

# Kuromoji と構文解析による 要求仕様書から状態遷移系への自動変換の試み

田幸 玄陽<sup>1</sup> 小形 真平<sup>1</sup> 岡野 浩三<sup>1</sup> 関澤 俊弦<sup>2</sup>

概要：本稿では、日本語で記述された要求仕様書を既存の形態素解析ツールを用いて構文解析を行い、その結果から自動で状態遷移系に変換する手法を提案する。提案する手法は構文解析の結果を XML 出力にし、モデル検査に必要な情報を XML パーサにより抽出することで、状態遷移系への自動変換の処理を行う。

## 1. はじめに

システム開発では、初期の段階で分析の対象となった要求が真のユーザ要求であることを確認しなければならない [1]。上流工程である要求仕様のほとんどは自然言語で記述されるため、曖昧な部分や矛盾を含んでいる可能性がある。このような観点から、できるだけ早期に要求仕様のあいまいな部分を検出、除去し、要求仕様書の品質向上の支援を行う必要がある。そこで本稿では、次の方法を提案する。日本語で記述された要求仕様書に対して形態素解析ツールである kuromoji を用いて形態素解析を行い、その出力結果から構文解析を行う。その構文解析の結果を XML 出力にし、XML パーサにより抽出することで、状態遷移系への自動変換の処理を行う。

### 1.1 話題沸騰ポット

本稿では、SESSAME が提供するソフトウェア技術者の教育を目的とした要求仕様書「話題沸騰ポット」[2] を対象に変換方法を文献 [3] に引き続き考察する。「話題沸騰ポット」には曖昧さを含む第 6 版と曖昧さを排除した第 7 版があり、今回は第 6 版の要求仕様書を対象に構文解析を行い、その結果をモデル検査することで曖昧検出を行うことを最終ゴールとした。また、曖昧性を除去した第 7 版との比較を行うことで変換の精度の向上を目指す。

## 2. 構文解析

### 2.1 Kuromoji

Kuromoji[4] は Java で書かれているオープンソースの日

本語形態素解析ツールである。主な機能を次に挙げる。

- 複合語の分割 複合語を複数の名詞群または動詞群に分割
- 品詞のタグ付け 単語分類の割り当て (名詞, 動詞, 助詞, 形容詞など)
- 見出し化 動詞の活用や形容詞に辞書の見出しを表示
- 読み方 漢字の読み方を抽出

### 2.2 CYK アルゴリズムによる構文解析

要求仕様書の曖昧性を検出するためにモデル検査を行う必要があるが、モデル検査の自動化をするためには要求仕様書を状態遷移系へと変換する必要がある。そのための情報を得るために構文解析を行い、条件やアクション、状態変数の取得をするための情報とする。構文解析は「複合語の分割」と「品詞のタグ付け」の機能を用いて行う。解析にはボトムアップ系の構文解析の最も基本的な方法である CYK アルゴリズム [5] を用いる。CYK アルゴリズムは容易でかつ、文の単語の並びに対し、あらかじめ用意しておいた句構造規則と一致するかどうかを動的計画法により調べる方法であるため、kuromoji による形態素解析の結果を使用できる。構文解析に用いた句構造規則の一例を示す。

- 名詞 + 助詞 → 名詞句
- 名詞 + 名詞句 → 名詞
- 名詞 + 動詞句 → 文

例えば、「給湯口」という単語が「名詞」で、「は」という単語の「助詞」が続く場合、「名詞句」という新たな品詞とみなす。最終的に入力全体を被覆する「文」が得られれば構文解析成功とする。

### 2.3 XML の出力

構文解析の結果を XML として出力する。形態素解析の

<sup>1</sup> 信州大学工学部  
Faculty of Engineering, Shinshu University

<sup>2</sup> 日本大学工学部  
College of Engineering, Nihon University

結果である単語と品詞の情報、構文解析結果の2つの情報からなる。以下はその例である。

```
<token id="1">
  <word>給湯</word>
  <part_of_speech>名詞</part_of_speech>
  <category>一般</category>
  <base_form>給湯</base_form>
</token>
<token id="2">
  <word>口</word>
  <part_of_speech>名詞</part_of_speech>
  <category>一般</category>
  <base_form>口</base_form>
</token>
<token id="3">
  <word>は</word>
  <part_of_speech>助詞</part_of_speech>
  <category>係助詞</category>
  <base_form>は</base_form>
</token>
<token id="4">
  <word>正面</word>
  <part_of_speech>名詞</part_of_speech>
  <category>一般</category>
  <base_form>正面</base_form>
</token>
<!-- 途中省略 -->
<token id="9">
  <word>ある</word>
  <part_of_speech>助動詞</part_of_speech>
  <base_form>ある</base_form>
</token>
<event>
  <noun_p>
    <noun>
      <word ref_id="1">給湯</word>
      <word ref_id="2">口</word>
    </noun>
    <word ref_id="3">は</word>
  </noun_p>
</event>
<action>
  <verb_p>
    <noun_p>
      <noun>
        <word ref_id="4">正面</word>
        <word ref_id="5">側</word>
      </noun>
      <word ref_id="6">に</word>
    </noun_p>
    <verb_p>
      <word ref_id="7">下向き</word>
    </verb_p>
    <verb>
      <word ref_id="8">で</word>
      <word ref_id="9">ある</word>
    </verb>
  </verb_p>
</action>
```

解析する話題沸騰ポットの要求文には、大きく分けて2種類存在している。「蓋はポット上部にあり、片手で操作できる。」のようなある特定の名詞を示し、その名詞の定義をする文と、「タイマボタンを押すと、タイマが起動状態となる。」のようなある特定のイベントを示し、そのイベントに対する処理を記述する文の2種類である。構文解析結果から要求文を < event > (条件節) と < action > (動作節) に分け、その二つのつながりから「定義」と「処理」に分類する。ここでは「定義」に分類された文はある状態変数の初期値や部品の役割を説明しており、「処理」に分類された文は状態変数の値が変わったときの処理の説明をするものである。表1に分類と例を示す。

表 1 要求文の分類

原文	条件節	動作節	分類
蓋はポット上部にあり、片手で操作できる。	蓋は	ポット上部にあり、片手で操作できる。	定義
タイマボタンを押すと、タイマが起動状態となる。	タイマボタンを押すと、	タイマが起動状態となる。	処理

なお「... できる状態/出来ない状態にする」という文は

2つの文にあらかじめ分解しておいた。

### 3. 状態遷移系への変換

NuSMV は、記号モデル検査アルゴリズムに基づいた高速な検証ができるモデル検査ツールである [6]。NuSMV の状態遷移系を表現する状態変数名および状態変数が持つ値の名前 (蓋センサー: on, off など) はユーザーが XML からの抽出結果を参考に手動で与える。

### 4. 成果と今後の展望

「話題沸騰ポット」の要求仕様書に対し構文解析を行い、その結果を xml に出力し、その要求が条件節か動作節かの自動判別を行った。また、要求仕様書に出てくる状態変数となり得そうな単語を手動で判別、整理しそれらを構文解析の結果から xml パーサを用いて自動抽出を行った。

今後は、状態変数になりうる単語の選別は手動で行っているため、それを自動で抽出できることを目指す。また、XML パーサで抽出したデータから NuSMV のモデル検査のコードを自動生成できるようにすることで、ソフトウェア開発の支援に結びつくように改良をする。

### 5. まとめ

ソフトウェア開発の支援を目的に、自然言語で記述された要求仕様書に構文解析を行い、その曖昧性を抽出する手法についての提案と展望を述べた。

### 謝辞

本研究は科学研究費補助金 16K00094 の支援による。

### 参考文献

- [1] 小泉寿男, 辻秀一, 吉田幸二, 中島毅: “ソフトウェア開発”, 株式会社オーム社 (2003).
- [2] 組込みシステム教育教材 話題沸騰ポット GOMA-1015 型 要求仕様書, [http://www.sesame.jp/workinggroup/WorkingGroup2/POT\\_Specification.htm](http://www.sesame.jp/workinggroup/WorkingGroup2/POT_Specification.htm) (2016/12/3 アクセス).
- [3] 遠藤健, 小形真平, 岡野浩三, 関澤俊弦: “自然語要求仕様記述の形式検証に向けて—話題沸騰ポットのモデル検査—”, ウィンターワークショップ 2015・イン・逗子論文集 pp.3-4 (2016).
- [4] Kuromoji-Atilika-Applied Search Innovation, <https://www.atilika.com/ja/products/kuromoji.html> (2016/12/3 アクセス).
- [5] T. Kasami: “An efficient recognition and syntax-analysis algorithm for context-free languages,” Scientific report AFCRL-65-758, Air Force Cambridge Research Lab, Bedford, MA (1965).
- [6] A. Cimatti, E. M. Clarke, E. Giunchiglia, F. Giunchiglia, M. Pistore, M. Roveri, R. Sebastiani, and A. Tacchella: “NuSMV 2: An OpenSource Tool for Symbolic Model Checking,” In Proceeding of International Conference on Computer-Aided Verification, pp.359-364 (2002).