研究テーマ：WWWを使った自然言語による質問応答システム

名列番号　０５１　　広谷　健一

１．まえがき

現在普及しているWWWの検索システムは単語をキーワードとしている。今回の研究ではWWW検索を行った結果から、入力された自然言語[1]の質問に基づく解答を得るシステムを考えてみた。

２．研究の意義

この研究は自然言語によるWWW検索のシミュレートでもある。それが可能になれば、検索精度が上昇するほか、音声認識と併用することで普段コンピュータを扱わない人でも検索が容易になることが考えられる。

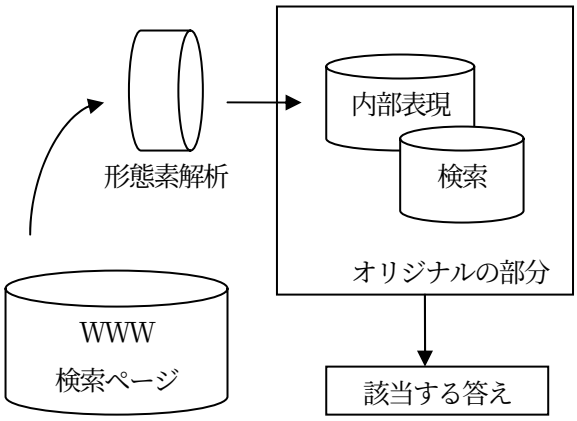
３．研究課題

この研究で必要なのは自然言語解析[2]と知識である。近年、WWWは莫大な情報量となり、これを答えの元になる知識として利用する。自然言語解析にはその中のひとつである形態素解析[3]という手段をとる。

４．研究方法

WWWの情報収集には一般の検索ページを利用し、形態素解析には専用のツール「茶筌」を使用する。この二つをベースとし、自作のプログラムをその上に階層的に連係させるシステムを考える。

形態素を考慮に入れて解答となる部分を検索するため、自然文からは直接検索を行わず、いったん形態素とその構成を記述した内部形式に変換する。そこから解答に該当する部分を抜き出す。



図．処理のイメージ

５．シミュレート

部分的なプログラム（上の図において自然文から内部形式に変換するもの, 内部形式から解答を検索するもの）を作って検証した結果、“ 該当する答え ” は検索された。

一部、質問のキーワードに一致するが質問の内容に沿わない解答も検索されてしまう。

６．まとめと今後の課題

今回の研究では自然言語解析に形態素解析を用いており、キーワードだけでなく形態素単位で検索できるようになったが、文の構成が異なっていても同じ意味の文などは検索されない。これを可能にするには自然言語解析の深い部分に踏み込まなくてはならない。係り受け解析[4]を取り入れるだけでも検索の精度は高まる。

また、自然言語解析が十分にできれば、“ 要約 ” が可能になり、複数の同じ意味の文を少数の文にまとめることができる。

---

[1]：自然言語とは人が読み書きする言葉。

[2]：自然言語解析には大きく分けて４段階あり、形態素解析, 係り受け解析, 意味解析, 文脈解析の順に流れる。

[3]：形態素解析とはプログラム言語でいうところの字句解析に相当する。つまり自然言語文から単語を切り出し、その品詞を同定する作業である。

[4]：形態素解析の次の段階であり、形態素同士の文法的な関係を同定する。（主語と述語の関係，副詞と述語の関係など）