

# 対象の特徴から名前を特定する手法の提案

田中貴浩 田添丈博  
鈴鹿工業高等専門学校

椎野努  
愛知工業大学

## 1. 背景と目的

本研究は対象の特徴から対象の名前を特定する手法を提案するものである。現在の検索システムは入力として対象の名前を要求するため、名前がわからない・思い出せない場合に利用するのが困難である。そこで、いくつかの特徴を示せば対象の名前を知ることができるシステムを構築し、検索システムへの入力インターフェースとしての利用を想定している。

名詞を修飾する形容詞はその名詞の特徴を表していることから、形容詞を「対象の特徴」として用いる。このように形容詞とそれに係る名詞の研究は過去にも扱われている[1]。

「(形容詞) + (名詞)」の組み合わせからデータベースを構築し、名前の特定に使用する。この収集データとなる「(形容詞) + (名詞)」の組み合わせを「セット」として定義する。例えば「リンゴ」に形容詞「赤い」が付随しているとき「赤いリンゴ」というセットが収集される。

データベースは複数のセットから構成され、形容詞に符合するかどうかという質問を用いて、データベース内の名詞を絞り込み、最終的に名詞をひとつに特定する。

## 2. 名前を特定する手法

名前の特定はユーザとの質問形式によって行われる。まず、ユーザに対象の特徴を想定してもらう。そして、形容詞に符合するかという質問を行い、その回答に応じたデータ処理によってデータベースの名詞を絞り込む。この質問はデータベース内の最も多い形容詞から選ばれる。それを繰り返し、様々な形容詞に対する回答から最終的にひとつの名詞を特定する。フローチャートを図1に示す。

データベースのすべての名詞に対し、回答により増減する初期値0の評価値を名詞候補に設定する。対象の特徴が質問形容詞に符合するならば評価値を増加、符合しないなら減少させる。また、重複したセットに関しては、それらのセット数について評価値の増減を行う。例えばデータベースにデータ「赤い+リンゴ」が10セットある場合、「赤い」かどうかという質問に対し「NO」と回答した時に「リンゴ」の評価値は10減少する。

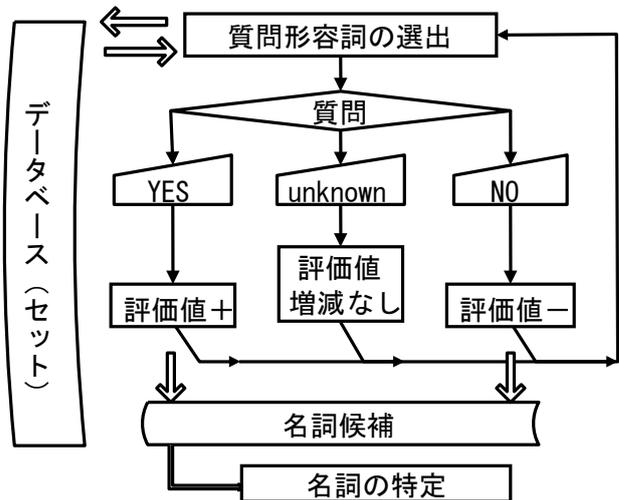


図1：名詞を特定するフローチャート

また、質問の意味がわからない、判断できない場合には質問を飛ばすことができ、評価値を増減させない。

規定回数の質問に回答した後、名詞候補内で評価値の一番高い名詞を想定された対象の名前と判断し出力する。

## 3. 実験と考察

実験ではデータベースの名詞を限定し、以下の10種を用いた。

犬	金魚	机	鉄	トンボ
パン	ヒマワリ	リンゴ	冷蔵庫	水

各名詞を含むセットをWWWから収集する。検索サイトを用いて「い+(名詞)」という形で検索し、検索結果の中からセットとなる組み合わせを手作業で探すという方法を採用した。

収集するセット数は上記の各名詞に対して20セット、計200セットを収集し、それをデータベースとした。

実験は10種の名詞をそれぞれ想定して、20題の質問を行い、ユーザとして回答した。ただし、回答「わからない」は質問数に含めないものとする。実験結果を以下に示す。

実験の結果、名詞6種が正解となった。表1に結果を示す。

表1：実験結果

犬	金魚	机	鉄	トンボ
誤	誤	誤	正	正
パン	ヒマワリ	リンゴ	冷蔵庫	水
正	正	正	誤	正

正解例として「リンゴ」の特定結果を表2に示す。「リンゴ」は他の名詞と比べ評価値が高く、回答による名詞特定が動作していると分かる。

表2：「リンゴ」の実験結果

質問	回答	質問	回答	名詞	評価値
おいしい	○	すっぱい	○	リンゴ	19
ながい	×	たかい	×	パン	6
甘い	○	ひよろい	×	水	2
長い	×	まあるい	○	金魚	-1
赤い	○	碧い	×	犬	-3
あまい	○	高い	×	机	-4
かしこい	×	あかい	○	ヒマワリ	-7
重い	×	あったかい	×	トンボ	-8
きたない	×	うすい	×	冷蔵庫	-8
細い	×	かたい	○	鉄	-10

次に、不正解だった名詞の検証を行う。4名詞が想定された名詞とは違う結果となった。

「犬」「金魚」「机」の3名詞は評価値が第2位であったが、「冷蔵庫」は第4位となった。「冷蔵庫」の特定結果を表3に示す。「冷蔵庫」の特定ではデータベースにセット「かしこい+冷蔵庫」が7つ含まれていたために「かしこい」の回答「NO」に対して評価値が7減少し、不正解となった。これは「かしこい冷蔵庫の使い方」を「かしこい+冷蔵庫」として収集したことが原因と分かった。

このような事態を避けるためには収集時の制限、もしくは正しく収集できる方法が必要となる。そこで、不正解の原因である「(形容詞)+(名詞)の(名詞)」という形に対して、それを除外または、「かしこい+使い方」として収集することで「冷蔵庫」の特定ができると考える。

表3：「冷蔵庫」の実験結果

質問	回答	質問	回答	名詞	評価値
デカイ	○	細い	×	ヒマワリ	8
おいしい	×	ごっつい	○	鉄	7
ながい	×	すっぱい	×	金魚	0
甘い	×	たかい	○	冷蔵庫	-1
長い	×	ひよろい	×	犬	-1
赤い	×	まあるい	×	机	-3
あまい	×	碧い	×	トンボ	-7
かしこい	×	かわゆい	×	水	-15
重い	○	高い	○	パン	-15
きたない	×	冷たい	○	リンゴ	-19

#### 4. データ収集法の改良

データベース構築のためのセット収集の際、「(形容詞)+(名詞)の(名詞)」という形を収集しないものとする。収集時「かしこい冷蔵庫の使い方、ほっそい金魚のフン、まあるいリンゴの飴、碧いトンボのピアス」などが収集されなかった。

改良されたデータベースを用いて、名詞10種、合計セット数200、質問20題によって実験を行った結果、名詞8種が正解となった。表4に結果を示す。

表4：実験結果(改良後)

犬	金魚	机	鉄	トンボ
正	正	正	正	正
パン	ヒマワリ	リンゴ	冷蔵庫	水
正	誤	正	誤	正

前回の実験と比べ、「犬」「金魚」「机」が正解になり、不正解は「ヒマワリ」「冷蔵庫」となった。「冷蔵庫」は最終評価値「9」で第2位となった。「冷蔵庫」の特定結果を表5に示す。

表5：「冷蔵庫」の実験結果(改良後)

質問	回答	質問	回答	名詞	評価値
デカイ	○	うすい	×	ヒマワリ	10
ながい	×	ごっつい	○	冷蔵庫	9
甘い	×	すっぱい	×	犬	1
あまい	×	ひよろい	×	鉄	1
おいしい	×	黒い	×	トンボ	-4
赤い	×	細い	×	水	-5
長い	×	重い	○	机	-6
可愛い	×	青い	×	金魚	-6
たかい	○	冷たい	○	リンゴ	-17
高い	○	アカい	×	パン	-20

#### 5. まとめ

検索システムへの入力インタフェースとしての利用を想定している名前特定システムは、ユーザとの質問形式での回答によりひとつの名詞を特定するものである。システム実現のための限定実験を行い、正解率上昇のためにデータ収集法を改良することで名前特定に改善が見られた。今後は名詞の数を増やし実験を行っていく。データベースに存在しない名詞に対する処理に関する問題、増加により生じる問題の調査と解決を課題とする。

#### 参考文献

- [1] 阿辺川武, 奥村学. “形容詞を用いた対象・属性名詞対の収集および分析.” 言語処理学会第12回年次大会論文集, 2006.