

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26年10月9日(木)～11(金) 於 KKR熊本+ホテルグレイトフル高千穂

日本語要求記述解析のための 調査・分析

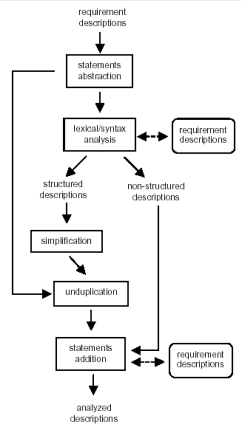
茨城工業高等専門学校
滝沢 陽三

テーマの概要

前回

- ◆ 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
(当時は「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

前回



支援手法 ASRED の流れ
(FIT2002)

記述の処理

前回

- ◆ 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

解析・修正は繰り返し行われる

(FIT2002)

前回

過去の研究発表

- ◆ 情報処理学会論文誌
 - ◆ 自然言語記述による要求仕様導出支援システムの提案(1997)
- ◆ 1997年度以降
 - ◆ 要求仕様導出支援システムにおける辞書構築手法 (SIGSE, 1998)
 - ◆ 顧客による要求分析のための手法とその支援システム (SIGSE, 2000)
 - ◆ ソフトウェア保守過程における要求記述の役割に関する考察 (学内彙報, 2002)
 - ◆ 保守過程における自然言語を用いた要求記述支援システムの検討 (SIGSE, 2002)
 - ◆ 顧客によるソフトウェア要求定義のための方法論の考察 (FIT, 2002)
 - ◆ 顧客によるソフトウェア要求定義の手順に関する考察 (学内彙報, 2003)

前回

その他の研究

- ◆ 安価なPCを用いたアプライアンス指向サーバの構築と運用 (学内彙報, 2006)
- ◆ 地域組織に対するプロジェクト支援に関する考察 (学内彙報, 2006)
- ◆ 学生教育と連動させた地域情報化支援 (工学教育, 2009)
- ◆ ディスクレスによるプログラミング演習室システムの開発 (学内彙報, 2010)

前回

その他の研究

- ◆ 複数の分野の混成グループによるソフトウェア開発の方法論に関する検討 (学内彙報、2011)
- ◆ 複数の分野の混成グループによるソフトウェア設計における開発方法論の導入 (学内彙報、2012)
- ◆ 機能限定ソフトウェアシステムのフレームワーク策定および開発環境に関する考察 (学内彙報、2012)
- ◆ フレームワークに基づいた機能限定ソフトウェアシステム設計ツールの開発 (学内彙報、2013)

前回

既存ソフトウェアに対する要求

- ◆ 設定値の変更で対応できるもの
- ◆ データベースの変更で対応できるもの
- ◆ モジュールの交換で対応できるもの
- ◆ システムの再設計が必要なもの
- ...

(FIT2002)

前回

テーマの位置付け

- ◆ 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
- ◆ 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?

<文> ::= <単文> 説点
<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>
| 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.

前回

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、
[047] b. キーボードを有すること。
[048] c. 同システムを稼働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国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。
```

前回

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。

```
[001] 1. はじめに
[002] 1.1 目的
[003] 国際会議のプログラム委員長 (PC) の業務を支援する情報システムの構築のためにこの文章は書かれた。
[004] よって、システムの設計者を含む開発グループおよびプログラム委員長自身が本文書の主たる読者となる。
[005] 1.2 スコープ
[006] 本システムは「PC 支援システム」と命名されている。
[007] その名のとおり多忙なプログラム委員長が行う業務のうち、自動化可能なものは自動化し、委員長の意思決定の助けとなる情報を自動作成することを目的とする。
[008] これによって、PC 自身の仕事量の削減、および仕事の遂行時間の短縮を可能とする。
```

前回

今後の予定

- ◆ 要求者自身に関わることは想定しない?
 - ◆ 開発者側からのアプローチに徹する可能性
- ◆ 要求者と対象ドメインの分類・詳細化?
 - ◆ 他の手法との関連

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

元の記述

- ◆ 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。
- ◆ 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。
- ◆ 同システムを稼働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。
- ◆ 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。
- ◆ 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、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。

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。

以前の構文規則

- <文> ::= <単文> 句点
- <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>
| 形容動詞 名詞 助詞 <動詞句>
- <動詞句> ::= 動詞 接続助詞 <単文>
| 動詞
| <単文>

改良した構文規則

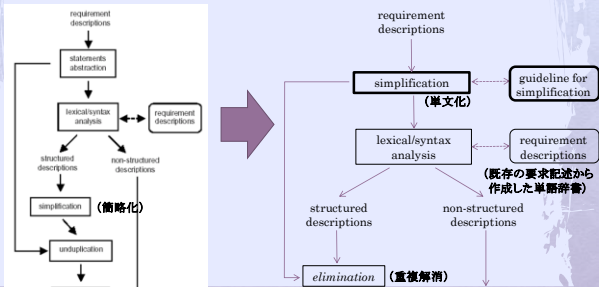
- <文> ::= <単文> 句点
- <単文> ::= <名詞句> <動詞句>
| 副詞 <動詞句>
- <名詞句> ::= 名詞 助詞
| 名詞 接続詞
| 名詞 慣用句
| 形容詞
- <動詞句> ::= 動詞 慣用句 <単文>
| 動詞 <単文>
| <単文>

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

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



今後の予定

- ◆ 関連研究との比較・検討
- ◆ 他の部分・分野の記述に対する調査・分析
- ◆ 他の言語の記述に対する調査・分析
- ◆ 単語(品詞)情報の(半)自動生成
- ◆ 複文・重文の単文化の(半)自動化の検討

```
$ cat reqsample-3.1.txt
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を移動させること。
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RAID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間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等のアジア語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。

$ ./req < reqsample-3.1.txt
NOUN POSTP NOUN POSTP VERB IDIOM OK
NOUN POSTP NOUN CONJ NOUN POSTP NOUN IDIOM NOUN POSTP VERB IDIOM OK
NOUN POSTP NOUN POSTP VERB IDIOM OK
NOUN POSTP NOUN CONJ NOUN POSTP VERB IDIOM OK
NOUN POSTP NOUN POSTP NOUN POSTP VERB IDIOM OK
NOUN POSTP NOUN POSTP NOUN POSTP VERB IDIOM VERB IDIOM OK
NOUN POSTP NOUN POSTP VERB IDIOM OK
NOUN POSTP NOUN POSTP NOUN POSTP VERB IDIOM OK
NOUN POSTP NOUN POSTP NOUN IDIOM VERB IDIOM OK
ADJ NOUN POSTP ADJ NOUN POSTP VERB IDIOM OK
NOUN IDIOM NOUN POSTP VERB IDIOM OK
NOUN POSTP NOUN POSTP NOUN POSTP VERB OK
NOUN CONJ NOUN POSTP NOUN POSTP VERB IDIOM OK
NOUN POSTP NOUN IDIOM NOUN POSTP VERB IDIOM OK
NOUN POSTP NOUN IDIOM NOUN POSTP VERB IDIOM OK
NOUN POSTP ADJ NOUN CONJ NOUN POSTP VERB IDIOM OK
```

"マウス" return(NOUN); "キーボード" return(NOUN); "入力" return(NOUN); "機能" return(NOUN); "GUI" return(NOUN); "システム" return(NOUN); "コンピュータ" return(NOUN); "2つ以上" return(NOUN); "ボタン" return(NOUN); "装置" return(NOUN); "二次記憶装置" return(NOUN); "RAID" return(NOUN); "3箇所以上" return(NOUN); "冗長化" return(NOUN); "局地災害" return(NOUN); "日" return(NOUN); "US-ASCII" return(NOUN); "ISO 8859-1 (Latin-1) Character Set" return(NOUN); "ISO-2022-JP等" return(NOUN); "アジア語" return(NOUN); "言語" return(NOUN); "MTA" return(NOUN); "MTA側" return(NOUN); "論文投稿窓口" return(NOUN); "査読結果受理窓口" return(NOUN); "メールアドレス" return(NOUN); "SMTP" return(NOUN); "ソフトウェア部品" return(NOUN);	"遂行する" return(VERB); "提供する" return(VERB); "移動させる" return(VERB); "有" return(VERB); "有する" return(VERB); "通信する" return(VERB); "準拠する" return(VERB); "行われている" return(VERB); "対応する" return(VERB); "配置されている" return(VERB); "準拠する" return(VERB); "扱えなくともよい" return(VERB); "取得できる" return(VERB); "登録されている" return(VERB); "提供されている" return(VERB); "冗長化された" return(ADJ); "異なる" return(ADJ); "必要な" return(ADJ); "および" return(CONJ); "と" return(CONJ); "の" return(POSTP); "に" return(POSTP); "を" return(POSTP); "は" return(POSTP); "が" return(POSTP); "の" return(POSTP);	"によって" return(IDIOM); "だけでなく" return(IDIOM); "なる" return(IDIOM); "こと" return(IDIOM); "ための" return(IDIOM); "ため" return(IDIOM); "." return(PERIOD);
---	--	---