

3T-10

イエローページにおける 業種分類推論方式

別所克人 岩瀬成人 戸部美春

NTT情報通信研究所

1. はじめに

店や企業等の情報を検索・案内するイエローページサービスにおいては、現在ツリー状に構成された業種分類を、ユーザーがメニュー形式で検索していく方式をとっている。しかしこの方式では、ユーザーの思っている業種名と、システムが用意している業種名とが一致しないことが多く、また業種数が多いため、検索行為が繁雑になるという問題があった。そこで、ユーザーが、通常の日本語文を入力するだけで、該当する業種を推論し、店・企業等の情報を案内する業種分類推論方式を研究している。

本方式では、入力文から自然言語処理を用いて商品・動詞・職業に相当するキーワードを抽出し、関連知識（業種＝商品＋動詞）を用いて、抽出したキーワードから業種を推論している（文献[1]）。

しかし、ユーザが入力する商品、職業の表現は複合語として表現される等多種多様なため、全ての表現パターンの商品、職業をシソーラスとして持つことは困難である。本稿では、商品、職業の多様な事例に対応可能な推論方式を提案する。

2. 業種推論の概要

図1に、業種推論システムの構成を示す。

推論システムは以下の4つのモジュールからなる。

- 1) 形態素解析：単語辞書を使って、入力文を単語単位に分割する。
- 2) 意味解析：単語間の係受け関係を調べ、入力文の意味構造を明らかにする。
- 3) 意図理解：ユーザーの意図を表現する単語群をキーワードとして抽出する。
- 4) 分野推論：シソーラス等の知識を用い、ユーザーの意図に沿った業種を推論する。

3)の意図理解においては、入力文からユーザーの意図として、動詞とその対象としての商品、そして職業をキーワードとして抽出する。例えば、「ひ

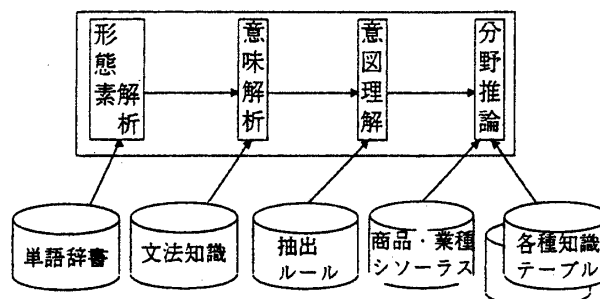


図1 業種推論システムのモジュール構成

らめの食べられるレストラン」では、商品キーワード「ひらめ」、動詞キーワード「食べる」、職業キーワード「レストラン」が抽出される。

4)の分野推論は次のように行う。

システムには、あらかじめ各業種を商品と動詞との組み合わせで表現した関連知識を定義しておく。抽出した商品キーワードで、商品シソーラスを検索する。検索した商品に関連があれば、その関連で対応づけられた業種を推論する。関連がなければさらに商品シソーラスを用いて上位商品に変換し、業種の推論を再実行する。

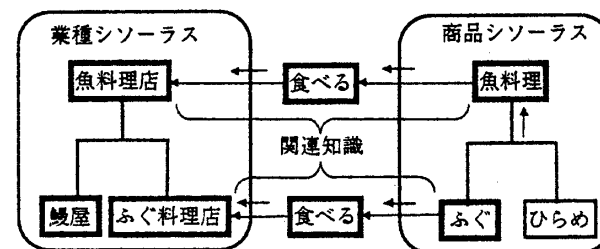


図2 分野推論の流れ

3. 複合語のキーワード処理

商品キーワードや職業キーワードは複合語の場合がある。複合語のキーワードを全てシソーラスに登録することは、複合語を構成する単語の組み合わせの数が膨大になるので非効率である。そこで、シソーラスの方には複合語でない商品のみを登録し、キーワードから適切なシソーラス中商品を導く曖昧処理が重要となってくる。

3.1 複合語の商品キーワード処理

複合語の商品キーワードには、主に次のような場合がある。

An Inference System of Business Category
for a Yellow Page Service

Katsuji Bessho, Shigehito Iwase, Miharu Tobe
NTT Information and Communication Systems
Laboratories

3-9-11 Midoricho, Musasino, Tokyo 180, Japan

1. 末尾語に重要な意味の単語がくる。
例：「フランス料理」
2. 末尾語が語尾相当の単語（主要な意味がない単語）である。例：「競馬情報」
3. 末尾語がサ変名詞である。例：「競馬予想」
4. 末尾語が動詞としての意味をもつ。
例：「ピアノの先生」（「先生」は「教える」に変換出来る）

そこで、図3のように、末尾語が語尾相当の単語なら削除し（2の対処）、末尾語が、意味的に対応する名詞と動詞のリストをもつ動詞変換テーブルに登録されているかサ変名詞である場合は動詞化を行う（3、4の対処）。それでも検索出来なかった場合は、一旦削除した末尾語を全てつけた上で前方の単語を削除して、同様に処理を続けていく（1の対処）。

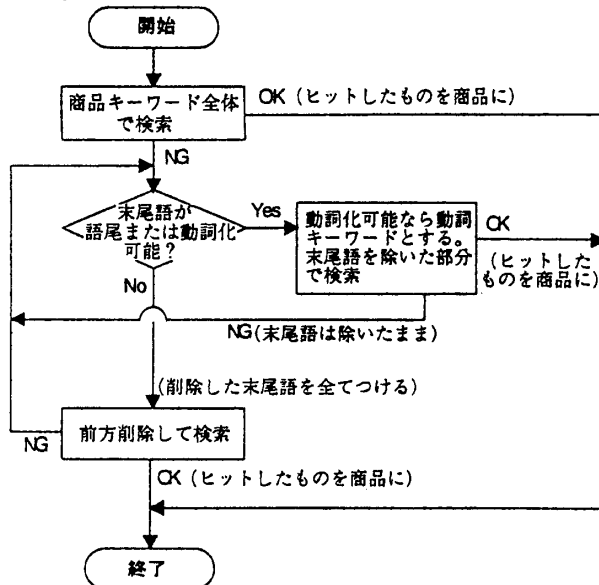


図3 商品キーワードからのシソーラス検索

3.2 複合語の職業キーワード処理

複合語の職業キーワードには、主に次のような場合がある。

1. 末尾語に重要な意味の単語がくる。
例：「温泉旅館」
2. 商品+職業語尾 例：「音楽教室」
3. 商品+サ変名詞+語尾 例：「CDレンタル店」
2や3の職業語尾やサ変名詞は、「教室→教える」、「レンタル→レンタルする」の様に動詞キーワードとして使用できることに注意する必要がある。

1の場合は、商品と同様に前方削除によりシソーラスを検索する。2、3の場合は、職業キーワードを商品部と動詞部に分離し、商品部は商品シソーラスを検索する。動詞部はサ変動詞化または、動詞変換テーブルにより動詞化して、動詞シソーラスを検索

する。但し、動詞と商品種別（物、事、人等）との不整合がある場合があるので、商品の最適変換を行うことにより、動詞に対応する商品を求め、推論を進める。商品最適変換では図4のように、ある商品がある種別に変換すると何の商品になるのかを記述した事象変換テーブルも用いる。

例：職業キーワード「調理師/専門学校」
 職業中商品：調理師
 職業中動詞：教える（専門学校→教える）
 商品最適変換：調理師（人）→調理（事）
 関連検索：調理+教える→料理学校

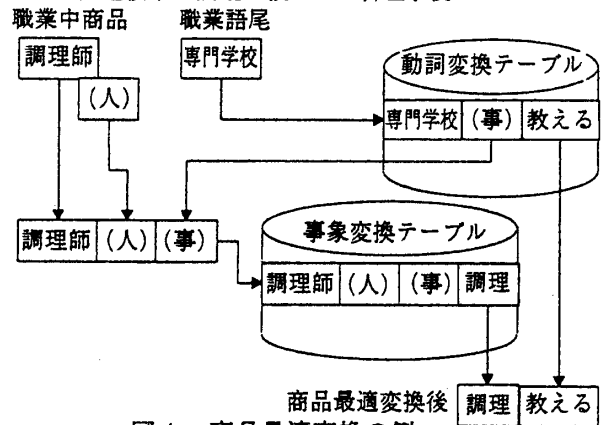


図4 商品最適変換の例

3.3 複数キーワードによる推論

入力文中に商品と職業あるいは商品と動詞と職業が全て含まれていた場合、推論に使用する商品候補は商品キーワードと職業キーワード中の商品がある。また、動詞候補は入力動詞と商品・職業キーワードから抽出される動詞とがある。このように複数の選択候補がある場合、最適な推論を行うため、適切な商品及び動詞の選択が必要になる。

例えば、入力文「布団をあつかっているレンタルショップ」を考える。関連知識として「布団+あうかう→寝具店」「布団+レンタルする→貸しふとん」があるが、推論において入力動詞よりも職業中の動詞との関連を優先させることにより、適切な業種「貸しふとん」を推論する。

4. おわりに

本稿では、商品と動詞から業種を推論する方式をベースとした時に、複数種類のキーワードや複合語としてのキーワードの曖昧処理方式について述べた。今後本方式による推論性能の評価を進めていく予定である。

参考文献

- [1]岩瀬, 大山：「自然言語処理技術を用いた職業別電話帳検索の高度化」, 電子情報通信学会論文誌, Vol. J74-D-II No.9 pp.1255-1263 (1991)