

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年6月20日現在

機関番号：83802

研究種目：基盤研究（C）

研究期間：2009～2010

課題番号：21590588

研究課題名（和文） シソーラス辞書を用いたがん患者家族のための連想学習につながる情報提供ツールの検討

研究課題名（英文） The information tool to support associative learning for cancer patients and their families using a thesaurus dictionary

研究代表者

石川 睦弓（ISHIKAWA MUTUMI）

静岡県立静岡がんセンター（研究所）・患者・家族支援研究部・部長

研究者番号：90324516

研究成果の概要（和文）：本研究の目的は、web 版がんよろず相談 Q&amp;A サイト

（<http://cancerqa.scchr.jp/>）を用いて、がん患者や家族の問題解決過程をサポートする情報ツールを検討することである。自由記載による 7885 人のがん体験者の悩みの調査結果を基に構築したがん患者の悩みデータベース（同サイトで公開）のテキストデータを用いた。テキストマイニングソフトを用いてデータ同士の関連性の分析を行い、作成したシソーラス辞書の一部をサイトの類似文章検索プログラムに実装した。

問題解決過程をサポートする情報ツールとして、利用者がアクセスした情報に関連する医療と暮らしにかかわる悩みや課題を包括的にわかりやすく表示する機能を追加し、試験的にサイトに実装した。

研究成果の概要（英文）：The purpose of this study is to support the problem-solving process of cancer patients and their families, using the information tool "Cancer consultation Q&A on the Web" (Shizuoka cancer center: <http://www.scchr.jp/>).

The relevant data was analyzed using text-mining software. Some thesaurus dictionaries were mounted in "the similar text retrieval program" of a site.

The information tool which collects information and supports a problem-solving process was examined. The function which displays the comprehensive information relevant to the information which the user accessed was added.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	900,000	270,000	1,170,000
2010年度	1,500,000	450,000	1,950,000
2011年度	700,000	210,000	910,000
総計	3,100,000	930,000	4,030,000

研究分野：医歯薬学

科研費の分科・細目：境界医学・医療社会学

キーワード：医療情報学

## 1. 研究開始当初の背景

現在、我が国では、国立がんセンターがん対策情報センターの「がん情報サービス」、米国国立がん研究所が提供するがん情報の翻訳版等を提供している「がん情報サイト」など、がんに関連した情報の集約・提供環境が整備されつつある。これらのサイトでは、主に疾患別の概要や治療、副作用症状への対応などの情報提供が行われている。

一方、がん患者や家族は、がんと診断された後、様々な時期に、様々な悩みや負担に直面する。悩みは、がん罹患や治療を受けることによる生活の変化や症状への対応、こころの問題、人間関係、就労や経済面の問題など多様で多岐にわたる。しかし、このような情報は多様であるが故に、広範囲のなかから探すようになり、また関連する情報を含めて収集するのは容易なことではない。

web版がんよろず相談 Q&A サイト

(<http://cancerqa.scchr.jp/>) では、2003年度に厚生労働科学研究の研究班が実施した全国がん体験者7,885名の悩みや負担等に関する実態調査結果を4階層の分類で体系的に整理した1万件余りのテキストデータからなる悩みデータベースを公開している。この調査は、がん体験者の悩みや負担を自由記載で調査したものであり、患者視点で具体的な多様な悩みの実態を明らかにした。

調査の結果、2万6千件の悩みの【要約文】が抽出され、これをwebで公開するために、さらに類似した【要約文】をまとめて1万件余りの短文（【統合文】）に集約し、悩みデータベースを構築した。このデータベースは、患者視点の悩みのデータベースであり、医療と暮らしにかかわるがん患者の悩みや負担を包括的にとらえている。「Web版がんよろず相談 Q&A」サイトでは、コンテンツの一つと

して、悩みデータベースを用いて、がん患者が抱える悩みや問題に近い【統合文】（悩みの短文）を検索し、公開している助言とあわせて閲覧することが可能である。Q&AのQuestion部分は、がん体験者自身の記述した文章をもとに作成された悩みの短文で構成されているので、利用者にとっては、具体的に自分の悩みや問題と合わせてイメージしやすいという特徴がある。また、悩みデータベースの検索方法の一つとして、インターネット検索に不慣れな者でも利用しやすいように、利用者が文章を入力して検索する自由文あいまい検索（「類似文章検索プログラム」）を開発し、実装した。本サイトの総訪問者数は、2011年の当サイトの利用状況を月平均でみると、訪問者数41,900名、助言参照（悩みと助言ページ閲覧）回数は、56,697件である。

そこで、悩みのデータベースのこのような特徴をいかして、利用者が自分の状況にあわせて、関連する情報もあわせて整理して入手し、がん患者や家族の問題解決過程をサポートする情報ツールの構築を検討することとした。

## 2. 研究の目的

がん医療の進歩により、がんの生存期間延長、治癒率は向上し、長期生存者も増加している。加えて、入院期間は短縮化し、通院治療が増加している。がん生存者は、いろいろな意味で不確かさを有するがんという「病い」を抱え、患者として診療を受け治療に参加する側面と、社会の中で一社会人として生活する側面のバランスをとりながら生活していかなければならない。

このような背景のなかで、がん患者や家族は情報へのアクセスを試みる。がん患者や家族にとっての情報は、病気や治療を理解する

知識としての情報だけではなく、生活領域も含めて生じた様々な問題に対処する行動を起こすための情報、不安や孤独感を癒すための共感できる情報などが含まれる。

2007年1月末に、インターネット媒体によるがん患者や家族を支援するためのツールとして、「Web版がんよろず相談Q&A」サイト (<http://cancerqa.scchr.jp/start>) を公開し、利用者や医療関係者、患者支援団体等の評価を得ながら改善を重ね運用してきた。この過程で、がん患者や家族をサポートする情報提供ツールについて考えると、利用者自身が、受動的に、断片的に、情報を閲覧するだけではなく、情報をつなぎ、考え、行動できるようなツールが必要であると考えた。

そこで、本研究では、「Web版がんよろず相談Q&A」 (<http://cancerqa.scchr.jp/start>) を用いて、がん患者や家族の問題解決過程をサポートする情報ツールの構築を検討する。

### 3. 研究の方法

#### (1) 悩みデータベースのデータクリーニング

悩みデータベースは、質問紙調査の患者の自由記述から、個人情報情報を削除後、助詞を整理した【要約文】と、複数の対象者の類似した悩みをまとめた【統合文】が存在する。患者の自由記述には、主語と述語の関係があいまいなものや、複数の悩みが並列で書かれていることが多い。【要約文】は、記述された内容の意味を損なわない範囲で、助詞を整理し、単文にすることで、係り受けが明確化されている。【統合文】はさらに、同じ意味と考えられる【要約文】をまとめたもので、個人の表現が、多くの患者が一般に用いる語に集約したデータとなっている。【統合文】データはクリーニング過程で患者が用いる一般的な語となっており、語の頻度や修飾関係を把握する目的にあっていてと考え、この2種のデータのうち、【統合文】データを用い

て、分析を行うこととした。これまでに進めてきた作業により、データのクリーニングはほぼ済んでいるが、データ分析過程での索引語の検出精度により、クリーニング作業を行う。

#### (2) データ分析

データの分析には、テキストマイニングソフト (Mining Assistant/R.2 (TRUSTIA /R.2 に同封) : ジャストシステム、Word Miner : 日本電子計算株式会社) の2種類のソフトウェアを使用する。この2つのソフトを用いて、結果を比較しながらクリーニングをすすめて精度をあげる。

第一に、形態素解析と索引語の検出を行う。悩みの短文についてその内容を特徴づける重要な単語である索引語を抽出する。そのために、まず形態素解析により、各単語に語形変化、品詞などの情報を付与し、単語の出現頻度もあわせてデータベース化する。抽出した索引語が悩みの短文の内容にどれだけ関係しているかを索引語の重要度として索引語に付与する。

第二に、類似内容の短文を抽出し、分類・関係付けを行う。主題となる語句は各グループの中からもっとも特徴的な言葉をテキストマイニングソフトで自動的に拾い出す。次に、主題分析機能を用いこの主題の関係性を確認し、重要と思われる主題やそれぞれの関係性を検討し、関係の深い名詞などを確認する。また、形容詞句と名詞の係り受けの関係に注目して分析を行う。これらの分析過程で得られたデータに基づき、シソーラス辞書を作成する。

#### (3) シソーラス辞書の検索システムへの実装

現在、web版がんよろず相談Q&Aサイト

(<http://cancerqa.scchr.jp/>) に実装している「類似文章検索プログラム」は、日本語自由文から各単語の係り受け関係を解析し、複合語を作成し、入力した文章の前後関係を判断してあらかじめ分類・統合された日本語文章データベースの複合語群を検索するシステムである。本プログラムは開発過程で一部類義語辞書を作成したが、本研究で作成した類義語を辞書に実装しプログラムの強化をはかる。

(4) Web版がんよろず相談Q&Aサイトの機能強化として、問題整理・問題解決を支援するツールを開発

テキストマイニングの過程で見いだした関係性を参照しながら、アクセスした悩みの短文に関連する情報を同様に悩みの短文で表示する方法を検討する。

#### 4. 研究成果

(1) 悩みの【統合文】は、9,871件で、一文書あたりの語句数は、最大176語、最小4語、平均27語であった。名詞句の出現頻度が高い語句は、手術1,060件、不安927件、治療818件、再発691件、仕事575件と続いた。係り受けの名詞句-形容詞句では、再発-不安73回、家族-心配49回、体調-悪い39回、転移-不安、医療費-高額などが上位にあがっていた。

悩みデータ同士の関連性の探索では、9,851件の悩みの短文データ全てで行ったところ整理がつきにくいいため、第1階層分類16の分類ごとに再分析を行い、全分類での分析と比較しながら、関連性を整理した。

(2) 利用者の情報入手支援をするための問題の整理、情報の集約へのナビゲーションとなるような手法、表示デザイン、パターンの検討を行った。テキストマイニングの過程で

見いだした関係性を参照しながら、アクセスした悩みの短文に関連する情報を同様に悩みの短文で表示することとした。現在、悩みのデータベースでは4階層の静岡分類を用いて悩みを分類しているが、最初にアクセスした情報(悩みと助言のページ)を中心とした関連情報を表示する際に、医療と暮らしにかかわる悩みや課題を包括的にわかりやすく表示するためにさらに項目ラベルを集約した。「診療(患者として診療を受ける)」、「日常生活(日常生活を営むために必要な活動や行動)」、「社会生活(就労や経済面など社会生活の問題)」、「関係性(家族など周囲の人々との関係)」、「こころの問題」の5項目とした。

新規に助言を作成する際、利用者が最初にアクセスした悩みと助言のページの助言の最後に、前述した5項目の悩みや課題にそって悩みに関連する情報を表示する機能を、試験的に追加した。

(3) 追加した機能を評価するため、logを用いて利用者の動きの変化を分析した。機能を追加する前の1ヶ月と追加後の1ヶ月のデータを用い、検索エンジンから直接サイトのQ&Aのページにアクセスした利用者の動きを分析した。

悩みのデータベースのなかでもアクセスの多い第4階層分類(以下、細分類とする)「抗がん剤による脱毛」に属する悩みの統合文を取り上げ、脱毛から他の悩みと助言ページに移動する動きを分析した。

「脱毛」の場合、導入前は、2番目に移動したページは同じ細分類に属する悩みと助言ページ移動が10と一番多かったが、導入後では、「治療が辛い(細分類ラベル)」が15と一番多く、他の分類への移動がみられていた。また、3番目、4番目の移動では、医

療費(4)や仕事(1)を選択している例もわず  
かではあるがみられていた。

しかし、追加する前1ヶ月の月平均ページ  
ビューは2.72ページ、追加後1ヶ月の月平  
均ページビューは2.15ページとこれまで同  
様平均2-3ページの間で推移していた。

また、関連する情報に移動するという点で  
は、他のページへの移動は導入後も少なく、  
導入前後で明らかな相違はみられなかった。

この理由として、現段階では、悩みに対し  
て作成した助言ページは、全体の悩みデー  
タースの一部であることや、表示する悩みの

【統合文】自体に直接最初にアクセスした

【統合文】の重要単語が含まれているとは限  
らず、助言のページを開いてみるまでは、関  
連している情報と判断しにくい点がある。そ  
こで、今後は、さらに利用者の情報集約や整  
理を支援する仕組みとして、悩みの短文では  
なく、最初の情報の重要単語を入れた複数語  
からなるKeywordを表示するなどの仕組みの  
再検討が必要と考えられた。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者に  
は下線)

〔雑誌論文〕(計0件)

〔学会発表〕(計2件)

- (1) 石川睦弓、がん患者／家族が求める情報、  
抱える悩み. 第24回がん看護学術集会.  
2010.2 . 静岡
- (2) 石川睦弓、北村有子. <Web版がんよろ  
ず相談Q&A>サイトの有用性と課題の検  
討. 第22回日本サイコオンコロジー学  
会. 2009年9月. 広島

〔図書〕(計0件)

〔産業財産権〕

○出願状況(計0件)

名称:

発明者:  
権利者:  
種類:  
番号:  
出願年月日:  
国内外の別:

○取得状況(計0件)

名称:  
発明者:  
権利者:  
種類:  
番号:  
取得年月日:  
国内外の別:

〔その他〕  
ホームページ等

#### 6. 研究組織

##### (1) 研究代表者

石川 睦弓 (ISHIKAWA MUTSUMI)  
静岡県立静岡がんセンター(研究所)・患  
者家族支援研究部・部長  
研究者番号: 90324516

##### (2) 研究分担者

北村 有子 (KITAMURA YUKO)  
静岡県立静岡がんセンター(研究所)・患  
者家族支援研究部・主任研究員  
研究者番号: 10364035

##### (3) 連携研究者

( )

研究者番号: