

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28年5月19日(木)～21日(土) 於 高崎

要求記述支援のための 事例分析

茨城工業高等専門学校
滝沢 陽三

1

背景・経緯

- 当初の研究目的・内容
(1990年代後半～2000年代前半)
 - 要求者自身による要求仕様化の支援手法開発
 - 要求者の意図を開発者に伝えることが主目的
 - 自然言語で仕様化することを想定
 - 自然言語は事前学習が必要ない記法
 - 限定的な自然言語処理技術による記述の洗練
 - 記述文章の単文化処理に基づく洗練が中心
 - 構文解析処理と単語辞書検索による情報導出
 - 用語出現に基づくドメイン固有の記述の追加支援
 - 支援システムの構成と運用
 - ドメイン（要求分野）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

2

背景・経緯

- その後の経緯
(2010年代前半)
 - 要求工学の観点でまとめ直し
 - 要求仕様化ではなく（顧客自身による）要求定義
 - 事例に基づく文法定義による記述の形式化
 - 単文化処理の改良
 - 形態素解析システム等による事例解析
 - 品詞情報を中心とした既存記述の分析

3

前回までの課題

- 事例記述の収集・分析
 - 「不適切な記述→失敗した開発」以外
 - 不適切な記述→成功した開発
 - 適切な記述→失敗した開発
 - 当面は「適切な記述」と判断された既存文章を中心に事例解析
- 「単文化」の有用性検討
 - 単文化することの意義
 - 元は「制限言語」による要求の仕様化を想定
 - 顧客による要求記述を想定し、別の観点での記述洗練を想定

4

研究目的の再定義

- 顧客自身が、提案依頼書に相当する記述から、機能要求を自然言語で定義する、手法および支援システムの開発
 - 支援システムの対象
 - RFPの洗練・詳細化、機能要求の導出
- 開発者のドメイン経験に依存していた、開発者による支援の部分をシステム化
 - ドメイン内部の知識・表現体系が必要
(これまでは単文化と記述追加支援を想定)
- あくまで顧客（要求者）視点
 - 開発者視点は設計・実装の詳細化となりがち

5

用語の確認（定義）

- 要求者≡顧客
 - 利害関係者（ステークホルダ）の一部
- 要求者は提案依頼書（RFP）までは支援なしに作成するものとする。
 - 必要な項目等の枠組みは存在
- 要求記述≡機能要求仕様書
 - RFPは機能要求の提案依頼

6

研究目的の再定義

- 当面の支援内容は「あいまいさ」の解消
 - 文章表現に対する修正支援として
 - 「過不足」は当研究の従来の記述追加支援でカバー
 - 「矛盾」は形式手法等の従来の関連研究でカバー
 - 状況によっては「ノイズ」の除去を加味
- 「あいまいさ」の基準の定義
 - 明確になりすぎると「設計」となってしまう。
 - 特定分野の十分に多い既存の要求定義記述より導出する仕組みを想定（そのための事例分析）
 - 十分には多くない分野の場合、要求定義促進も兼ねた手法・システムを想定しなければならない可能性あり

7

事例分析（1）

- 記述文書中の品詞分布の傾向調査
 - 元は単文化のための文法定義改良に必要なものとして実施
 - 今回は単文化以外の記述洗練の手がかりを得るための分析として実施
- MeCabによるSubtypeまでの品詞情報42種類を用いて記述内の割合を調査
 - 概ね「一割以上=多い」とみなす。
- 「要求」記述にこだわらず、様々な分野の記述を解析
 - システム仕様書、提案依頼書（RFP）、携帯アプリレビュー、文学作品（青空文庫）、等

8

事例分析（1）

- 仕様書等ほど「名詞＋する／できる」「名詞＋を行う」を使用。
 - サ変接続名詞数／（自立動詞＋非自立動詞）
 - 仕様書等：1.0～3.0
 - レビュー文章：0.5前後
 - 文学作品（青空文庫）：0.1～0.2
 - あくまで傾向であり、数値自体は特に意味はなし
 - MeCabは「名詞＋する」を「サ変接続名詞＋自立動詞」とする。
 - 日本語（文化）固有の意味限定用法？
 - 裏付を調査中（サ変接続となり得ない名詞との関係等）。
 - 英語では「名詞は全て動詞となり得る」。
 - 要求記述における多様な「モード」の存在の可能性
 - 「ドメイン」だけでなく「モード」の観点でも記述チェックすることで、あいまいさ解消（やノイズ除去）の仕組みを構築可能？

9

用語の確認（定義）

- モード
 - フランス語のmode、英語のmood
 - 本来の意味は「流行」に近い。
 - 言語学においては（今回の場合は）記述者の意図を表現するための記述方法を指す。
 - 客観的な意味内容は変わらないが、記述者の意図によって追加される動詞・述語表現。
 - 【例】客観的な意味内容「書く」
 - モードによる追加：「書け」「書きます」「書こう」
 - （敬称・丁寧等の表現に限らないことに注意）
- ここでは、個々の要求定義活動、または、ドメインごとに存在する表現体系と定義
 - 前者については、個々の要求定義文書ごとに体系が大きくことなる可能性あり

10

事例分析（2）

- 新聞記事の品詞分布の傾向
 - 専門用語が多い（相対的に造語的なサ変接続名詞が増える）記事ほど仕様書等傾向に近い可能性
 - 仕様書等に近い
 - 「プログラミング 必修化を創造力育てる一助に」
 - 「肉フェスで49人が食中毒症状 東京・お台場、3人入院」
 - 「外添都知事に政治資金規正法違反の重大疑惑！」（要約）
 - 仕様書等とレビュー文章の中間
 - 「ふるさと納税 富裕層の節税策なのか」
 - 「男児の遺体見つからず2人を処分保留 新潟地検」
 - レビュー文章に近い
 - 「首から血、高齢男性が室内で死亡…札幌の民家」
 - 「フィリピン保険金殺人の犯人が逮捕直前に全てを告白！」（要約）
 - 文学作品（青空文庫）に近いものは見当たらず
 - 「名詞＋する」等が多いのは現在の日本語文に共通の特徴？

11

事例分析（3）

- 品詞割合の平均情報量の比較
 - 品詞（Subtype）42種類を対象
 - 最大エントロピー5.39
 - 結果：仕様書等の記述ほど平均情報量が低い。
 - 仕様書等（・新聞記事）4.0未満
 - レビュー文章・文学作品（青空文庫）4.0以上
 - 考察
 - 一般動詞からサ変接続名詞への使用の偏りが原因と解釈すれば、先の分析の解釈と矛盾しない。
 - 品詞の分類方法に依存するため、多少の値の増減では「仕様書度合い」は判断できない。
 - 特に単語数が少ない記述の性質を大まかに知る手段として有効である可能性あり
 - レビュー文章と文学作品（青空文庫）で差が全くない理由は不明

12

用語の確認

- 平均情報量（エントロピー）
 - ある事象系の情報の無秩序さを示す尺度（要求のあいまいさとはもちろん関係がない）
- 記述変化に伴う平均情報量の変化の例
 - 上がる（最大エントロピーに向かう）場合
 - より強い暗号化が行われている。
 - 下がる（冗長度が低下する）場合
 - より強い圧縮が行われている。

13

事例分析（4）

- 形式的な記述の「改善」前後の品詞分布や平均情報量の変化の調査
 - RFPサンプル2記述で比較
 - 結果（改善前→改善後）
 - いずれも「サ変接続名詞数／（自立動詞＋非自立動詞）」が減少（名詞約3%減・動詞1～3%増）
 - 平均情報量も増加
 - 考察
 - 改善⇔形式化と解釈すると先の分析・解釈と矛盾
 - 「非形式的な記述の追加・残存」に伴う変化と解釈？（実際、分析・設計で要求定義に立ち戻ることが多かった）
 - 「不適切な記述→成功した開発」の事例に相当する可能性

14

事例分析（5）

- 「あいまいさをなくす」前後の品詞分布や平均情報量の変化の調査
 - 「話題沸騰ボット要求仕様書」6版と7版で比較
 - 結果（6版→7版）
 - 名詞・動詞の割合はあまり変化せず（むしろ微増）
 - 助動詞2.5%減少、記号3%増加
 - 平均情報量は減少（Subtype、品詞のいずれでも）
 - 考察
 - あいまいさをなくす⇔形式化と解釈すると先と矛盾
 - 6版と7版で体裁が異なる
 - 体裁や用語の変更で「あいまいさを減少」していると推測
 - 助動詞減少は「ですます調→だ・である調」の影響大
 - 「叙述形式→表形式」に伴い、括弧等の記号が大幅増加
 - 文章自体も大幅追加・削除

15

事例分析（6）

- 一文あたりの助詞数の傾向
 - 文書の種類に関わらずまちまち
 - （助詞数 | 格助詞数 | 係助詞数）／（1読点 | 1句読点）で比較
- その他の助詞分布に関する傾向
 - ほとんどの文書で格助詞10%前後
 - 非形式的な文書で係助詞が多め（未確認）
 - 句読点ごと（擬似的に1文）の助詞数/格助詞数/係助詞数等は文書ごとにまちまち

16

事例分析：余談

- 要求記述に関する事例の収集は困難
 - 内容の機密性よりもむしろ「要求が明確になっていなかった」（⇔記述自体が存在しない）こと自体の機密性が大きい可能性
 - 既存の汎用ソフトウェア/システムパッケージの導入のみが「要求」として想定される（結果、導入のみで稼働できない）ケースも。
 - 開発者が要求定義に関わっている場合、設計・実装の段階で多くの要求を定義し直すことが多い。
 - 要求内容のレビューも少なく、要求の過不足分が残ったまま要求定義が終了
- 比較対象としての、要求記述や仕様書以外の文書収集も困難
 - 電子媒体も複写不可であることが多く、サンプル的な手入力では限界

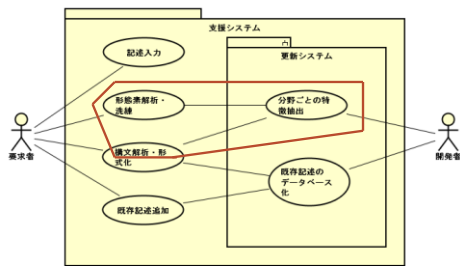
17

今後の予定

- 更なる事例分析・考察
 - 別の観点からの分析と考察
 - 文型の傾向調査等（単文・複文・重文等）
 - 要求分析・設計への発展の追跡（仮）
 - 他の自然言語（英語等）の各種文書の分析
 - 分析ツールの整理・発展
 - 構文木生成・分析ツールを含む。
- 手法の整理と支援ツールの開発
 - 「モード」情報等に基づく記述の洗練支援の検討
 - 「ドメイン情報に基づく記述の追加支援」と統合？
 - 事例分析（ツール）に基づく結果情報のデータベース化手法の検討
 - 事例分析ツールからの発展の可能性

18

今後の予定（手法の整理）



19

今後の予定（手法の整理）

□ 支援ツール実装の可能性

- 選択候補は「類語辞典」情報から生成可能？

- 用語の統一の観点も含む（「先生／教員」等）

□ 支援の例

- 「トークが過去ログから表示される現象がまだ直ってない」→修正／改善／回復
 - 以前の状態を考慮して選択
- 「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」→含有／保有／携行／有（する）／保持／維持／残存／携帯／所持／所蔵／所有
 - より限定された意味による他、主語との共起率を考慮して選択（変更不要の判定を含む）

□ 更新ツール実装の可能性

- MeCabで用いる辞書情報の半自動更新を考慮

20