

情報処理学会ソフトウェア工学研究会要求工学ワーキンググループ  
第11回東京サブワークショップ  
平成27年7月24日(金) 於 国立情報学研究所

# 自然言語を用いた 顧客による要求定義の支援手法

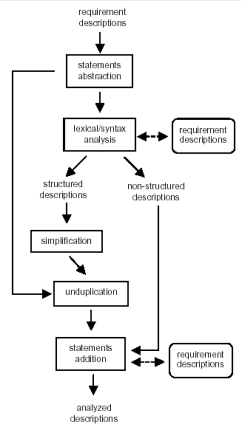
茨城工業高等専門学校  
滝沢 陽三

## テーマの概要

- ◆ 要求者 (stakeholders) 自身による要求仕様化の支援
  - ◆ 要求者の意図を開発者に伝えることが主目的  
(当初は「要求工学」として意識せず)
- ◆ 自然言語で仕様することを想定
  - ◆ 自然言語は事前学習が必要のない記法
  - ◆ 限定的な自然言語処理技術による記述の洗練
  - ◆ 構文解析処理と単語辞書検索による情報導出
- ◆ 支援システムの構成と運用
  - ◆ ドメインごとに定義された単語辞書構築
  - ◆ 開発者による支援と要求者の学習を組み入れ

## 記述の処理

- ◆ 辞書情報や顧客・開発者の判断に依存
- ◆ 開発者は解析後と辞書情報にのみ関与
- ◆ 開発者による変更は顧客の確認が必要



支援手法 ASRED の流れ  
(FIT2002)

## テーマの位置付け

- ◆ 要求獲得支援としての側面
  - ◆ 要求者自身が要求獲得・抽出に関わる(作成・品質向上させる)ことができるか?あるいは、責任があるか?
    - ◆ 契約(書)レベルの問題?
- ◆ 開発用の記法(現在のUML等)の理解を促す学習ツールの側面
  - ◆ 構文が極めて限定的
  - ◆ 当初は形式手法と関連させた研究
  - ◆ 形式化の前段階としての自然言語処理

<文> ::= <単文> 読点  
<単文> ::= <名詞 助詞 動詞句>  
| 副詞 動詞句  
| 接続詞 動詞句  
| 連体詞 名詞 助詞 動詞句  
| 形容詞 名詞 助詞 動詞句  
<動詞句> ::= 動詞 接続助詞 <単文>  
| 動詞  
| <単文>

図 5 構文解析で使った構文規則  
Fig. 5 Syntax rule for syntax analysis.

## 形式化構文の改良(拡張)

- ◆ これまでの構文はモデル文章を想定
  - ◆ 従来は単語検索による記述追加がメイン
  - ◆ 「記述例ありき」による構文の再検討の必要性
- ◆ 2007年のWG報告付録を基に改良
  - ◆ さしあたり「3.1外部インタフェース要求」を参照
  - ◆ 構文の改良を通した手順・システムの再検討
- ◆ 付録記述の修正
  - ◆ 単文化
    - ◆ 従来、複文・重文を想定していないため
    - ◆ 付録記述の多くは既に単文、または、単文しやすい形式
  - ◆ カンマの削除

## 付録A 要求仕様書の例

行頭の数字は行番号である。5章の分析は以下の3節以降を対象としている。3節において行番号の後に#がある行は行数として数えていない。よって分析対象となるのは045~110行の66文となる。

----- # 3. 要求事項について  
----- # 3.1 外部インタフェース要求  
----- # 3.1.1 ユーザーインタフェース  
[045] a. 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。  
----- # 3.1.2 ハードウェアインタフェース  
[046] a. 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、  
[047] b. キーボードを有すること。  
[048] c. 同システムを稼動させるコンピュータはMTAと通信するための装置を持つこと。  
[049] d. 二次記憶装置はRAID1に準拠し3箇所以上の冗長化が行われていること。  
[050] e. 局地災害に対処するため冗長化された記憶装置は異なる箇所に配置されていること。  
----- # 3.1.3 ソフトウェアインタフェース  
[051] a. US-ASCIIだけでなく ISO 8859-1 (Latin-1) Character Set を扱えること。  
[052] b. ISO-2022-JP 等。アジア圏の言語は扱えなくともよい。

## 修正した記述

- ◆ システムはGUIを提供すること。
- ◆ 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。
- ◆ コンピュータはシステムAを稼働させること。
- ◆ コンピュータはマウスおよびキーボードを有すること。
- ◆ マウスは2つ以上のボタンを持つこと。
- ◆ コンピュータはMTAと通信するための装置を持つこと。
- ◆ 二次記憶装置はRAID1に準拠すること。
- ◆ 二次記憶装置は3箇所以上の冗長化が行われていること。
- ◆ 局地災害に対処するため二次記憶装置は冗長化されること。
- ◆ 冗長化された二次記憶装置は異なる国に配置されていること。
- ◆ US-ASCIIだけでなくISO 8859-1 (Latin-1) Character Setを扱えること。
- ◆ ISO-2022-JP等のアジア圏の言語は扱えなくともよい。
- ◆ 日付および時刻の情報を取得できること。
- ◆ MTA側に論文投稿窓口となるメールアドレスが登録されていること。
- ◆ MTA側に査読結果受理窓口となるメールアドレスが登録されていること。
- ◆ SMTPに必要な装置やソフトウェア部品が提供されていること。

7

## 改良した構文規則

<文> ::= <単文> 句点

<単文> ::= <名詞句> <動詞句>  
| 副詞 <動詞句>

<名詞句> ::= 名詞 助詞  
| 名詞 接続詞  
| 名詞 慣用句  
| 形容詞

<動詞句> ::= 動詞 慣用句 <単文>  
| 動詞 <単文>  
| <単文>

8

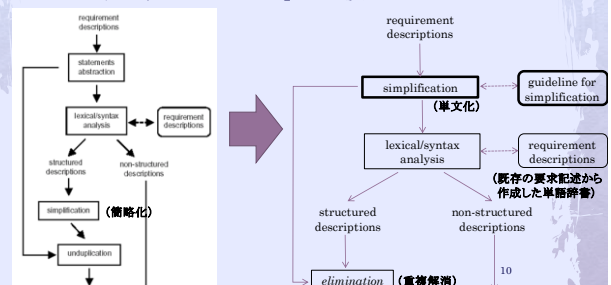
## 構文規則の改良・実現

- ◆ 「名詞句」の導入
- ◆ 「慣用句」を導入
  - ◆ 以前は削除対象、次段階で必要に応じて記述追加
- ◆ いくつかの品詞を削除・統合
  - ◆ 接続助詞と接続詞の統合、等
- ◆ 今回はyacc/lexで実現
  - ◆ 以前は日本語文字処理の都合上、Schemeインタプリタ上で実現
- ◆ 自然言語処理というよりはプログラミング言語処理に近い形で実現
  - ◆ 「単文化ガイドライン」の検討

9

## 要求記述の形式化手順の再検討

- ◆ 要求記述部分抽出と簡略化を統合(単文化)
  - ◆ 「単文化ガイドライン」や支援ツールの必要性



## 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

11

## 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

12

## (参考)3文献のmecab解析結果比較

[illegible]

13

## 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＞

14

## 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」

15

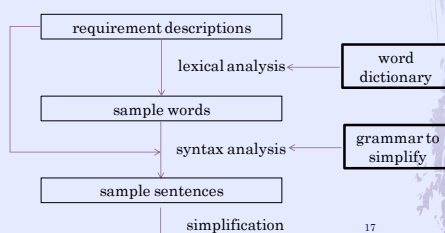
## 単文化前記述の構文解析例

- ◆ 「同システムを稼働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「稼働」  
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」

16

## (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



17

## 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」

18

## 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」

19

## 課題・問題点

- ◆ 不十分な単文化記述・認識への対応
  - ◆ 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
  - ◆ 手作業による修正
- ◆ 修飾表現の削除に伴う問題点の分析・検討
  - ◆ 形容詞や副詞は本当に削除して良いか？
  - ◆ 名詞句や動詞句による修飾表現の扱いは？
- ◆ 別言語への手法の適用可能性
  - ◆ 品詞・文法表現の違いによる手法変更の是非
  - ◆ 原文解析に基づく分析・検討

20

## 課題・問題点

- ◆ 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

21