

XML データベースにおけるタグ抽出機構の研究

在宅介護分野への応用

北郷 航[†] 末田直道[‡] 新田恒美^{††}

[†]大分大学大学院工学研究科知能情報システム工学専攻

[‡]大分大学工学部知能情報システム工学科

^{††}デジタルアンネット株式会社

1. はじめに

現在、介護分野では在宅介護サービスが注目されている。しかし、在宅介護サービスを行う際、事前に作成するサービス計画書(図1)[1]を利用者のニーズにあわせて作成できる経験豊富な人材が不足している。そこで、過去の事例をもとにして経験の有無に関わらずにサービス計画書の作成を支援するシステム(居宅介護サービス計画支援システム)の構築が求められる。

本研究では、XML データとして格納されている過去の事例を参照する際、詳細な類似検索を可能にするために、過去の事例のドキュメントに対し形態素解析と構文解析を用い、意味情報を持つ要素を付加する箇所を抽出する機構(タグ抽出機構)の構築に取り組んだ。

本稿では、居宅介護サービス支援システムの概要とタグ抽出機構の概要を述べる。また実験では、構文解析と抽出ルールを用いてドキュメントに要素を追加し、抽出の精度を示し、構築した抽出ルールの検証を行う。

生活全般の解決すべき課題(ニーズ)	援助目標	時期	対応目標	概要
!! 室内における歩行補助を要する 併せて、外出時の自立を促したい	機能回復訓練により、 自立歩行の達成を図る。		歩行補助具を用い、歩行練習を行う。	2週間以内

図1 居宅介護サービス計画

2. 居宅介護サービス計画支援システム

本節では、居宅介護サービス計画支援システム(図2)の構成要素を説明する。

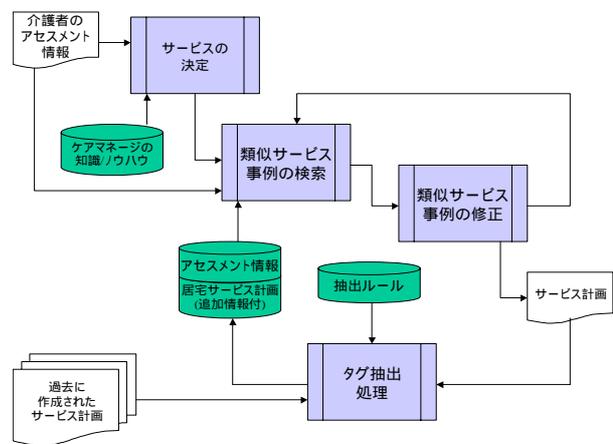


図2 居宅介護サービス計画支援システム

(1)サービスの決定

居宅サービス計画は、個々の利用者の特性とニーズに応じて作成されることが必要であるため、居宅サービス計画の作成に先立って、利用者のアセスメントを行う。このアセスメントと過去のアセスメントから抽出された知識・ノウハウの決定木からサービス項目の決定を行う。

(2)類似サービス事例の検索

過去の事例類似事例を検索することにより決定したサービス項目の内容に類似したサービスの内容や注意点、留意する点などの情報を取り出す。

(3)類似サービス事例の修正

検索し取り出した情報をもとにサービス内容を今回作成している被介護者に適した形に修正を行う。

(4)タグ抽出処理

過去に作成された大量の居宅介護サービス計画(テキストデータ)に対して形態素解析・構文解析を行い新しい要素の内容を取り出す。また、作成した居宅サービス計画にも同様の処理を行う。要素の内容を取り出し、タグを追加した後は、XML データベースに格納(追加)する。

3. タグ抽出機構

本節では、前節で説明した構成要素のうちタグ

A Tag Extraction System for XML database
 - An application to the field of home care -
 Hokugou Wataru[†] Naomichi Sueda[‡] Tsunemi Nitta^{††}
[†]Graduate School of Engineering Oita University
[‡]Department of Computer Science and Intelligent Systems, Oita University
^{††}DigitalianNet & Co.,Ltd

抽出処理について述べる。図3はタグ抽出機構の概要である。過去に作成された大量の居宅介護サービス計画や居宅介護サービス支援システムによって作成されたサービス計画書を次回以降の居宅介護サービス計画作成の際、類似検索による参照が行えるように、要素内容を抽出する。まず、要素ごとにテキスト内容を形態素解析し、品詞に分解する。分解された文書は構文解析ルールをもとに構文解析が行われる。その後、タグ抽出ルールにより、タグ抽出が行われる。タグ抽出ルールとは、品詞の意味や文の構成から任意の要素に当てはまる内容を抽出するルールである。

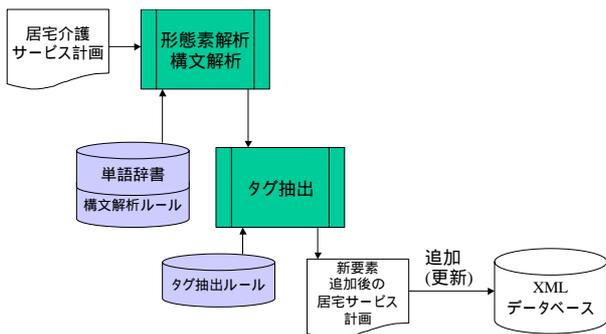


図3 タグ抽出機構

4. 実験

居宅サービス計画書のXMLファイルを用いて、要素の内容となる語や文書を抽出できるか確認する。今回の実験で使用したXMLデータは、被介護者5名の居宅サービス計画書(図1)をXMLデータ化したものである。

要素内容の抽出

居宅介護サービス計画には、「生活全般の解決すべき課題(ニーズ)」、「援助目標」、「援助内容」といった項目がある。今回の実験では、「生活全般の解決すべき課題(ニーズ)」の項目に書かれているテキストについてのみ要素内容の抽出を行う。テキストからは「病名」、「原因」と「要望」の3つのタグの要素内容を抽出する。以下は、抽出するタグの要素内容の説明である。

(1) 病名

居宅サービス計画中の病名を現す記述部分の抽出。病名の判別には概念辞書を用いる。

(2) 原因

被介護者が特定の介護サービスを受ける原因または、理由となる事象の記述部分を抽出。名詞「ため」、「せい」や助詞「より」などをキーとして前後の文を解析する。

(3) 要望

被介護者または、家族が居宅介護サービスに対して望むことを記述している部分の抽出。助動詞「たい」、形容詞「ほしい」など要望を表す語をキーとして前後の文脈を解析する。

実験結果

実験結果を表1、抽出結果例を図4に示す。表の「実際の要素内容数」とは、人手によって抽出された要素数で、「抽出した要素内容数」とは、タグ

	実際の要素内容数	抽出した要素内容数	*1
病名	20	13	13
原因	42	36	23
要望	39	28	20

*1 抽出した要素内容のうち有効な要素内容数

表1 実験結果

抽出機構によって抽出された要素数である。そして、「抽出した要素内容のうち有効な要素内容数」とはタグ抽出機構で抽出されたもののうち検索に用いることのできる要素数である。

「病名」タグに関しては、一部、辞書に登録されていない病名が抽出できなかったが、抽出されたすべての要素内容は、有効な要素内容と言え、うまく抽出できたといえる。「原因」・「要望」タグに関しては、現在の抽出ルールでは抽出できない表現があることがわかった。また、有効でない要素内容があった。これは、抽出された文書が助詞から開始されているなど文書としておかしいものである。このことから、これらのタグについては抽出ルールの強化が必要である。

```

<needs>
  右半身麻痺のため、和式トイレでの使用が困難である。
</needs>

↓

<needs>
  <原因>
    <病名>右半身麻痺</病名>のため
  </原因>、和式トイレでの使用が困難である。
</needs>

```

図4 抽出結果例

5. おわりに

本稿では、形態素解析と構文解析を用い、意味情報を持つ要素を付加する箇所を抽出するタグ抽出機構について説明した。あるタグ要素を表す内容をテキストから抽出し、そこにタグを追加することで、特定の質問に対して類似検索を可能となる。今後の課題として、抽出ルールの強化、さらに詳細な抽出を可能にするタグの追加、電子辞書の調整などがあげられる。また、実際に類似検索を行いどのような結果を得られるか検証する予定である。

参考文献

[1] 社会法人全国社会協議会, 新居宅サービス計画ガイドライン