に音が出ない	(タイトル)	
2. コントロールパネルの [サウンド] から CHIMES WAV ファイルをテストした場合、ボリューム設定に関わらず	コントロール パネル の [サウンド] から CHIMES.WAV ファイルを テストした場合、ボリューム設定に関わらず、音は出ません。	妥当
3. 音楽の再生時に USB スピーカーからポップ音が出る	(タイトル)	
4. YAMAHA YSTMS55D USB スピーカセットのインストール後、スピーカのボリュームコントロールノブを使っても、非常に音が小さい、または、音が出ない	YAMAHA YSTMS55D USB スピーカセットのインストール後、スピーカのボリューム コントロール ノブを使っても、非常に音が小さい、または、音が出ないことがあります。	冗長
5. Windows サウンド (.WAV) ファイルを再生時に	Windows サウンド (.WAV) ファイルを再生時に、音が出ない。	妥当

環境 (サウンドデバイス名, アプリケーション名, ファイルの種類など) が最重要情報であると判断した。2番, 5番の選択肢は, 再生するファイルの種類を過不足なく述べているため「妥当」と判断した。一方, 4番の選択肢は, サウンドデバイス名を含んでいるものの, それ以外の発生条件や「非常に音が小さい」といった情報を余分に含んでいるため「冗長」と判断した。

提案手法による代表文の圧縮率が比較的小さかったのは「冗長」な状況説明文が多いのが大きな要因であった。具体的には, 表 14の 4番の選択肢のように, ユーザが選択肢を指定する上で重要な情報を含まないセグメントが, 削除されずに状況説明文に含まれてしまったものが多かった。より適切な選択肢を得るためには, 選択肢の代表文どうしを比較して何が最も重要な情報かを認識し, それを優先して選択肢に含める一方, それ以外の情報は除外することが必要である。また, 4番の選択肢では「非常に音が小さい」と「音が出ない」を並列節として扱うことによって, 両者がともに聞き返しにとって冗長であることを認識して削除する必要がある。

「不十分」と評価された状況説明文については, 状況説明文抽出の対象となるテキストの代表文 (3.5.2項) がテキストの内容をよく表していないものが多かった。これは, 主にユーザ質問文とテキストのマッチングに関する問題である。しかし, テキストの中には, ユーザ質問文と

マッチする 1 文だけを抽出しても, 良い代表文が得られないものもある。例えば, ユーザが遭遇する問題 (「エラーが発生する」など) と, ユーザの具体的な状況 (エラーメッセージなど) が, それぞれ別々の文に書かれている場合は, 提案手法はうまくいかない。このような場合は, 文脈解析などのより深い言語処理が必要となる。

6 おわりに

本論文では, 大規模テキスト知識ベースを利用する対話的質問応答システムを提案した。システムを実際に運用し, 得られた対話ログに基づいてシステムの評価を行い, 対話セッションの成功率 76%, 妥当な状況説明文の割合 61% という結果を得た。また, 曖昧な質問への聞き返しとして対話カードと状況説明文の抽出を組み合わせる本システムの枠組みは, 有効に機能していることを示した。

今後の課題としては, 対話カードの自動的な作成と, 対話のコンテキストの利用があげられる。対話カードの作成は, 現在はすべて人手で行っているが, 曖昧な質問を十分にカバーする対話カード集合の構築にはコストがかかるので, 自動的に作成する手法が必要である。また, 対話のコンテキストの利用については, 収集した対話ログをより詳細に分析することで, 研究を進める予定である。

参考文献

- Ferret L., Grau B., Plantet M. H., Illouz G., Jacquemin C., and Masson N. Lecuyer, P. (2001). “QALC: the Question-Answering system of LIMSI-CNRS.” In *The Ninth Text REtrieval Conference (TREC-9)*, pp. 235–244. NIST Special Publication.
- Fukumoto J., Kato T., and Masui F. (2002). “Question Answering Challenge (QAC-1) Question answering evaluation at NTCIR Workshop 3.” In *Working Notes of the Third NTCIR Workshop Meeting, Part IV: Question Answering Challenge (QAC1)*, pp. 1–10. National Institute of Informatics.
- Hammond K., Burke R., Martin C., and Lytinen S. (1995). “FAQ Finder: A Case-Based Approach to Knowledge Navigation.” In *Proceedings of the 11th Conference on Artificial Intelligence for Applications*.
- Harabagiu S., Moldovan D., Pasca M., Mihalcea R., Surdeanu M., Bunescu R., Gîrju R., Rus V., and Morărescu P. (2001). “The role of lexico-semantic feedback in open-domain textual question-answering.” In *Proceedings of the Association for Computational Linguistics*.

- Hearst M. A. and Pedersen, J. O. (1996). "Reexamining the Cluster Hypothesis: Scatter/Gather on Retrieval Results." In *Proceedings of ACM SIGIR*.
- Kawahara D., Kaji N., and Kurohashi S. (2002). "Question and Answering System based on Predicate-Argument Matching." In *Working Notes of the Third NTCIR Workshop Meeting, Part IV: Question Answering Challenge (QAC1)*, pp. 21–24. National Institute of Informatics.
- Kurohashi S. and Higasa W. (2000). "Dialogue Helpsystem based on Flexible Matching of User Query with Natural Language Knowledge Base." In *Proceedings of 1st ACL SIGdial Workshop on Discourse and Dialogue*, pp. 141–149 HongKong.
- Kurohashi S. and Nagao M. (1994). "A syntactic analysis method of long Japanese sentences based on the detection of conjunctive structures." *Computational Linguistics*, **20** (4).
- 黒橋禎夫, 長尾真 (1999). 日本語形態素解析システム JUMAN version 3.61 使用説明書. 京都大学大学院情報学研究科, <http://www-lab25.kuee.kyoto-u.ac.jp/nl-resource/juman.html>.
- Lagus K. (2000). *Text Mining with the WEBSOM*. No. 110 in Acta Polytechnica Scandinavica, Mathematics and Computing Series. D.Sc.(Tech) Thesis, Helsinki University of Technology, Espoo, Finland.
- Murata M., Utiyama M., and Isahara H. (2002). "A Question-Answering System Using Unit Estimation and Probabilistic Near-Terms IR." In *Working Notes of the Third NTCIR Workshop Meeting, Part IV: Question Answering Challenge (QAC1)*, pp. 47–54. National Institute of Informatics.
- 西村英樹, 河野浩之, 長谷川利治 (1996). "WWW データ資源検索システムの実装と評価." 電子情報通信学会技術研究報告 DE96-54, pp. 1–6.
- NIST and DARPA (2001). *The Ninth Text REtrieval Conference (TREC-9)*. NIST Special Publication.
- Oddy R. N. (1977). "Information retrieval through man-machine dialogue." *Journal of Documentation*, **33** (1), pp. 1–14.
- Rocchio J. (1971). "Relevance Feedback in Information Retrieval." In Salton, G. (Ed.), *The SMART Retrieval System: Experiments in Automatic Document Processing*, chap. 14, pp. 313–323. Prentice-Hall.
- Takano A., Niwa Y., Nishioka S., Iwayama M., Hisamitsu T., Imaichi O., and Sakurai H. (2000). "Associate Information Access Using DualNAVI." In *Proceedings of Kyoto International Conference on Digital Libraries 2000 (ICDL '00)*, pp. 285–289.
- Wilensky R., Arens Y., and Chin D. (1984). "Talking to UNIX in English: An Overview of UC." *Communications of the ACM*, **27** (6), pp. 574–593.

Zhai C., Tong X., Milic-Frayling N., and Evans D. A. (1996). "Evaluation of Syntactic Phrase Indexing — CLARIT NLP Track Report." In *The Fifth Text REtrieval Conference (TREC-5)*, pp. 347–358. NIST Special Publication.

略歴

清田 陽司: 1998年京都大学工学部電気工学第二学科卒業。2000年同大学院情報学研究科修士課程修了。2003年同大学院情報学研究科博士後期課程単位認定退学。同年、東京大学大学院情報理工学系研究科産学官連携研究員、現在に至る。質問応答システム、情報検索、自動要約の研究に従事。

黒橋 禎夫: 1989年京都大学工学部電気工学第二学科卒業。1994年同大学院博士課程修了。Pennsylvania大学客員研究員、京都大学工学部助手、京都大学大学院情報学研究科講師を経て、2001年東京大学大学院情報理工学系研究科助教授、現在に至る。自然言語処理、知識情報処理の研究に従事。

木戸 冬子: 1997年マイクロソフト株式会社入社。1998年埼玉大学大学院理工学研究科入学(在学中)。University Program担当。自然言語処理技術を用いたサポートシステムの効率化を目的としたストレリチアプロジェクトのリーダーに従事。2001年科学技術振興事業団による理科教育用のデジタルコンテンツ開発にあたっては、埼玉大学、お茶の水女子大学、東京学芸大学との共同開発プロジェクトのリーダーを担当した。現在は、University Program担当として自然言語処理を中心とした大学との共同研究を担当している。

(2002年11月22日受付)

(2003年3月8日再受付)

(2003年4月10日採録)