

民生機器における知識データベースの構築と応用

Knowledge Database and Its Application for Consumer Electronic Products

梶 雅代, 堀内 直明, 莪山 真一

Masayo Kaji, Naoaki Horiuchi, Shinichi Gayama

要 旨 民生機器において利用可能な知識データベースを構築し、それを用いた新たな情報検索手法を開発した。本研究における知識データベースとは、機器に関連する大量のテキスト情報を解析して得た結果をもとに、機器固有の「知識」を体系化したデータの集合を指す。これを利用することによって、システムはさまざまな検索入力に対して機器固有の「知識」を反映した検索結果を出力することができる。本手法をカーナビゲーションシステムの「ドライブプランナー」機能に適用した結果、指定条件（同行者、観光目的）を満たす観光スポットの推薦を行うことができた。

Summary We developed a “Knowledge Database” which is highly suited to consumer electronic products and we also proposed an information retrieval system as a practical use of its architecture. The Knowledge Database we defined has a well-organized data structure which comprises a set of product-specific information generated based on the analysis of a great deal of text data related to the product.

It was expected that the database would enable the information retrieval system to provide appropriate results that reflected some product-specific knowledge. As a result of applying the method to the “Drive Planner” functionality on the car navigation system, we succeeded in retrieving a variety of spots for sight seeing that could meet specified requirements such as region, purpose of trip and the presence of passengers.

キーワード : 知識データベース, 情報検索, 意味定義, 意味関係, 情報推薦, ドライブプランナー, 観光スポット情報

1. まえがき

近年、インターネットの高速化と情報蓄積装置の大容量化に伴い、多種多様な情報をだれもが大量に、かつ容易に取得できるようになった。また、インターネット利用環境の整備が進むにつれ、同様の環境は家庭内に留まらず移動中や公共の場においても実現されつつある。ところが、大量の情報蓄積が可能となり情報の利用範囲（目的、シーン）が拡大する一方で、蓄積された情報から所望の内容を的確に見出す（検索する）ことが次第に困難になってきた。そのため、最近の情報技術分野では、情報検索を効率化し、利用者支援のための技術開発が盛んに行われている。

大規模なデータを対象とした情報検索には、膨大なデータを効率的に取り扱うためのインデクシングや、その情報と検索結果を直感的に把握するための視

覚化が重要である。これらを実現するには、あらゆる情報の意味を体系化することが必要である。一方で、情報を体系化することは情報自体を理解することに等しく、コンピュータプログラムを知的に動作させるための手段となる。すなわち、体系化された情報はコンピュータの持つ一種の「知識」と考えることができる。情報技術分野では、この「知識」を体系化したものを知識データベースと呼び、情報検索支援や顧客管理における利用が見込まれている。

知識データベースを構築するためには、さまざまな情報を「知識」として形式的（意味表現）かつ体系的（意味体系）に書き表す必要があるが、我々が取り扱う情報はテキスト、画像、音声、映像など多様な表現形態を取っており、それらをすべて体系化することは一般に困難である。

しかしながら、どのような形態の情報であっても、それらを表現する共通指針を見出せば各情報内容の形式化と全体の体系化が可能となる。例えば、我々はさまざまな事象を説明する際に「言葉」を用いるが、それと同様に情報内容を表現する場合にはテキスト情報を用いることができる。これは、膨大な情報を記述し活用するためのメタデータ(情報を説明する情報)として近年注目されている。また、テキスト情報の意味内容を理解することは、情報検索システムのみならず、質問応答システム、音声対話システムといった情報と人間とを結ぶインターフェースを実現する上で極めて重要である。このようなことから、テキスト情報とその処理技術は、情報機器のユーザインターフェース支援を含めた今後のマルチメディア情報の知識データベース化の基盤となっている。

本研究は、民生機器において利用可能な知識データベースを構築し、新しい情報検索機能および情報推薦機能を実現することを目的とする。具外的な手段としては、民生機器に関連する大量のテキスト情報を分析し、得られた結果をもとに機器固有の「知識」を体系化したデータベースを用いる。これを利用することによって、システムはさまざまな検索入力情報に対して機器の「知識」を反映した検索結果を出力することができる。

本稿では、まず民生機器における情報検索機能とその課題について述べ、次に我々が提案する民生機器における知識データベースの構築および活用方法を解説する。そして、一つの応用例としてカーナビゲーションシステムの機能である「ドライブプランナー^(注1)」⁽²⁾⁽³⁾に適用した結果について報告する。

2. 民生機器における情報検索機能

2.1 従来の課題

カーナビゲーションシステムを始めとする民生機器の情報検索機能は、入力したキーワードに合致する、あるいはそれを含む情報を提示するものが一般的であるが、ユーザが入力したキーワードが検索対象(目的とする情報)の記述内容に含まれない場合には検索結果を出力することができない。また、これらの方法では、同じ意味であっても異なった読みやつづりを持つ単語には対応できず、検索漏れや誤った検索結果を出力するという問題があった。

2.2 知識データベース適用によって期待される効果

上述の通り、知識データベースはさまざまな種類

の情報の意味表現と意味体系から形成される。これを情報検索に適用する利点は、情報の特徴や意味内容に基づいて、内容に即した検索や再利用が効率的に行えることである。例えば、カーナビゲーションシステムは、地図情報や観光スポット情報など大量の情報をもとにさまざまな動作を行うが、多機能化が進んだ上に情報量も増大したことから、情報検索に要するユーザの労力は極めて大きい。

そこで、知識データベースを適用すれば、あいまいな入力に対しても機器内部の体系化された情報や、走行目的、走行位置などの状況を反映した明確な検索結果の取得が可能となり、機器の使い勝手を大きく向上させることが期待できる。

3. 民生機器における知識データベース構築と活用

本章では、我々の提案する情報検索システムを枠組みとして、民生機器における知識データベース構築と活用方法について説明する。

3.1 全体構成

図1に、知識データベースとそれを含む情報検索システムの構成を示す。2つのデータベースのうち、D1が本研究の中心となる機器情報の意味表現と意味体系を保持する知識データベースであり、D2が情報検索の対象となる機器情報そのものを保持するデータベースである。また、①～④の過程について以下に記述する。

- ① 検索に用いる複数の入力単語群を取得する。
- ② D1を用いて、①で得た単語群と有意な意味関係にある他の単語群を抽出する。
- ③ ①②で得た単語群をもとにD2の機器情報群を検索する。
- ④ ③で得た機器情報群を同時に計算された該当スコアを用いて並び替え、出力する。

3.2項では、まず知識データベース(D1)の構築方法について解説する。次に3.3項で「知識データベースの活用と関連単語出力」と題して、①②を説明する。最後に3.4項の「機器情報検索と出力順序の決定」において、③④を解説する。

3.2 知識データベース構築方法

知識データベースを構築するには、次の4つの過程を必要とする。

- (1) テキスト情報の収集
- (2) テキスト解析
- (3) 知識の抽出(意味表現の形式化)

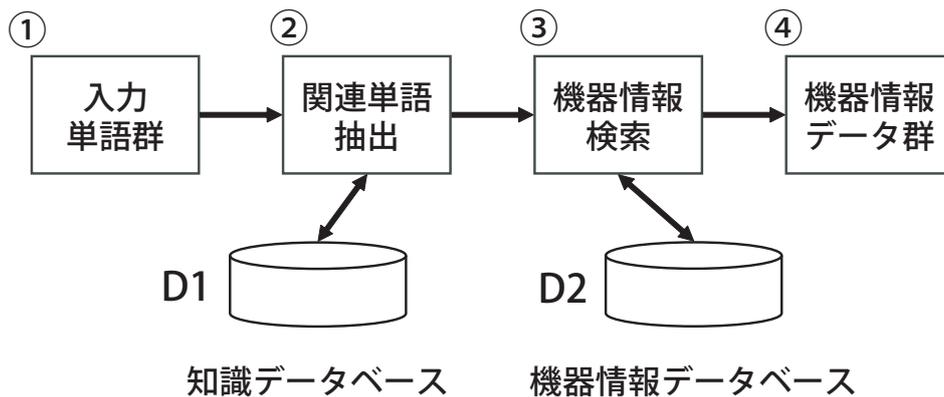


図1 知識データベースを用いた情報検索システムの構成

Fig.1 Information Retrieval System with Knowledge Database.

(4) 知識の関連付けとデータベース化 (意味体系の構築)

各項目について順に解説する。

(1) テキスト情報の収集

知識データベースを構築するには、まず対象とするシステムに関連する大量のテキスト情報が必要である。さらに、対象とするシステムが民生機器である場合、テキスト情報の内容も重要である。なぜなら、インターネット上の情報検索支援や顧客管理システムが一般的で幅広い表現を網羅する必要があるのに対して、民生機器は機器の性質に特化した固有の情報源を扱わなければいけないからである。その上、民生機器の情報には日常生活において一般的でない表現も多く見られ、ユーザの理解を容易にするためにはこれらを木目細かく処理する必要がある。つまり、民生機器を対象とする場合には、情報量ではなく情報の特質に重点を置いた知識の体系化が重要といえる。

そこで本研究では、民生機器が内部に保有する情報群と、機器固有の情報を一般的な表現で記述したさまざまなテキスト情報を利用した。情報の範囲を機器内部に限定することは、同機器を扱う上で確実性の高い知識を抽出できるという利点がある。また、機器固有の表現を一般化する情報を加えて情報量を拡張することは、知識データベースを用いた民生機器の情報検索操作において、入力単語の許容範囲を拡大させる効果がある。

(2) テキスト解析 (形態素解析 / 構文解析)

項目 (1) で収集したテキスト情報群を知識として扱うためには、各情報の意味を表現する手段が必要である。そこで本研究では、テキスト解析によりそれを取

得することを試みた。テキスト解析方法としては、代表的な言語解析手法である形態素解析技術と構文解析技術を採用した。

まず、形態素解析を用いて対象とするすべてのテキスト情報を複数の単語群に分解する。次に、形態素解析によって得られた単語群に対し構文解析を施すことによって、各単語間の相互関係 (係り受け) を抽出する。係り受けとは、国文法において「文末単語以外のすべての単語と後に出現する他の単語との修飾関係」のことである。以上のように、本研究ではテキスト情報の意味表現手段として、各情報から抽出した単語群とその係り受け関係を用いる。

(3) 知識の抽出 (意味表現の形式化)

次に、項目 (1) で収集したテキスト情報群から知識の抽出、すなわち意味表現の形式化を行う。項目 (2) では、意味表現の手段としてテキスト情報に含まれる単語群とその係り受け関係を抽出した。単語間の係り受け関係がそれらを含むテキスト情報の断片的な意味を表すことは「係り受け」の定義上自明である。そこで本研究では、特に単語間の係り受け関係に注目し、【意味定義】【意味関係】の対応関係を用いて意味表現の定式化を行うこととした。【意味定義】は、意味関係に用いる単語群、【意味関係】は、意味定義に基づく係り受け関係を表す。レストラン情報の例を用いて意味表現について以下に説明する。

(例) レストラン情報

(1) 収集したテキスト情報

(店名) ○△レストラン

(その他の情報) ファミリーレストラン / 10 時～

22時 / 火曜定休日 / 東京

(店の紹介)

家族連れでもくつろげるゆったり空間がおすすめ
個室あり。

(2) テキスト解析結果

形態素解析結果：ファミリー、レストラン、家
族連れ、ゆったり、空間、個室…

係り受け解析結果：家族連れ—おすすめ、くつ
ろげる—空間、ゆったり—空間、…

(3) 意味表現の定式化

【意味定義】ファミリー、レストラン、家族連れ、
ゆったり、空間、個室…

【意味関係】家族連れ—おすすめ、くつろげる—
空間、ゆったり—空間、…

まず、(1) で収集したレストランの店名や紹介文な
どのテキスト情報に対して (2) でテキスト解析を施す
と、形態素解析結果の単語群と、構文解析結果の係り
受け関係群が得られる。次に、(2) で得られた結果を (3)
に示す形式で意味表現の定式化を行う。本項目では一
例を挙げたが、他のテキスト情報についても同様の手
段で意味表現の形式化を行う。

(4) 知識の関連付けとデータベース化 (意味体系の構築)

本項目では、知識データベース構築の最終段階で
ある知識の関連付け、すなわち対象としたテキスト情
報全体に対する意味体系の構築について記述する。

本研究で言う意味体系とは、項目 (3) で形式化した
テキスト情報の意味表現を分析し、共通の意味定義 (単
語群) をもつ意味関係 (係り受け関係) を互いに関連

付けるとともに、各々の意味関係の重要度を数値化し
たデータの集合を指す。以下、3つの意味表現の例を
用いて具体的な方法を示す。

(例1) レストラン情報の意味表現の例

【意味定義】レストラン、家族、ゆったり、空間、
個室…

【意味関係】家族—レストラン、家族—個室、ゆっ
たり—空間…

(例2) 結婚式情報の意味表現の例

【意味定義】結婚式、教会、カップル、週末…

【意味関係】結婚式—週末、教会—カップル…

(例3) 遊園地情報の意味表現の例

【意味定義】遊園地、乗り物、カップル、家族、週末、
興奮…

【意味関係】遊園地—カップル、遊園地—家族、
週末—家族、興奮—乗り物…

まず、意味関係の関連付けについて説明する。例1
と例3の意味定義のように「家族」という共通の単語
が存在する場合、それを含む意味関係として例1の「
家族—レストラン」「家族—個室」、例3の「遊園地
—家族」「週末—家族」の各々の意味関係を「家族」
を中心として関連付ける。一方、例2と例3の意味
定義には「週末」と「カップル」が存在する。この場
合には、例2の「結婚式—週末」と例3の「週末—
家族」を「週末」を中心として関連付けるとともに、
例2の「教会—カップル」と例3の「遊園地—カッ
プル」を「カップル」を中心として関連付ける。この
例を視覚化したものを図2に示す。

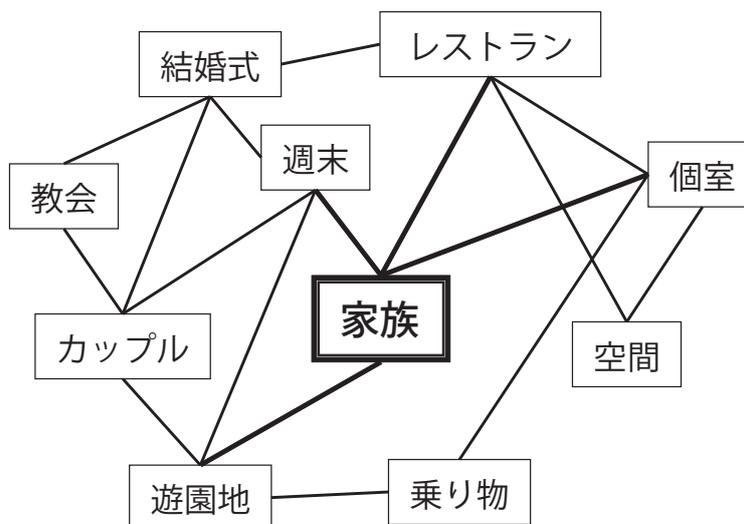


図2 意味関係の関連付けの視覚化

Fig.2 Visualization of Semantic Relations.

次に、意味関係の重要度の定義と数値化手法について説明する。本研究においては、形式化した意味関係の重要度を「各々の意味関係が、対象とするシステムにおいてどれほど特徴的であるか」を示す指針として定義する。この定義に基づき、すべての意味関係について次に示す「特徴度合い」を演算して、各意味関係の重要度とする。

【特徴度合い】

意味関係 (係り受け関係) の評価値

上記における評価値とは、以下に示す2種類の範囲における意味関係の出現頻度 (F_1 と F_2) を用いて計算した値である。

【出現頻度 F_1 】

テキスト情報全体においてある特定のデータが出現する回数

【出現頻度 F_2 】

特定のテキスト情報においてある特定のデータが出現する回数

いま、知識データベースに用いる N 個のテキスト情報を $S_n (n = 1, \dots, N)$ とし、項目 (2)(3) を経て得られた M 個の意味関係を $R_m (m = 1, \dots, M)$ とする。また、あるデータ Z に対する出現頻度 F_1 を $F_1(Z, S)$ 、出現頻度 F_2 を $F_2(Z, n)$ (ただし、 S はテキスト情報全体、 n は特定のテキストを指すインデックス) とする。このとき、 m 番目の意味関係 R_m の特徴度合い X_m を、 S と出現頻度 F_1 、 F_2 と単語の重要度を数値化する際に用いられる TF-IDF^{付録} 法を用いて以下の式で定義する。

$$X_m = \sum_{n=1}^N F_2(R_m, n) \log\left(\frac{N}{F_1(R_m, S)}\right) \quad (\text{式 1})$$

本項目では、式 1 で求める特徴度合いを対象とするシステムにおける意味関係 (係り受け関係) の重要度を示す値として定義した。一方、意味関係を構成する意味定義 (単語) 群に着目した場合、特徴度合いは同じく「意味定義の関連度合い」を表すと考えることができる。

最後に、以上で求められる重要度を、表 1 に示す形式でデータベースに格納し、知識データベースの構築を完了する。

3.3 知識データベースの活用と関連単語出力

本項では 3.2 項で構築した知識データベースを利用して、入力単語群に対する関連単語を出力する手順を説明する (図 1 ①②)。ここで、関連単語とは、入力単語と 3.2 項で示した意味関係にある単語 (意味定義)

表 1 意味表現による知識データベース

Table.1 Semantic Representation Architecture in Knowledge Database.

意味関係	重要度
家族-レストラン	20.5
家族-個室	10.1
教会-カップル	9.4
カップル-遊園地	14.6

対のことを指す。次に、図 3 のフローチャートに沿って各手順を説明する。

(関連単語抽出手順)

手順 1 入力単語群から単語をひとつ選択する。

手順 2 知識データベース内に同一の単語が存在するか否かを検索する。

手順 3 同一の単語が存在した場合、それに関連付けられている他の単語と関連度合いそれぞれ関連単語、重要度として取得する。

手順 4 すべての入力単語に対して手順 1 ~ 3 を繰り返す。

手順 5 手順 3 で得られるすべての関連単語と重要度をまとめて、関連単語群とする。

ここで、3.2 項 (4) の例で構築した知識データベー

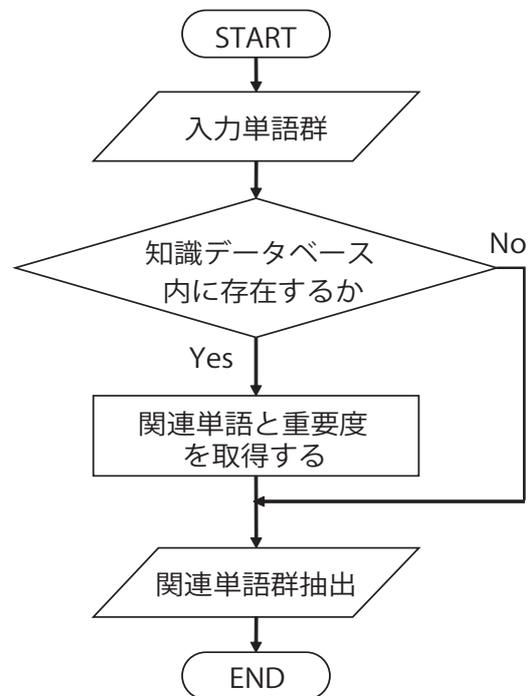


図 3 関連単語抽出アルゴリズム

Fig.3 Algorithm for Relative Keywords Extraction

スと入力単語群「家族」「カップル」を例にとると、上記手順は以下のように具体化される。

手順1 単語「家族」を選択する。

手順2 知識データベース内に単語「家族」が存在する。

手順3 単語「家族」に関連付けられている関連単語、重要度として、「レストラン:A」「個室:B」「週末:C」「遊園地:D」を取得する。

手順4 同様に、単語「カップル」から「教会:E」「結婚式:F」「週末:C」「遊園地:D」をそれぞれ関連単語、重要度として取得する。

本例では、関連単語群として「レストラン:A」「個室:B」「週末:C」「遊園地:D」「教会:E」「結婚式:F」が得られる。なお、各関連単語はそれぞれ、入力単語群との意味関係における重要度(A~F)を保持している。

3.4 機器情報検索と出力順序の決定

本項では、3.3項で抽出した関連単語群と重要度を用いて機器情報を検索し、出力順序を決定する手順について説明する(図1③④)。本手順では、図1の機器情報データベース(D2)を検索対象とする。D2の機器情報データベースは、情報検索の対象となる様々な機器情報を格納しているものとする。ここで言う機器情報とは、対象機器に関する情報であって、かつテキスト情報として扱えるものであれば種類は問わない。例えば、番組情報であれば番組詳細内容、グルメスポットであれば、店の紹介文といったものである。

次に具体的な手順を示す。ここでは、機器情報データベースに格納されているN個の機器情報データを $D_n(n=1, \dots, N)$ とし、3.3項より得られるM個の関連単語を $R_m(m=1, \dots, M)$ 、その重要度を $X_m(m=1, \dots, M)$ とする。

(情報検索手順)

手順1 関連単語群を取得する。

手順2 機器情報データ D_n に関連単語 R_m が含まれていれば、機器情報データD全体における評価値を3.2項で定義した出現頻度 F_1, F_2 を用いた式2で計算する。

$$\sum_{n=1}^N F_2(R_m, n) \log\left(\frac{N}{F_1(R_m, D)}\right) * X_m \quad (\text{式2})$$

手順3 機器情報データ群を手順2で計算した評価値の高い順に並び替える。

手順4 機器情報データ群が結果として出力される。

このように、対象とする機器に本手順を適用すれば、入力単語数が少ない場合でも知識データベースから得られた関連単語群が検索結果に反映されるため、結果として関連情報を含む幅広い機器情報を適切な順序で出力することができる。

4. 応用例

本章では、知識データベースを用いた情報検索手法をカーナビゲーションシステムの機能である「ドライブプランナー」に適用した例について報告する。

4.1 ドライブプランナーとは

ドライブプランナーとは、ユーザが指定した所定の条件に合致した観光スポット情報を推薦し、「ドライブプラン」として旅行行程表に近い形で提示する機能である。同機能はその他にも多くの特徴を備えているが本稿での説明は割愛する。詳しくは文献(1)(2)(3)を参照されたい。

4.2 知識データベースの応用

ドライブプランナーでは、指定された出発地や目的地、同行者、季節、観光目的、同機能の利用頻度、走行状況など多数のパラメータから各観光スポット情報の評価値(以下、おすすめ度)を算出し、それに応じて複数の観光スポット情報の検索と表示順序の調整を行う。

本研究の知識データベースを用いた情報検索手法は、上記おすすめ度の計算手法の一つとして用いられた。同手法ではまず、全国の観光スポット情報を知識データベース化し、それをもとに日本各地の観光エリア、同行者、観光目的(見る、食べる、遊ぶ、買う)を反映した各観光スポット情報のおすすめ度の算出と、表示順の並び替えを行う。

ところで、本適用例で対象としたのはカーナビゲーションシステムという民生機器であり、「観光スポット推薦」という特有の機能である。ここで、3.2項の(1)で述べた民生機器特有の留意点に沿って施した2点の工夫事項について解説する。

① データベースの範囲を限定する。

観光スポットには、文化的な特色や特産物、名物など地域ごとに異なる特色が見られる。そこで本適用例では、観光スポット推薦に用いる知識データベースをドライブプランナーのエリア情報(図4)にしたがって個別に構築することとした。これにより、各地域の特色が他の地域に影響されることなく各データベース内の特徴度合いとして保持されるため、各観光スポッ

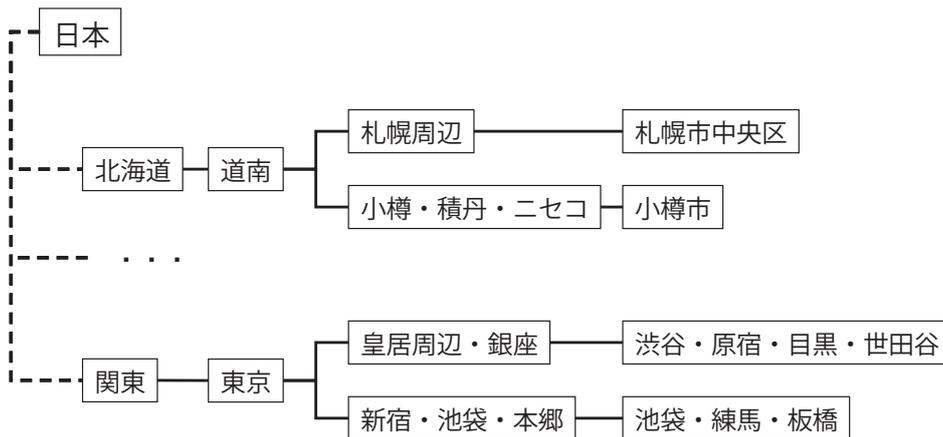


図4 ドライブプランナーエリア情報

Fig.4 Hierarchy of Segmentation Areas in 'Drive Planner'

ト情報のおすすめ度の算出においても適切な出力が可能となった。

② 一般用語を追加し、知識データベースを拡張する。

知識データベースを用いた観光スポット推薦では、同行者や観光目的を定義する単語群(家族, カップル, 友達, 見る, 遊ぶ, 食べる, 買う)を知識データベースに入力し、得られた出力(関連単語群)を用いて該当する観光スポット情報を検索する。しかしながら、本処理では、入力する各単語が知識データベースの保持するいずれかの意味定義(単語)に合致することを前提としている。そのため、知識データベースの構築に用いたテキスト情報(すべての観光スポット情報)や、検索対象とする観光スポット情報に上記の単語群が含まれていない場合には、同行者や観光目的を反映した観光スポット情報の推薦が不可能となる。

例えば、「家族」と「食べる」を入力単語とし、観光スポット情報として以下に示す「レストラン1」と「レストラン2」が存在するものとする。

* レストラン1

ファミリーでも利用しやすいカジュアルな雰囲気のお店。

* レストラン2

家族連れでもくつろげるゆったり空間。

ここで、「家族」という単語を含む「レストラン2」は推薦結果として提示されるが、それを含まない「レストラン1」は検索結果に該当しないため、提示されない。この現象は、知識データベース構築に用いたテキスト情報からは「家族-ファミリー」という意味定義と意味関係が得られなかったために起こる。そこで本適用例では、

知識データベースの構築の際に、既存の観光スポット情報に加えて「家族-ファミリー」に代表される同行者や観光目的に対する連想語句を付加した。その結果、本例の場合にも「家族」に対して「レストラン1」の情報を出力することが可能となった。

以上のように、民生機器における知識データベースは、対象とする情報の範囲を分割あるいは限定することや、機器特有の用語に対して一般用語を追加し、意味定義と意味関係を拡張することによって所定の目的に特化した活用が可能となる。

4.3 推薦結果の一例

本項では、実際の処理手順にしたがった観光スポット情報の推薦例について説明する。

- ① 同行者を「家族」に設定する。
- ② 目的地エリアを設定する。
- ③ 観光スポット情報を推薦、提示する(「食べる」ジャンルについてのみ提示)

実際のシステムでは、まず①で入力された「家族」を保持し、次に②で指定された対象エリアに該当する知識データベースを選択する。その後、キーワード「家族」を入力として知識データベースを駆動し、関連単語群を出力する。以下に出力された単語群を示す。

「家族」「ファミリー」「食べる」「ランチ」「和食」「くつろげる」「楽しめる」「雰囲気」「有機野菜」...

以上の関連単語群を用いて③の処理を行った結果、例えば以下の観光スポット情報が推薦される。

* 洋食レストラン

有機野菜を使ったランチがおすすめ。家族みんなで楽しめます。

* 和食処〇〇

家族連れでも安心してくつろげる。

* イタリア料理△〇

グループやファミリーなら個室でゆっくりと。

このように、店の意図する紹介文から知識データベースを構築し、活用することで、指定条件である同行者、観光目的に合った内容の観光スポット情報(レストラン)を推薦することができた。

5. まとめ

本稿では、民生機器において利用可能な知識データベースの構築方法と、それをを用いた新たな情報検索手法について解説した。また、本手法をカーナビゲーションシステムの「ドライブプランナー」機能に適用し、指定条件(同行者、観光目的)を満たす観光スポット情報の推薦が可能であることを確認した。

本研究では、まずマルチメディア情報の一つであるテキストに着目した。一方、昨今の民生機器は、音楽や画像、映像といったエンターテインメント性が高く、また高度な解析を必要とする情報を広く扱うようになったため、テキスト情報のみの対応では不十分である。今後は、これまで取り組んできたテキスト情報に加え、他の様々なマルチメディア情報を用いた知識データベースの構築方法について研究を進めていきたい。

(注1):パイオニア製カーナビゲーションシステム「サイバーナビ」(AVIC-VH009 など)に搭載

付録: TF-IDF アルゴリズム

テキスト情報の特徴度合いを数値化する方法の一つとして、TF-IDF 法がある。

TF (Term Frequency) とは、ある文書における単語の出現頻度 (tf) である。一方、IDF (Inverse Document Frequency) は、全文書数 (N) に対してある単語を含む文書数 (df) の逆数の対数である。ここで TF-IDF は両者の積を取ることで得られる数値であり、以下の式で表される。

$$TF-IDF = tf \times \log\left(\frac{N}{df}\right)$$

特定の文書中に高頻度で出現する単語、すなわち tf 値が高い単語は同文書において特徴的と言える。一方、多くの文書で多数出現する単語は概ね一般用語である場合が多いため、 df 値が高い単語は全体として特

徴的ではないと考えられる。よって本アルゴリズムでは、TF-IDF の値として tf 値と df 値の逆数を取った値の積が高い単語ほど重要単語であるものとする。つまり、ある単語が特定の文書において特徴的であるほど TF-IDF は大きくなり、汎用的であるほど TF-IDF 値は小さくなる。

参考文献

- (1) 市原, 杉江, 山崎, 井上, 塩田: カーナビゲーションにおけるドライブプラン作成システムの開発, 情報処理学会第 66 回全国大会論文集, 5-225 ~ 228, 2004-03
- (2) 市原, 杉江, 山崎, 井上, 塩田: カーナビゲーションにおけるドライブプラン作成および行程管理システム, 自動車技術会春期学術講演会前刷集 No74-04, p17-22, 2004-05
- (3) 市原, 杉江, 山崎, 井上, 塩田: ドライブプランナーの開発, PIONEER R&D Vol.14, No.3, p.1 ~ 10 (2004)

筆者紹介

梶 雅代 (かじ まさよ)

技術開発本部 総合研究所 情報メディア技術研究部に所属。主な経歴は、エージェント技術の研究開発、自然言語処理技術の研究に従事し、現在にいたる。マルチメディア関連の人工知能応用に興味を持つ。

堀内 直明 (ほりうち なおあき)

技術開発本部 総合研究所 情報メディア技術研究部に所属。主な経歴は、エージェント技術の研究開発に従事し、現在にいたる。人工知能とヒューマンマシンインターフェースの研究に興味を持つ。

莪山 真一 (がやま しんいち)

技術開発本部 総合研究所 情報メディア技術研究部に所属。主な経歴は、アクティブノイズキャンセラの開発、デジタル無線変復調技術の開発、エージェント技術、音楽構造化技術の研究を経て、現在に至る。ソフトウェアによるアルゴリズム研究とアプリケーション開発。特に、マルチメディア情報の特徴抽出と検索技術に興味を持つ。