

使いやすい言語処理のための実装例の提案

荒牧 英治^{*1} 森田 瑞樹^{*2} 仲村 哲明^{*1} 島本 裕美子^{*1} 宮部 真衣^{*1}

^{*1}京都大学 デザイン学ユニット ^{*2}東京大学 知の構造化センター

Toward Easy to use Medical Natural Language Processing Application

Eiji ARAMAKI^{*1} Mizuki MORITA^{*2} Tetsuaki NAKAMURA^{*1}
Ymiko SHIMAMOTO^{*1} Mai MIYABE^{*1}

^{*1}The Unit of Design, Kyoto University

^{*2}Center for Knowledge Structuring, The University of Tokyo

So far, each Natural Language Processing (NLP) system utilizes each data format. However, the medical application demands various constrains. This demand makes the development of each system a heavy burden. To deal with this problem, we represent various NLP tasks as a general task based on the single XML framework. This paper also proposes three NLP application examples using the proposed framework.

Keywords: Natural Language Processing , De-identification, User Interface, XML, NLP Application

1. はじめに

電子カルテの普及に伴い、医療情報分野において自然言語処理の導入の機会が増している。例えば、これまで病名の自動コーディング、類似症例検索、検査名の正規化など多様な試みが行われてきた。

しかし、研究レベルと応用性に必要なレベルの差異が大きく、現場での利用が困難な場合もある。本当に使えるシステムにするためには、単なる自然言語処理としての精度だけでなく、ヒューマン・コンピュータ・インタラクションや協調作業など関連分野の知見を取り込んだ適切な実装を付与することも必要である。しかし、利用シーンごとに最適な実装は異なり、また、求められる言語処理もコーディング、変換や検索など多様であり、個別に開発すると大きなコストがかかる。例えば、病名の自動ICDコーディングをwebブラウザ上で実装したとする。しかし、インターネットなど外部ネットワークを使えない場合には、これをスタンドアローンのアプリケーションとする必要性が生じる。

また、携帯端末から使うためにiPhoneなどのスマートフォンに移植する必要性が生じる場合もある。これら実行環境ごとに、移植を行うのに一定の手間がかかる。さらに、求められる処理もコーディングだけでなく、文章要約や翻訳など様々な処理が考えられる。このように、実行環境の数と求める言語処理の数の掛け算で生じるコストがかかるのが現状である。

そこで、本研究では、主要な言語処理を扱える一般化したデータ構造を提案する。提案するデータ構造はXMLやjsonで表現可能なシンプルなものである。本研究では、Web、タッチデバイス、コピー機といった3つの実行環境にて、匿名化、言い換え、翻訳という3つの言語処理を実装し、提案するデータ構造のフィージビリティを検証する。

2. 手法と結果

機械翻訳、自動病名コーディング、文章要約などさまざまな言語処理技術が医療分野でも適応されつつある。これらの処理内容や目的は多様であるが、多くの言語処理は、文章を入力として、その一部の文字列に対し

て、文字列リストを付与し出力するという形では一般化できる。

例えば、病名のコーディングは入力病名またはカルテ文章であり、それに対して候補となるコードのリストを返す。文章補完は、入力が入力文字列であり、それに対して尤度の高い後続文字列を返す。機械翻訳(自動翻訳)は、入力が入力文であり、それに対して翻訳結果英文の候補を返す。以上のように、主要な言語処理の一般化例を表1に示す。このように言語処理応用の多くは、ある文字からある文字までの間に対して(表の例ではXMLではsタグで表現)、n個の文字列を付与(c表の例ではcand属性で表現)するという形で表現できる。

本研究では、この構造を表現するXML形式を用い、3つのアプリケーションにてフィージビリティを検証した。実装は、タッチパネル上で動くスレートタイプ、紙を入力とし言語処理結果を付与してプリントするOCRタイプ、エクセル・アドインとして1クリックで実行可能となるアドイン・タイプの3つである。実装した言語処理は、匿名化、言い換え(難解語の簡易な言い換え)、症状表現の抽出であり、いずれも同一のデータ構造で動作可能であった。

アプリケーションの概観を図1から図4に示す。これらのツールの一部はウェブにて公開している¹⁾。

表1 様々な言語処理の入力と出力の枠組み。

処理	入力	出力
病名コーディング	病名 (例: 胃癌)	ICD コードのリスト <s cand=" C169">胃癌</s>
文章補完 サジェスト	文の途中まで (例: 早期胃)	後続する文 早期 <s cand=" 胃がん"></s>
機械翻訳	日本語文 (例: 胃癌)	英語文 <s cand=" stomach">胃 <s cand="cancer">癌</s>
文章要約	文章 (例: 腹部膨隆がなく、板状硬もない)	不要部を削除した文章 <s cand=" 平坦">腹部膨隆がなく</s>, <s cand=" 軟">板状硬もない</s>
匿名化	文章 (例: 横浜市在住の 70 歳男性)	個人情報等を*に置き換えた文章 <s>横浜</s>*市在住の<s>70</s>*歳男性.
文章簡単化	文章 (例: 麻疹に罹患した場合)	難解語に言い換え候補を付与した文章 麻疹に<s cand=" 病気がかる こと">罹患</s>した場合
副作用抽出	文章 (例: 胃がムカムカして)	副作用表現に MedDRA を付与した文章 <s cand="悪心">胃がムカムカ</s>して

(a)

At this time, she decided that she needed the prolapse fixed and actually wrote to the **North Hospital** to try and find a Gynecologist. Her letter was referred to Dr. **Earlamarg Pare** who saw her in the office and recommended vaginal hysterectomy. Patient has never had a pessary and refused one and denies nausea, fever, chills, dysuria, or stress incontinence on her current admission. PAST MEDICAL HISTORY: Significant for a lower left saphenous vein thrombosis diagnosed with lower extremity non-invasives and on May 2, 1993 with no evidence of deep venous thrombosis and clearance by Dr. **Ribreecheamner** for further intervention.

(b)

At this time, she decided that she needed the prolapse fixed and actually wrote to the **North Hospital** to try and find a Gynecologist. Her letter was referred to Dr. **Earlamarg Pare** who saw her in the office and recommended vaginal hysterectomy. Patient has never had a pessary and refused one and denies nausea, fever, chills, dysuria, or stress incontinence on her current admission. PAST MEDICAL HISTORY: Significant for a lower left saphenous vein thrombosis diagnosed with lower extremity non-invasives and on May 2, 1993 with no evidence of deep venous thrombosis and clearance by Dr. **Ribreecheamner** for further intervention.

(c)

At this time, she decided that she needed the prolapse fixed and actually wrote to the **North Hospital** to try and find a Gynecologist. Her letter was referred to Dr. **Earlamarg Pare** who saw her in the office and recommended vaginal hysterectomy. Patient has never had a pessary and refused one and denies nausea, fever, chills, dysuria, or stress incontinence on her current admission. PAST MEDICAL HISTORY: Significant for a lower left saphenous vein thrombosis diagnosed with lower extremity non-invasives and on May 2, 1993 with no evidence of deep venous thrombosis and clearance by Dr. **Ribreecheamner** for further intervention.

(d)

At this time, she decided that she needed the prolapse fixed and actually wrote to the **North Hospital** to try and find a Gynecologist. Her letter was referred to Dr. **Earlamarg Pare** who saw her in the office and recommended vaginal hysterectomy. Patient has never had a pessary and refused one and denies nausea, fever, chills, dysuria, or stress incontinence on her current admission. PAST MEDICAL HISTORY: Significant for a lower left saphenous vein thrombosis diagnosed with lower extremity non-invasives and on May 2, 1993 with no evidence of deep venous thrombosis and clearance by Dr. **Ribreecheamner** for further intervention.



図1 タッチデバイス上の匿名化実装

(a)ドラッグで選択範囲を選択.(b)匿名化すべき箇所がハイライトされる.(c) ハイライトされた箇所が匿名化されるべき箇所なら,こするマウス・ジェスチャを行うと,「***」と匿名化される.(d) すべてに適応ならば,丸を描くマウス・ジェスチャですべてが匿名化される.



図2 コピー上の匿名化実装

(a)コピーを構成する機械(PC+スキャナー+プリンター). (b)入力用紙. (c)出力(a)コピーを構成する機械(PC+スキャナー+プリンター). (b)入力用紙. (c)出力用紙 (a)コピーを構成する機械(PC+スキャナー+プリンター). (b)入力用紙. (c)出力用紙

細菌よりも小さく、電子顕微鏡でないと見えない病原体

病気にかかること

麻疹ウイルスによる脳症

高熱や重病のために脳に障害を起こした状態で、脳実質に炎症が見られないもの

15 歳以上で麻疹に罹患した場合、成人麻疹と表現されます。

小児期に自然感染した場合に比べて、肺炎や脳炎など重篤

な合併症を起こしやすく、入院が必要となることもあります。

国内では予防接種が行われていることもあるのですが、一部の年齢層では予防接種の効果が低く、再接種が推奨されることもあります。

麻疹にかかったことがあるかどうかは血液検査を行うことである程度判明しますが、血液検査にもEIA法、NT法、PA法など様々な方法があります。このうち、最も推奨されるのはEIA法という検査法です。医療機関によっては、EIA法以外の検査法を用いていることもありますので、もし自身が検査を希望される場合にはEIA法かどうかを確認してください。

酵素を用いた免疫測定法

EIA法で麻疹抗体価を測定、次のステップは？

病状が非常に重いこと

脳の炎症性疾患の総称

ある病気が原因となって起こる別の病気

ウイルス粒子を利用した抗体の測定法

あらかじめ特異抗原を結合させたゼラチン粒子を用いた検査法

細菌やウイルスと戦い、からだを守ってくれる。人間のからだの中で作られる物質

図3 コピー上の言い換え実装

難解な表現に注釈(赤い四角領域)が付与されて出力される。

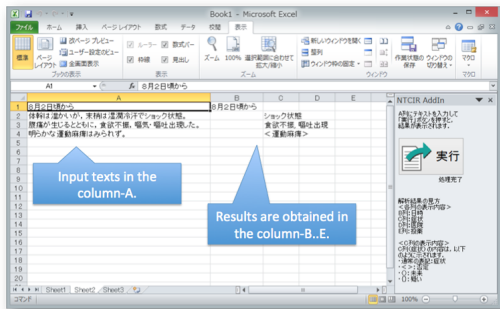


図4 エクセル・アドインとしての症状表現抽出実装

3. おわりに

医療分野での情報技術の応用に際しては、様々な制約がある。例えば、施設で統一したOSバージョンにすると

いう制約、ネット環境に接続不可能な場合があるという制約、紙媒体が使われうという制約、専用デバイスが使われる制約など、施設単位で独自の制約は様々である。このため、医療分野での実装は、一般の言語処理に比べて開発コストが高くなる傾向がある。今後、本研究で提案する枠組みが一般化すると、ある処理を手軽に複数の環境で試すことができ、言語処理の利活用が進むと考えられる。

ただし、提案する枠組みは今回対象とした言語処理に対応しているものの最適なものである保証はない。今後、様々な実装／処理に対応していくなかで、精査する必要がある。

最後に、ある一つのプログラムを容易にさまざまな実装で試せる可搬性は、医療システムの開発を加速するものと信じている。

参考文献

[1] MedNLP. <http://mednlp.jp>.