

## ユーザの要求に応じたニュース記事の表形式要約

西口駿祐†<sup>1</sup> 芋野美紗子†<sup>2</sup> 土屋誠司†<sup>3</sup> 渡部広一†<sup>4</sup>

会話によるコミュニケーションによって人のパートナーとなる知的ロボットの実現が望まれている。人とロボットの会話にニュース見出しを用いることで時事情報に関する話題を提供できる。しかしニュース見出しは抽象的な表現であるため、人との会話に適用することは難しい。また人によって興味を引く情報は異なってくる。そこで人の興味を引く時事情報を具体的に提供するために、ユーザの要求に応じてニュース記事を表形式に整理する手法を提案する。

### Organization In A Table Format From News Articles According To User's Request

SHUNSUKE NISHIGUCHI†<sup>1</sup> MISAKO IMONO†<sup>2</sup>  
SEIJI TSUTIYA†<sup>3</sup> HIROKAZU WATABE†<sup>4</sup>

Developing a robot to coexist with people, people hope that adopting a conversational interface to a way of communication between people. By incorporating headlines collected from Web-News sites into a conversation between people and robot, people think that robots are more familiar to them. To present the concrete headline, this study proposes Concrete Meaning Generation Method of headline with a Concept-Base and a Calculation Method of Degree of Association.

#### 1. はじめに

近年、情報技術の発達によりユーザ同士がインターネット上で情報を共有でき、その中にはテキスト情報も多く存在している。そのためユーザは大量の情報を入手することが可能となっている。しかし、情報が膨大であるため、ユーザは求める情報を的確に選択することが困難となっている。そのため、膨大であるテキスト情報をよりコンパクトにまとめることが求められている。その手法として検索、分類、要約があげられる。膨大な情報からキーワードを指定して検索することで、キーワードに関連性のある情報を取得できる。また、情報の内容によって分野に分類することで数多く存在するテキスト情報を少数にまとめることが可能である。しかしテキスト情報を少数にまとめても、1つのテキスト情報には多くの情報が存在するため、そこから瞬時に求める内容を知ることは難しい。よってテキスト情報を自動で要約する技術が求められている。

これまで様々な要約に関する研究が行われてきた。例えば文章から内容を表すと考えられる複数の文を抽出し、抽出した文を文章の要約とする手法<sup>[1]</sup>などが存在するが、文単位による抽出方法の場合は修飾語や助詞などの情報が含まれるため、文章要約の圧縮に限界が生じる。また抽出した文を羅列しているため、文の前後のつながりが悪いなど可読性に問題点が存在した。そこで近年では文単位でなく、

文章の内容を表すと考えられる語（重要語）を抽出し、重要語を組み合わせて生成した一文を文章の要約とする研究<sup>[2]</sup>が行われている。しかし重要語を組み合わせて生成した文は重要語間の助詞が不適切であることや重要語同士の係り受け関係がわからない可能性がある。さらにユーザが重要語についての知識を持っていない場合、要約した文の内容を把握することが困難であると考えられる。また重要な情報として認識する内容はユーザによって異なるため、抽出した重要語がユーザの知りたい情報であるか分からない。例えばA社が新作ゲームを販売したという記事の場合、ゲームの値段やタイトルについて知りたいユーザとゲームのシステム内容について知りたいユーザでは、抽出すべき重要語が異なると考えられる。

そこで本稿では Web ニュース記事から重要語を抽出して表形式に要約し、ユーザの要求に応じて表を出力するシステムの提案を行う。表形式に要約することで様々な利点が考えられる。例えば表には格納する語が何であるかを表す項目が存在するため、助詞などを省いた名詞だけで内容を把握でき、文以上に圧縮して記事内容を要約できる。またユーザが格納されている語を知らない場合でも、その語を格納している項目からその語が何であるかを把握できる。その上でユーザが他の重要語に対して詳細を求めた場合は表の左端に格納している重要語と項目を変更して表を再生成することで柔軟に対応できると考えられる。よってユーザごとの異なった要求に応じた表を提供できると考えられる。

本システムでは以下の処理を行うことで Web ニュース

†1†2 同志社大学大学院 工学研究科  
Graduate School of Engineering, Doshisha University  
†3†4 同志社大学 理工学部  
Faculty of Science and Engineering, Doshisha University

記事から表を生成する。

- 1)新聞社の Web サイトに存在する分野とニュース見出し文、その記事本文から名詞を取得し、名詞から重要語を決定する。
- 2)重要語に対応する項目を生成し、重要語を格納する。
- 3)生成した項目と重要語から表を生成する。
- 4)ユーザの要求に沿って表を改変する。

表は格納する語に対してその語が何であるかを表す項目を付与することで助詞などの品詞がない名詞から内容を把握できると考えられる。そのため、本稿では記事から助詞などの品詞を省いて名詞を取得している。上記の処理によって、ニュース見出し文とその記事本文をユーザの要求に応じた表に要約する。例として図 1 にニュース見出し文「告発報復の異動、判決後も」とその記事本文を入力した場合の処理を示す。

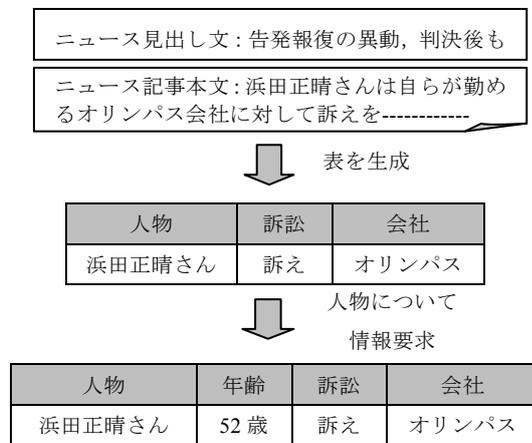


図 1 本システムの処理

図 1 より、入力したニュース記事から「浜田正晴さん」と「訴え」「オリンパス」が重要語として取得される。その後、「人物」「訴訟」「会社」が項目として生成される。よって人物に「浜田正晴さん」、会社に「オリンパス」、訴訟に「訴え」を格納した表が出力される。またユーザが「浜田正晴さん」について詳細を要求した場合は、浜田正晴さんの「年齢」が含まれた表に改変し、浜田正晴さんについてより具体的な情報がわかるようになっている。このシステムによって、ユーザの要求に応じた情報要約が可能になると考えられる。

## 2. 使用技術

### 2.1 時語知識ベース

時語知識ベースとは、時間判断システム<sup>[3]</sup>内に存在する知識ベースのことである。時語知識ベースには、「大正」や「月曜日」などの明示的に時間を表現する時語と、その語がいつ始まりいつ終わるのかなどを格納したテーブルが登録されている。表 1 に時語知識ベース内に存在する年号テーブルを示す。

表 1 時語知識ベース内の年号テーブル

ID	名称	ふりがな	時間軸	始	終
1	昭和	しょうわ	年	1926	1989
2	大正	たいしょう	年	1912	1926

表 1 は昭和, 大正を時語として格納している。昭和は 1926 年に始まり 1989 年に終わったため、時間軸に年を、始項目に 1926、終項目に 1989 を格納している。

### 2.2 NTT シソーラス

NTT シソーラス<sup>[4]</sup>とは単語の意味や概念を分類、整理して用語を階層的に体系化したものである。各節点に相当する語をノード、ノードに含まれる語をリーフと呼ぶ。NTT シソーラスには一般名詞の意味的用法を表す約 2700 個のノードの上位下位関係・全体部分関係が木構造で示されたものであり、約 13 万語のリーフが登録されている。

## 3. 提案システムの流れ

提案システムではあらかじめ、ニュース記事の分野ごとに必要と考えられる項目（初期項目）を設定しておく。まずニュース見出し文とその記事本文を入力し、ニュース記事から名詞を取得する。そして取得した名詞から初期項目に格納する名詞（初期名詞）を抽出する。次に初期名詞の係り受け関係を調べ、その結果からニュース記事の内容を表す重要語を取得する。その後 NTT シソーラスと生成する項目の候補となる語を格納している項目知識ベースを用いて、重要語の上位ノードから項目として適切な上位ノードを決定し、項目として生成する。そして生成した項目に重要語を格納することで、ニュース記事を表形式に要約して出力する。また、ユーザが詳しく知りたい項目または格納されている名詞を指定した場合、上記と同様に指定された名詞（指定名詞）に対する係り受け関係の結果から項目を生成する。そして指定名詞と重要語を項目に格納し、表を再出力する。図 2 に提案システムの流れを示す。

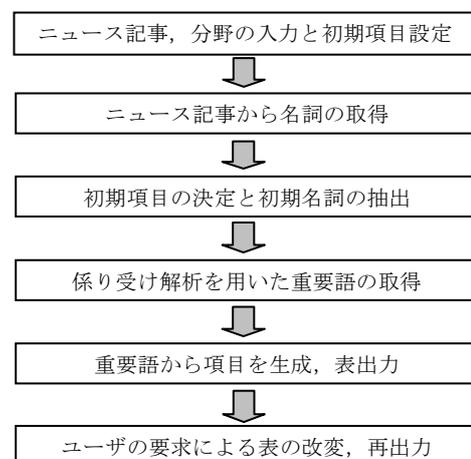


図 2 提案システムの流れ

各々の詳しい処理について、3.1~3.6 節で説明する。

### 3.1 ニュース記事、分野の入力と初期項目設定

本稿では「asahi.com (朝日新聞)<sup>[5]</sup>」に存在するニュース見出し文とその本文(ニュース記事)を研究対象とする。分野は、Web ニュースサイトでニュース記事が分類されている 17 個の分野を用いる。

Web ニュースサイトはニュース記事を内容によって分野に分類しているため、分野ごとにニュース記事から必ず生成される項目が存在すると考えられる。よって、あらかじめ分野ごとに必ず生成されると考える項目(初期項目)を目視で設定する。例えば「産業・経済」分野の場合はニュース見出し文「ローソンとヤフー、共同で食品宅配サービス開始へ」や「お肌のシミを測定、ソニーが新技術」など、記事に会社名が含まれるため「会社」項目を初期項目とする。

分野によっては初期項目を複数設定しており、「政治」分野の場合、初期項目は「人物」と「選挙」である。これは、「政治」分野のニュース見出し文および記事本文に人物が存在しない場合に「選挙」項目を初期項目として設定することを意味する。「音楽」分野の場合はニュース記事内に歌手が存在しなければ「楽器」項目を代わりに設定する。ただし「訃報」分野の場合は「日時・人物・職業・原因」が記事に必ず存在するため、この 4 つを初期項目として同時に設定した。また「科学」分野はニュース見出し文「A さん、ノーベル賞受賞式」や「地球の海面、20 年で 11 ミリ上昇」のように様々な内容の記事が入っており、「社会その他」分野は「事件・事故」「災害・交通」「裁判」「訃報」に分類されなかった社会に関する記事がすべて入っているため、必ず生成されると考えられる項目を決定することができない。そこで「科学」分野と「社会その他」分野についてはニュース記事における名詞の出現頻度によって初期項目を決定する。具体的にはニュース記事内で出現頻度が最も多い名詞が時語知識ベースに存在するかを調べ、存在する場合は初期項目を日時とする。また、出現頻度が最も多い名詞が時語知識ベースに存在しない場合はシソーラスを用いて名詞の 1 つ上位のノードを取得する。そして取得した上位ノードが、生成する項目の候補となる語を格納している項目知識ベースに存在する場合はその上位ノードを初期項目として設定する。項目知識ベースとは、NTT シソーラスに存在するノードから目視によって項目として不適切と考えられるノードを削除し、残ったノードを格納した知識ベースのことである。例えば「遊び」「人」「宝石」「紛争」「動物」など 2048 語が項目知識ベースに格納されている。ここで取得した上位ノードが項目知識ベースに存在しない場合はさらに上位のノードを取得し、項目知識ベースに存在するか検索する。この処理を繰り返して名詞の出現頻度による初期項目を決定する。

### 3.2 ニュース記事から名詞の取得

表に格納する重要語をニュース記事から抽出するため

に、ニュース記事に対して茶筌<sup>[6]</sup>による形態素解析を行い名詞を取得する。ただし、名詞には「携帯電話」や「人名」などの複合語が存在するため、複合語は 1 つの名詞と判断する。例えば「携帯電話」の場合は「携帯」と「電話」ではなく、「携帯電話」として名詞が取得される。

また、ニュース見出し文や記事本文中に「」で括られた部分は「外れ馬券も経費と認めて」のような発言もしくは「スマートキッチン」という商品名などの固有名詞である場合が多い。そこで、「」内に助詞が含まれている場合は「」の内容を発言と判断して削除し、名詞のみの場合は 1 つの名詞として取得する。

### 3.3 初期項目の決定と初期名詞の抽出

あらかじめ設定した項目(初期項目)に格納する名詞(初期名詞)を 3.2 節で取得した名詞から抽出する。ただし初期項目の個数によって初期名詞の抽出方法が異なる。

初期項目が 1 つの分野における初期名詞の抽出方法はまず各名詞の上位ノードに初期項目が存在するかを調べる。そして上位ノードに初期項目が存在する名詞を初期名詞として抽出する。上位ノードに初期項目が存在しない名詞は抽出しない。例えば「スポーツ」分野の場合、上位ノードに初期項目である「スポーツ」が含まれる名詞を抽出する。これにより初期名詞がスポーツ名、初期項目がスポーツとなる。

初期項目を複数設定した分野の場合、優先度の高い順に初期項目を調査し、初期名詞の抽出を行う。例えば「裁判」分野ではまず各名詞の上位ノードに「罪」が存在するかを調べる。上位ノードに「罪」が存在する場合は「罪」を初期項目に決定し、その名詞を初期名詞として抽出する。取得した上位ノードに「罪」が存在せず、「訴訟」が存在する場合は「訴訟」を初期項目として決定し、その名詞を初期名詞として抽出する。例えば「A 氏が所得税法違反で起訴された」という文の場合、初期名詞は「所得税法違反」が抽出され、所得税法違反の上位ノードに罪が存在するため初期項目は「罪」となる。また、「3 月に C さんが国を提訴」という文の場合、初期名詞は「提訴」が抽出され、提訴の上位ノードに訴訟が存在するため初期項目は「訴訟」となる。

### 3.4 係り受け解析を用いた重要語の取得

「南瓜<sup>[7]</sup>」を用いて係り受けの解析を行い、初期名詞と他の文節間の関係がどうなっているかを調べる。ここで、初期名詞が存在する文は記事の中で最も記事の内容を表す重要な文であると考えられる。よって初期名詞に係っている名詞と、その名詞に係っている名詞を重要語として取得する。

### 3.5 重要語から項目を生成、表出力

3.4 節で取得した重要語から項目を生成する。まず重要語の 1 つ上位のノードを取得し、その上位ノードが項目知識ベースに存在するかを調べる。重要語の上位ノードが

項目知識ベースに存在する場合は、その上位ノードを重要語の項目として生成する。取得した上位ノードが項目知識ベースに存在しない場合はさらに上位のノードを取得し、項目知識ベースに存在するかを調べる。項目知識ベース内に存在する上位ノードを取得するまでこの処理を繰り返し、取得した場合はその上位ノードを項目として生成する。

これまでの処理によって初期項目を決定し、初期項目に格納する初期名詞が抽出された。また、初期名詞から重要語を取得し、重要語の項目が生成された。次に初期項目と初期名詞、および重要語とその項目を組み合わせることで表生成を行う。例として裁判分野に存在するニュース見出し文「外れ馬券も経費と認めて」脱税、無罪主張」とその記事本文を提案システムにかけた場合の出力結果を図3に示す。

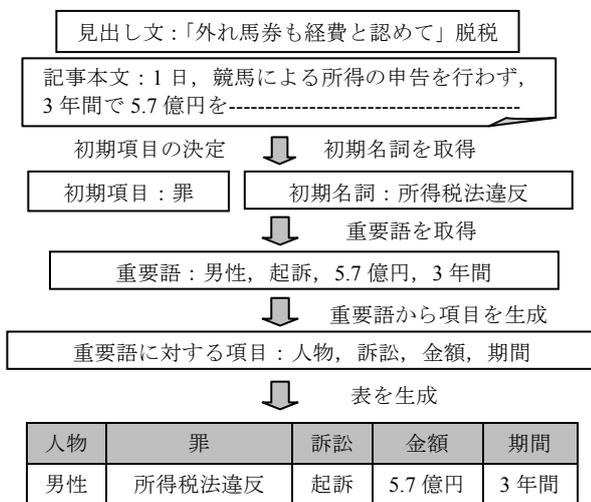


図3 提案システムの出力結果

図3の入力記事は裁判分野であるため、上位ノードに罪を持つ所得税法違反が存在することから「罪」を初期項目として決定し、「所得税法違反」を初期名詞として抽出する。次に所得税法違反を含む文の係り受け関係を調べ、所得税法違反に係っている「起訴」と起訴に係っている「男性」「5.7億円」「3年間」を重要語として取得する。そして重要語の項目として「人物」や「訴訟」が取得され、表を生成する。図3より、男性が3年間で5.7億円の所得税法違反により起訴された内容の結果が出力される。

### 3.6 ユーザの要求による表の改変、再出力

生成した表に格納されている名詞について、ユーザがより詳しく知りたいと要求した場合にその要求に応じて表を改変する。まずユーザが3.5節で生成された表から詳しく知りたい名詞を指定する。次にユーザが指定した名詞(指定名詞)の記事から検索し、指定名詞が含まれる一文を抽出する。そして指定名詞の係り受け関係を調べ、指定名詞に係っている名詞を重要語として取得する。その後、NTTシソーラスを用いて重要語の1つ上位のノードを取得し、項目知識ベースに存在するかを調べる。上位ノードが項目

知識ベースに存在する場合はその上位ノードを重要語の項目として生成する。取得した上位ノードが項目知識ベースに存在しない場合はさらに上位のノードを取得し、項目知識ベースに存在するかを調べる。項目知識ベース内に存在する上位ノードが取得できるまでこの処理を繰り返し、取得した場合はその上位ノードを重要語の項目として生成する。そして指定名詞と重要語、および生成した項目を組み合わせることで表の再出力を行う。例として図3に存在する表に対してユーザが「男性」についてより具体的に知りたいと指定した場合の出力結果を図4に示す。

人物	罪	訴訟	金額	期間
男性	所得税法違反	起訴	5.7億円	3年間

↓ 男性から重要語を取得

重要語: 39歳, 会社員, 大阪市

↓ 重要語から項目を生成

項目: 人物, 年齢, 職業, 地域

↓ 表生成

人物	年齢	職業	地域	罪	訴訟
男性	39歳	会社員	大阪市	所得税法違反	起訴

図4 提案システムの出力結果

ユーザが「男性」を指定した場合、指定名詞は男性となる。次に男性を含む文の記事から抽出し、男性について係り受け関係を調べる。そして男性に係っている「39歳」「会社員」「大阪市」「所得税法違反」「起訴」を重要語として取得する。その後、指定名詞と重要語から「人物」「年齢」「職業」「地域」「罪」「訴訟」を項目として生成する。指定名詞と重要語、および生成した項目を組み合わせることで図4の表が出力される。これにより、所得税法違反で起訴された男性が年齢39歳の大阪市に住む会社員であることが把握できる。

## 4. 実験・評価

### 4.1 評価実験

朝日新聞社のWebニュースサイトからニュース見出し文およびその記事本文を分野ごとに10件ずつ、合計170件取得し、テストセットとして扱う。このニュース記事170件から本システムによって表出力した後、被験者3人に①「生成した表から記事本文の内容が把握できるか」②「改変した表はユーザの要求に正しく沿っているか」の2つの観点で評価を行った。①は、本システムによって生成した表とニュース見出し文を比較して同じ内容を表すと考える物同士で線引きを行い、2人以上が○の場合に正解、それ以外は不正解とした(図5)。②は生成した表から一語を指定して改変した表と記事本文を比較して、表の改変部分は指定した語を説明しているか○、×で評価した(図6)。また、目視によって③「項目とその項目に格納されている名

詞の正当性はとれているか」④「重要語から生成した項目と分野の正当性はとれているか」の2つの観点でも評価を行った。

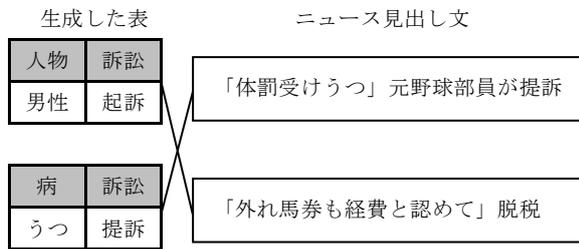


図5 評価方法①の一例

図5は、ニュース記事3件の見出し文と、そのニュース記事から生成した表である。ニュース見出し文「告発報復の異動、判決後も」は「判決」から表の「訴え」を連想することで線引きが可能であると考えられる。ニュース見出し文「外れ馬券も経費と認めて」脱税は表から「脱税」と同じ違反行為である「所得税法違反」が連想されることにより、図5の線引きが可能であると考えられる。ニュース見出し文「体罰受けうつ」元野球部員が提訴は「提訴」が生成した表にも存在するために線引きされると考えられる。

人物	罪	訴訟
男性	所得税法違反	起訴



人物	年齢	職業	地域	罪	訴訟
男性	39歳	会社員	大阪市	所得税法違反	起訴

図6 評価方法②の一例

図6は、生成した表からユーザが男性について詳細を要求した場合に表を改変して再生成した図である。ユーザが指定した「男性」に対して「39歳」「会社員」「大阪市」が取得された。男性に対して改変した表が男性を説明している場合は○、表に格納されている名詞が1つでも男性を説明していない場合は×で評価を行った。3人のうち2人以上が○とした場合はユーザの要求に応じた表を生成できたと考え、正解の評価とした。○が2人に満たなかった場合はユーザの要求に応じた表を生成できなかったと考え、不正解の表であると評価した。評価方法③では生成した項目とその項目に格納されている名詞との正当性を目視で評価する。例えば罪項目に所得税法違反、スポーツ項目にサッカーがされている場合に、サッカーとスポーツ、所得税法違反と罪は正当性があるかを目視で評価する。評価方法④では分野とその分野で生成した項目との正当性を目視で評価する。例えば訃報分野において、年齢項目を生成した場合は記事内容に適していると考えられるが、結婚項目だと

訃報分野では不適切であると考えられる。

#### 4.2 実験結果と考察

評価実験を行った結果、①の評価については66.7%、②の評価は48.3%、③の評価については69.1%、④の評価については48.7%の精度を得た。図7は、評価実験で使用した朝日新聞社Webサイトの「訃報」分野のニュース見出し文「元福岡県弁護士会長の池永満さん死去」とその記事本文を本システムに入力した場合の出力結果である。

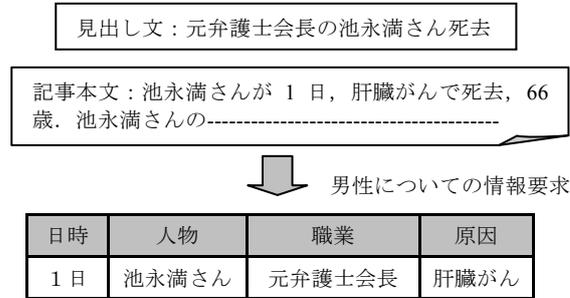


図7 本システムの出力結果

図7より、「日時・人物・職業・原因」項目が生成され、対応する名詞が格納された表が出力された。

評価方法①の結果について、図7よりニュース見出し文「元福岡県弁護士会長の池永満さん死去」とその記事本文から重要語を抽出して表形式に要約することで1日に元福岡県弁護士会長の池永満さんが肝臓がんで亡くなったという内容が瞬時に把握できる。本稿では生成した表のうち66.7%の確率で、表から記事内容を把握することができた。文形式ではなく表形式に要約した場合でも記事内容を把握できることが示された。よって文の形式で要約を行うよりも圧縮した要約が可能になったと考えられる。

次に図7で出力された表より、ユーザが池永満さんについての詳細を要求した場合の出力結果を図8に示す。

日時	人物	職業	原因
1日	池永満さん	元弁護士会長	肝臓がん



池永満さんの情報詳細

人物	年齢	日時	武器	宗教儀式	日時
池永満さん	66歳	1日	肝臓がん	通夜	2日

図8 池永満さんを指定した場合の出力結果

図8より、ユーザが「池永満さん」についての詳細を要求した場合、「池永満さん」に係っている「年齢」や「通夜」、通夜の「日時」を格納した表が出力された。評価方法②では、指定した「池永満さん」と生成された「年齢」「通夜」「日時」を比較して評価を行った。図8では、池永満さんを説明している年齢は正解であると評価されるが、池永満さんを説明していない宗教儀式、日時は不正解と評価され

る。よって生成された表が1つでも間違っているため本稿の評価は不正解となる。この結果から指定名詞の詳細を表す語は指定名詞に係っている語が重要であるといえる。

評価方法③の結果より、重要語の項目はNTTシソーラスによって重要語の上位ノードから生成しているため、約70%の評価結果となった。例えば図8では66歳という名詞は年齢であるために年齢項目は正解であると評価した。しかし、肝臓がんという名詞からは武器項目が生成されてしまったため間違いであると評価した。本評価結果より、表の名詞が分からない場合は、名詞を格納している項目を確認することで名詞の概要が把握できると考えられる。

評価方法④の結果より、約50%の確率で分野に対して適切な項目を生成することができた。図8では計報分野に対して宗教儀式や日時という項目が生成された。計報分野では人物がなくなった日時に関する記事が存在すると考えられるため、宗教儀式項目は計報分野に対して適切な項目であるといえる。

また評価結果によって、本システムではいくつかの失敗が存在することがわかった。失敗内容を考察した結果2つの改善案が考えられる。

1つ目は、ニュース見出し文および記事本文から名詞を抽出する際に固有名詞に対する処理が行うことである。例えば記事本文に存在する固有名詞「きやりーばみゆばみゆ」に対し、人名ではなく動詞や未知語と判断されてしまった。このような結果は主に芸能や文化の分野で多く見られた。これは概念ベースやNTTシソーラス内に固有名詞が定義されていないため、固有名詞が未知語と判断されたためであると考えられる。よって、未知語と判断された名詞については、Webを用いて概念ベースやNTTシソーラス内に存在する名詞に置き換えて処理することで精度が向上すると考えられる。

2つ目は、多義を持つ語の判別である。例えば「Aさんがんで倒れる」という文が記事にあった場合、「がん」は武器である銃と病気である癌の2つの意味合いが考えられるため、生成される項目に「武器」という結果が出力されてしまった。そこで、生成された項目とその分野の関連の強さを求め、関連性の高いものを項目とすることで精度の向上が期待できると考えられる。

## 5. おわりに

今回、ニュース見出し文とその記事本文をユーザの要求に応じて表形式に要約する研究を行った。本研究では茶筌による形態素解析と南瓜による係り受け解析を用いてニュース記事の構文解析を行い、ニュース記事から抽出した名詞から項目を生成して、表を出力するシステムを構築した。また本システムでは、生成した表からユーザが情報の詳細を要求した場合、ユーザの要求に応じて表を改変、再出力

も行った。その結果、66.7%の確率で要約した表から内容が把握でき、48.3%の確率でユーザの要求に応じた表を生成することができた。本システムによって、ユーザが興味のある時事情報を表形式に要約することが可能になる。

## 参考文献

- 1) Edmundson, H. "New Methods in Automatic Abstracting." *Journal of the ACM*, 16(1969).
- 2) 畑山満美子, 松尾義博, 白井諭, "重要語句抽出による新聞記事自動要約", 自然言語処理, Vol.9, No.4, pp55-73, 2002.
- 3) 岩瀬元秀, 渡部広一, 河岡司: "文の意味理解に基づく常識的時間判断システムの構築", 信学技報, AI2006-52, Vol.106, No.587, pp.1-8, 2007.
- 4) NTTコミュニケーション科学研究所監修, "日本語語彙体系", 岩波書店, 1997.
- 5) "asahi.com: 朝日新聞社の速報ニュースサイト", <http://www.asahi.com/>
- 6) ChaSen - 形態素解析器, <http://chasen-legacy.sourceforge.jp/>, 奈良先端科学技術大学院大学情報科学研究科自然言語処理学講座(松本研究室), 2011.
- 7) 松本裕治: "形態素解析システム「南瓜」" <http://chasen.naist.jp/hiki/ChaSen/>