

情報処理学会ソフトウェア工学研究会要求工学ワークショップ
平成27年5月21日(木)～22(金) 於 知床第一ホテル

日本語要求記述解析ツールの 開発

茨城工業高等専門学校
滝沢 陽三

問題点

- ◆ 分析・解析双方のための、単語(品詞)情報の(半)自動生成方法の必要性
- ◆ 単文化ガイドラインの不備
 - ◆ 接続詞や修飾関係以外の形式でも複文となることがある。
 - ◆ 例:「GUIはマウスおよびキーボードの入力によって機能を遂行すること。」
 - ◆ 実際の記述の分析に基づいた再定義が必要
- ◆ ツール化に向けた、複文・重文の単文化の(半)自動化方法の必要性

形態素解析の改良

- ◆ mecabの導入
 - ◆ 「分かち書き」だけでなく品詞情報も付加
 - ◆ IPA品詞体系を使用
 - ◆ lex用ファイルを自動生成
- ◆ 「WG報告付録」該当記述全てを用いてlex用ファイルを生成
 - 原文解析可能な、最も単純な文法を定義
 - 単文化ガイドラインを再定義

(参考)3文献のmecab解析結果比較

文	品詞	形態素	品詞	形態素	品詞	形態素
GUIはマウスおよびキーボードの入力によって機能を遂行すること。	名詞	GUI	助詞	は	名詞	マウス
			助詞	および		キーボード
			助詞	の	名詞	入力
			助詞	によって	動詞	機能
			助詞	を	助詞	こと
						。

単文化前記述の構文解析用文法

- ◆ mecab原文解析結果を参考に定義
 - ◆ 定義トークン: 名詞、助詞、動詞(+助動詞)、句点
 - ◆ 上記以外の単語は解析対象外
 - ◆ 結果的に、名詞句か動詞句が連続した形式
- ◆ 記述 ::= <文> 句点
- ◆ 文 ::= <名詞句> | <動詞句>
- ◆ 名詞句 ::= <名詞群> | <名詞群> <助詞群>
| <名詞群> <文> | <名詞群> <助詞群> <文>
- ◆ 動詞句 ::= <動詞> | <動詞> <文>
| <動詞> <助詞> <文>

単文化前記述の構文解析例

- ◆ 「マウスおよびキーボードの入力によって機能を遂行するGUIを提供すること。」
名詞句「マウス(および)キーボードの」
名詞句「入力によって」
名詞句「機能を」
名詞句「遂行」
動詞句「する」
名詞句「GUIを」
名詞句「提供」
動詞句「する」
名詞句「こと」

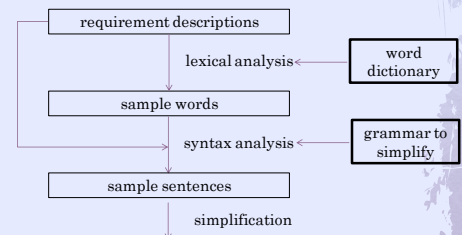
単文化前記述の構文解析例

- ◆ 「同システムを稼動させるコンピュータは2 つ以上のボタンを持つマウスとキーボードを有すること.」
 - 名詞句「(同)システムを」
 - 名詞句「稼動」
 - 動詞句「さ」
 - 動詞句「せる」
- 名詞句「コンピュータは」
- 名詞句「2」
- 動詞句「つ」(助動詞(動詞)として認識)
- 名詞句「以上の」
- 名詞句「ボタンを」
- 動詞句「持つ」
- 名詞句「マウスと」
- 名詞句「キーボードを」
- 動詞句「有する」
- 名詞句「こと」

(当面の)単文化手法の再定義

- ◆ 先の文法を原文に適用

→ 「<名詞句>*<動詞句>」を単文化ガイドラインとして参照



単文化前記述の構文解析例

- ◆ 「マウスおよびキーボードの入力によって機能を遂行するGUIを提供すること.」

「マウス(および)キーボードの入力によって機能を遂行する」

「GUIを提供する」

「こと」

単文化前記述の構文解析例

- ◆ 「同システムを稼動させるコンピュータは2 つ以上のボタンを持つマウスとキーボードを有すること.」

「同システムを稼動させる」

「コンピュータは2つ」

「以上のボタンを持つ」

「マウスとキーボードを有する」

「こと」

課題・問題点

- ◆ 不十分な単文化記述への対応
 - ◆ 品詞の細分類情報を用いた解析
 - ◆ 別の単語辞書情報を用いた解析
- ◆ yacc/lexからの脱却
 - ◆ yaccの文法認識に限界がある。
例: (なぜか) 「(<名詞句>*<動詞句>)」が認識できない。
 - ◆ lexの単語認識に限界がある。
例: 「対応付けられる」→「対応付け(動詞)」「られる(動詞)」
「対応付けの更新」→「対応(名詞)」「付け(名詞)」の「更新」
- ◆ 統合システムとしての開発・適用
 - ◆ 今回は一連のコマンドツール群として開発
 - ◆ 異なる文章・分野の要求記述の適用