

2024年 12月 3日

報道関係者各位

国立大学法人 奈良先端科学技術大学院大学

## 医療データをコンピュータで解析し、適格な関連情報を提供する 大規模医療辞書を開発 核となるデータ「JMED-DICT mini」を公開 ～医療支援システムの基盤として幅広い利活用に期待～

### 【概要】

奈良先端科学技術大学院大学（学長：塩崎一裕）先端科学技術研究科の荒牧英治教授を責任研究者とした共同研究グループは、医療データ活用を促進するための大規模医療辞書「JMED-DICT」（<https://sip3-d2.naist.jp/jmed-dict.html>）の開発を進めています。膨大な医療データから、情報科学の自然言語処理技術などにより、医療用語を抽出し、分類・整理して国際的な医療統計基準などと紐づけることで、医療者らに関連する医療データを提供する辞書です。今回、そのコア部分（約1万語）を「JMED-DICT mini」として公開しました。

JMED-DICTは、研究機関での医療データ解析や、診断支援、医療の質や効率を向上させるDX（デジタルトランスフォーメーション）化などの医療アプリケーションへの応用を目的に設計されています。これまでにない広範な医療用語・医学概念を収載したこの辞書は、将来的に医療支援システムの基盤として活用されることが期待されます。

### 【背景と目的】

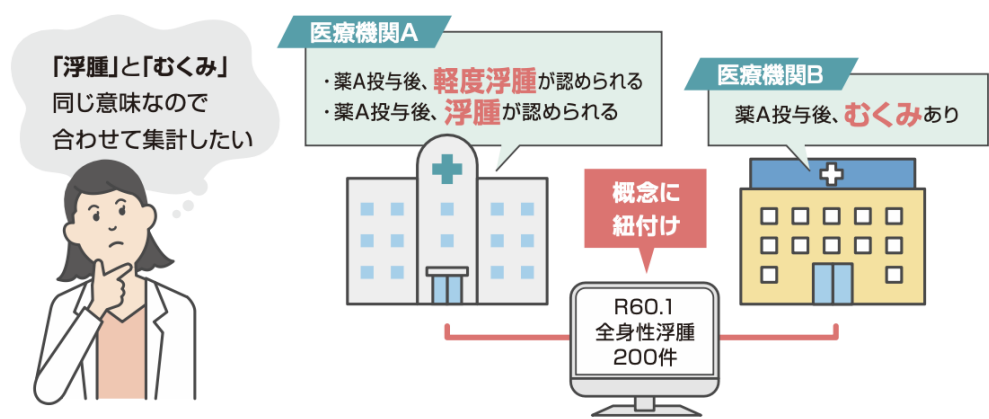
医療データの活用（図1, 2）には、電子カルテ（EHR）（注1）や患者記録（PHR）（注2）などに記述される様々なテキストを構造化し、医療知識を抽出する必要があります。医療テキストから医療知識を抽出するには、自然言語処理の技術を利用してテキストの処理を行う解析基盤が必要ですが、その開発のためには、医療用語や医学概念を統合・構造化した辞書が不可欠です。

既存のリソースに目を向けると、医学概念をコンピュータが理解しやすい方法で整理する巨大オントロジーがすでに存在している一方で、その維持管理に膨大なメンテナンスコストが必要とされています。そこで、本研究グループは、自然言語処理技術を応用し、従来よりも大規模・高品質でありながらも省コストで実現することを目指して辞書の開発を進めています。

**ユーザー**  
医療者

**目的**  
正確なイベントの発生頻度を把握

**精度**  
● ● ● ● ●



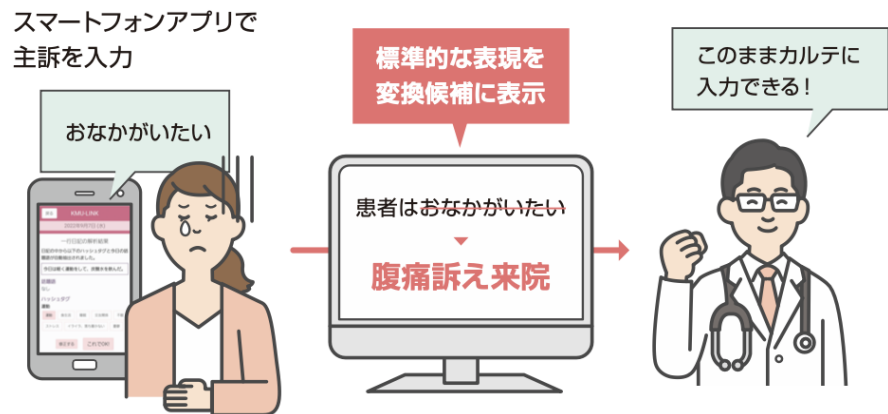
ICDなど国際的な体系に用語を紐付け、発生頻度を集計できる

図 1：医療統計への応用例

**ユーザー**  
医療者

**目的**  
カルテや診療記録の入力効率化

**精度**  
● ● ● ● ●



患者の記述がカルテに記載できるようまとめて、入力の手間が減る

図 2：カルテ入力への応用例

### 【研究の内容】

JMED-DICT は、次のようにして構築しました。まず、既存の医療テキストから、自然言語処理技術である NER（注 3）を利用して医療用語を抽出しました。次に、抽出された医療用語を分類・集計し、出現頻度順に JMED-DICT のエン트리として収載しました。さらに、それぞれのエントリに対して読み仮名や用語理解の難易度など様々なアプリケーションでの利用が想定される属性情報と、研究や医療業務に必要な、既存の分類体系へのリンク情報を付与しました。このようにして構築した初期データに対して、人手による情報の確認・修正と、プログラムによる自動推定を繰り返すことで半自動的にデータ全体の品質を向上させています。

JMED-DICT は以下 4 つのサブデータから構成されます。

1. **病名データ**：症状や病名を収載。国際疾病分類の ICD-10 や 11、標準病名などと紐付けます。
2. **医薬品データ**：医薬品名を収載。一般名や、薬効・作用部位などによって分類された ATC コードなどと紐付けます。
3. **検査データ**：検査表現を収載。代表的な臨床検査項目の分類と紐付けます。
4. **部位データ**：人体部分表現を収載。代表的な解剖学辞書の表現と紐付けます。

今回公開する内容は、全データ中から病名データおよび医薬品データの商用利用可能部分を抽出したもので、今後、公開されるフルバージョンのコア部分（頻度が高く最も利用可能性が高い部分）となります。

### 【今後の展開】

引き続きメンテナンス作業を進め、より大規模かつ高品質な辞書データの構築に向けて継続的にアップデートを続けていく予定です。また、辞書を応用した新たな言語処理技術の開発も進め、論文等で成果を発信します。

### 【用語解説】

(注1) EHR: Electric Health Record; 医療機関が患者情報を電子化して記録したデータ。

(注2) PHR: Personal Health Record; 患者個人の健康に関する情報を電子化して記録したデータ。

(注3) NER: Named Entity Recognition; テキスト中から固有表現(固有名詞や数量)を抽出する技術。固有表現抽出。

### 【お問い合わせ先】

<研究に関すること>

奈良先端科学技術大学院大学 先端科学技術研究科 情報科学領域

ソーシャル・コンピューティング研究室 JMED-DICT 作業班

TEL: 0743-72-5250 E-mail: sip3-d2@ml.naist.ac.jp

研究紹介ホームページ: <https://sip3-d2.naist.jp/index.html>

<報道に関すること>

奈良先端科学技術大学院大学 企画総務課 渉外企画係

TEL: 0743-72-5112 FAX: 0743-72-5011 E-mail: s-kikaku@ad.naist.jp