

# ハッシングを使ったアナグラム検索システム

近畿大学大学院 生物理工学研究科  
電子システム情報工学専攻 知能情報工学研究室  
中野 尚則

## 1. はじめに

ことわざ等で1部の単語を置き換えて、一種のアナグラムと称して利用されています。例えば「背中に水」は「寝耳に水」のアナグラムとなる。

アナグラムから元の諺を検索するシステムを設計試作した。諺を文節に分解し、その各文節をハッシュ鍵(KEY)とするハッシュ表を設計製作し、そのハッシュ表から元の諺全文が検索できるように設計した。現在ハッシュ表には190の諺が573個の文節に分解され記憶されている。ハッシュ表の大きさは文節の数の3倍を目安として設計した。

同じ文節が異なる諺で使われる事もあるので諺の何番目の文節であるかもハッシュ表に記憶することで、諺を絞り込むことを可能にした。

## 2. 文節検索

### 2.1. これまでの検索システム

これまでのデータ検索システムは漢字などを含めた文字数や読みの文字数などのヒット数から近い文章を出力させていました。しかし、その文字数の違いが大きければ検索することはできませんでした。

### 2.2. アナグラムと文節

最近では、ことわざをその場の雰囲気に応じて置き換えたアナグラムが利用される

ことがある。

しかし、最近のデータ検索システムではことわざを一つの文として記憶しており、そのことわざとアナグラムの変化によっては正確に検索できないという場合が発生します。

元のことわざとアナグラムの間には助詞、助動詞や形容詞等語尾が変化していないという特徴がある。本来、助詞、助動詞や形容詞等語尾変化をする単語はその語尾変化によって意味や雰囲気が大きく異なってしまう、アナグラムとして聞こえなくなってしまう場合があるからである。

そこで、今回のことわざ検索システムではことわざを文節毎に分けて助詞、助動詞、形容詞等の語尾変化も含めて文節としています。

各文節をハッシュ鍵とするオープンハッシュ表を作成し、文節と一致したことわざでその出現順位が一致したことわざを候補のことわざとする。

候補のことわざから助詞、助動詞、形容詞等を照合して、候補のことわざと異なる文節が最小のことわざ $k$ を出力する。例えば「背中に水」

このようにして助詞、助動詞、形容詞等語尾だけを変化させたアナグラムも語尾変化させた文節だけが異なるアナグラムの元のことわざとして検索できる。また、元のことわざとの文字の違いが大きいアナグラムであってもアナグラムから元のことわざ

を検索することができる。

### 3. データ検索の工夫

ことわざのアナグラム検索システムを試作するにあたって文節ごとに分割して検索を行う、文節検索を実現した。実際には諺のデータベースを入力するときに、人が文節に区切って入力することで、機械的に文章から文節に区切る困難さを避けている。文節に区切ることによって助詞、助動詞、形容詞等語尾変化をする単語をその語尾変化のまま検索語の一部として利用することができる。

#### 3.1. ハッシングによる検索

文節をハッシュ鍵とするオープンハッシュの表を図1のように設計した。オープンハッシュの表の長さをn、各文節をsとする。各文節sから自然数g(s)を計算し、 $h(s) = \text{mod}_n(g(s))$ とすることでハッシュしている。ハッシュ表のインデックスh(s)に文節sを含むことわざのことわざ表に記憶されているインデックスがリンクドリストで記憶されている。データ検索アルゴリズムの高速化を実現した。

データ検索アルゴリズムにデータ構造ハッシングを利用しました。

データ構造ハッシングは、あらかじめN個(N:任意の数)の配列を準備しておき、記憶する文字列の数値をNで割った余りの数値をPとしたとき、P番目の配列に格納するというアルゴリズムです。

そうすることで、データ検索スピードを向上することができます。

### 4. おわりに

アナグラムから元のことわざを検索でき

ることによって、元々、ことわざに興味のなかった人でもことわざに興味を持ってもらえるのではないかと思います。

このシステムでは検索語が正確に入力されていることを前提としたシステムとなっています。

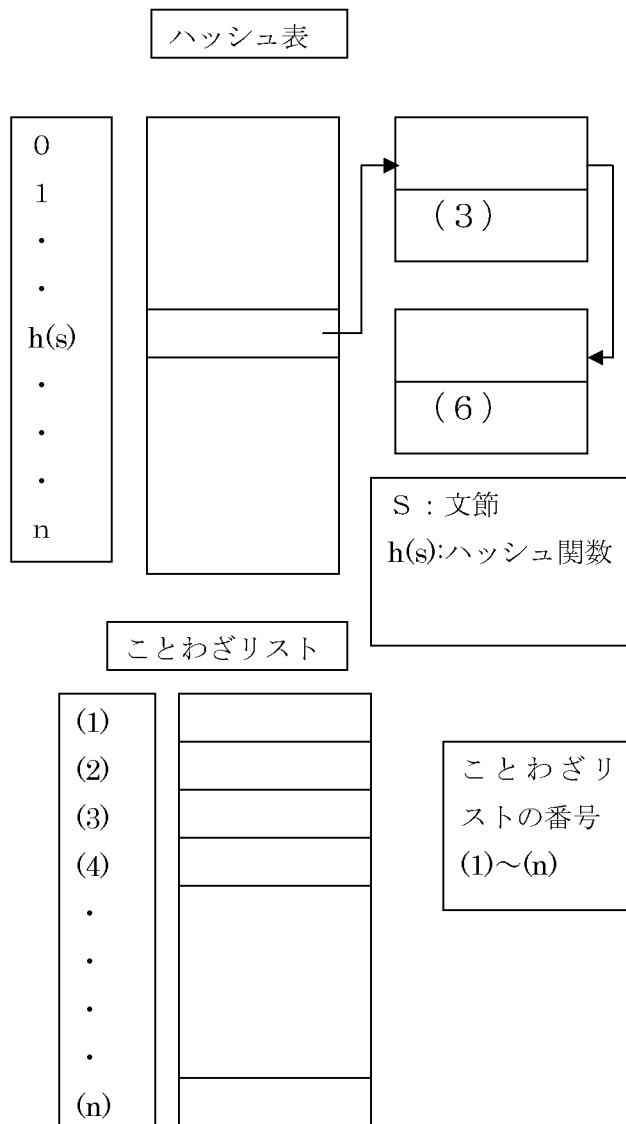


図1 ハッシュ表とそのデータ構造

#### 参考文献

[1]野崎,野下訳「アルゴリズムの設計と解析 I, II」サイエンス社 1977,pp100