

知的ヘルプシステムにおける 入力文とヘルプテキストのマッチング

鈴木 達彦[†] 柿間 俊高[†] 吉良 徹[†] 荒木 亮[‡] 杉本 徹[†]

芝浦工業大学 工学部 情報工学科[†]

芝浦工業大学大学院 工学研究科 電気電子情報工学専攻[‡]

1. 背景と目的

現在、多種多様なソフトウェアが開発され、ユーザに利用されている。ソフトウェアには使い方の説明をするためにヘルプ機能が備わっているが、ユーザはヘルプ文書の検索や理解に多くの時間を費やしている。これはヘルプ文書特有の言葉遣いが原因と考えられる。そこで自然言語文の意味理解と対話要素を取り入れ、ユーザの知識に合わせて言い換えたヘルプテキストを提供するヘルプシステムが提案されている[1]。

本研究ではより容易に構築可能な辞書を用いて、同様のヘルプシステムを開発することを目標とし、入力文とヘルプテキストの解析およびそれらのマッチングの手法を提案する。

2. 知的ヘルプシステム

システムの処理の流れを図 1 に示す。

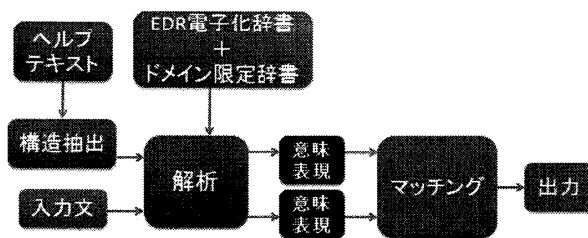


図 1 システムの処理の流れ

ユーザが入力した文とヘルプから抽出したヘルプテキストに対して自然言語解析を行い、解析結果として意味表現を生成する。意味表現同士のマッチングを行い、その結果としてヘルプのタイトル候補の一覧を出力する。

本システムの特徴は EDR 電子化辞書[2]とドメイン限定辞書[3]を使用することにより、ヘルプ文書特有の言葉の解析およびマッチングを行っている点である。

Matching of input sentences and help texts in an intelligent help system

[†]Tatsuhiko Suzuki, Toshitaka Kakima, Toru Kira, Toru Sugimoto

Department of Information Science and Engineering, Shibaura Institute of Technology

[‡]Ryo Araki

Graduate School of Engineering, Shibaura Institute of Technology

3. 研究内容

3.1 入力文の調査

理研対話コーパス[4] から、ユーザが発話した Word の操作方法に関する質問文を収集した。これらの文を知的ヘルプシステムに入力されるサンプル文と仮定してその傾向を調査した。

3.1.1 ユーザ特有の言葉遣いの調査

ユーザからの質問文とヘルプテキストでは、意味が同じでも異なる単語を用いることがある。両者を比較することで、ユーザが使用する単語とヘルプテキストで使用される単語の関係をまとめた。まとめた単語の関係は対応表にして、文の解析で利用する。

3.1.2 ユーザの希望を表す表現の調査

ユーザ入力文はヘルプテキスト文のような形式的な文とは違い、様々な文のパターンを持っている。特に、「知りたい内容を表す表現」+「希望を表す表現」という表現が多かったため、希望を表す表現のパターンを収集した。

3.2 ヘルプページの構造抽出

一般にヘルプページは 1 ページあたりのテキスト量が多く、その構造は入れ子になっている。そこで構造を考慮した解析、マッチングを行えるようにヘルプページの構造抽出を行う。

今回 Microsoft Word 2007 付属のヘルプテキストで「表」、「図やグラフを使った作業」というカテゴリーに属す 89 ページを対象とし、タイトル、概要、トピックタイトル、サブタイトル、手順部分の抽出を試みた。

その結果、タイトル、概要、トピックタイトル、サブタイトルに関して約 8 割を解析できる形で抽出することができた。しかし、手順部分に関しては、未対応のタグで囲まれている箇所、表が含まれている箇所などが抽出できず、約 6 割の抽出率にとどまった。この結果を踏まえて、本研究ではヘルプテキストのうち手順を除いた文を解析対象とする。

3.3 文の解析

本研究室で開発中の自然言語解析システム SEM を用いて入力文とヘルプテキストを解析す

る。SEM が行う解析処理は、形態素解析、係り受け解析、電子化辞書を用いた単語解析と意味解析からなる。電子化辞書には EDR 電子化辞書とドメイン限定辞書を使用し、入力文の調査で得られたユーザの言葉遣いに関する対応表を参照する。入力文を解析する際には、ユーザの希望を表す表現を事前に削っておく。意味解析では格フレーム辞書を用いたフィルタリングによって概念の候補を絞るとともに、共起辞書も参照して深層格の付与を行う。出力される解析結果には形態素情報や係り受け情報に加え、木構造で表された意味構造や概念同士の関係を表す表層格・深層格が格納される。

例として「図形をクリックして文書に挿入します。」という文の解析結果を図 2 に示す。

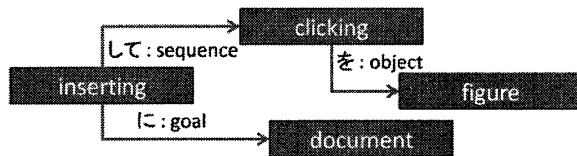


図 2 解析結果の例

3.4 マッチング

本研究では、解析結果に格納されている概念と深層格の情報を利用してマッチングを行う。この際に、対応する概念同士の類似度は、概念体系上の距離に基づいて定める。マッチングのアルゴリズムについて図 3 を例に説明する。ここでは、入力文を「罫線やセルを表に入れたいんだけど。」、ヘルプテキストを「表にセルを追加または削除する。」としている。

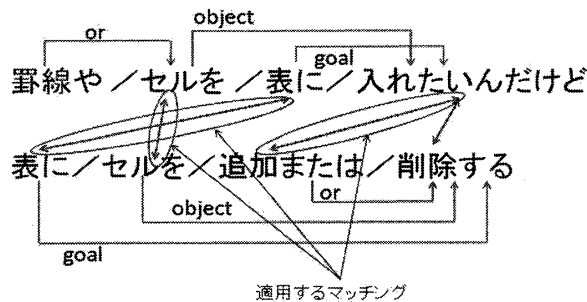


図 3 マッチングの例

1. 入力側とヘルプ側の、木構造における根となる概念を比較し類似度を計算する。根から伸びる枝の深層格が「or」、「and」の場合、それらも比較対象とし、類似度が高い組合せを根の類似度とする。図 3 の場合は、入力側「入れたい」とヘルプ側「追加する」、「削除する」の類似度を計算する。
2. 1 で対象となった概念から伸びる枝の深層格をチェックする。そして、深層格が等

しい節点の概念同士を比較し類似度を計算する。その節点から伸びる枝の深層格が「or」、「and」、「modifier」の場合、それらも比較の対象とする。図 3 では入力側の「セル」、「罫線」とヘルプ側の「セル」を比較し、入力側の「表」とヘルプ側の「表」を比較する。

3. 1, 2 で算出された類似度を足し合わせ、適用したマッチング回数で割ったものを全体の類似度とする。

4. 考察

入力文の調査で収集したサンプル文の中から、25 文を選択し、ヘルプテキストとのマッチングを行った。図 4 にサンプル文の一部を示す。

影の色変えることできるんですか
シャドウをかけたい
文字に色をつけたい

図 4 サンプル文の例

ドメイン限定辞書を解析に用いているため、ヘルプ特有の言葉を含む場合でも、正解のヘルプテキストとのマッチングを行うことができた。しかし、ドメイン限定辞書に概念が存在しない単語を含む文の場合は、正解のヘルプテキストの類似度が低くなるケースがあった。また、2 つの辞書を用いて単語の概念を付与するため、候補となる概念が多い場合、マッチングの結果が出力されるまでに 1 分以上かかるケースがあった。

また、深層格が同じ単語を比較しているのに、文の構造が異なる場合に、正解のヘルプテキストとそれ以外のヘルプテキストとの類似度が等しくなるケースがあった。例えば入力文「文字に色をつけたい」と正解のヘルプテキスト「文字の色を変更する」を比較する際に、「文字に」と「文字の」の比較が行われない。そのため、異なるヘルプテキスト「図の色を変更する」の類似度と同じ値になった。

これらの問題点を解決することが今後の課題である。

参考文献

- [1] 岩下志乃 他：マニュアルテキストを用いた個人化ヘルプシステム、第 18 回人工知能学会全国大会、2004。
- [2] 日本電子化研究所：EDR 電子化辞書第二版、2001。
- [3] 荒木亮 他：知的ヘルプシステムのためのドメイン限定辞書構築手法の提案、第 72 回情報処理学会全国大会、2010。
- [4] 理研ワープロ操作対話音声コーパス、音声資源コンソーシアム、<http://research.nii.ac.jp/src/index.html>。