

アイデアインテグレーション環境

4W-6

佐藤 研治、村木 一至
NEC 情報メディア研究所

1. はじめに

人が計算機を用いてアイデアを発想し、整理し、具体化していくという機会が増えてきている。計算機は当初は数値計算をするものとして、その後は情報を保存/検索するものとして認識され、更に推論や高度の知能を持つものになると期待されたがそうはなっていない。現在は、あくまで計算機は情報保存/検索装置であり、人間にしかできない高度な判断と計算機の高速度な情報処理能力を組み合わせた知識処理への要求が高まっている。筆者らは、計算機内に記録された情報を活用するために、「文章作成」という場面においてテキスト資産を容易に再利用させるエピソードのインデクス化手法等の提案[1, 2]を行ってきた。この手法を用いることで、ユーザの既知情報への到達はより容易になったが、アイデアを発想するという場面では、ユーザを未知情報へ遭遇させることが必要となる。

本稿で提案する環境は、アイデアの発想場面の1つとして、文章作成場面を支援対象とする。また、未知情報として最新情報が望まれることも多いと考え、プレーンなテキストが蓄積されている情報源を対象とする。文章作成場面のまさに文章を入力しつつある状況で、入力直後の文字列を現在思考中のアイデアの関連情報として捉える。そして、関連を持つ未知情報を自動検索し提示する環境となっている。この環境によりユーザが明示的な検索要求を示していない場合でも、未知情報に遭遇することが可能で、その後の明示的な検索要求やアイデアへの発展が期待できることになる。

2. アイデアインテグレーション環境

論文や特許を書く際に、通常まえもって先行技術のサーチを行ない、その技術を押さえた上で、自らの文章作成に取りかかる。しかしこのサーチは、頭の中の技術をキーワードとして表現し、そのキーワードを用いて検索を行なう手法が現在一般的である。このため、これから書こうとしている文章と密接な関連があるが、キーワード的には異なる文献は検索されない。また、文章を書き進めると思考の整

Idea Integration System
Kenji Satoh, Kazunori Muraki
Information Technology Research Laboratories,
NEC Corp.

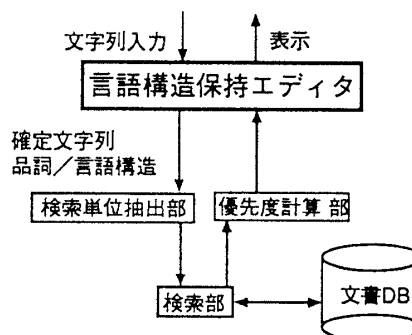


図1. システムの構成図

理が行なわれ、技術的により良いものへと変化するというのも誰しも経験することである。

もし文章を先に作成したと考えたと、その作成文章は実は検索キーワードの羅列であり、更にそのキーワード間の関連性まで文として指示されているものとなると考えられる。ここで、文章全体を書き終ってから検索するのではなく、入力直後の文字列を用いて検索すると以下のメリットがある。

1. 入力情報と検索文書の検索箇所の対応関係が明白
2. 入力情報に対する記憶が鮮明な内に情報を提供可能

この文書を入力しつつ検索が行なわれる環境を、1つのシステムとして実現した。このシステムの構成図を図1に示す。

2.1. 検索単位の抽出

文字列の入力に伴い、入力中の文から順次検索要素を取り出し検索を行なう際に、検索の要素として考えられるものは、「単語」「文節」「格要素」「節」「文」等がある。この中で、単語や文節ではあまりに多くの文が検索されると考えられ、また関連性の乏しい文が多くなってしまふ。また、逆に文同士では表層での全体の一致は滅多に起こらないと考えられる。日本語の基本的な文の組み立てでは、同一用言に対する格要素の組合せで節ができていたため、本システムでは基本的な検索の単位を「格要素」とし、同一用言での複数格の一致は、複数の検索結果が存在した場合の優先付け手段としている。

文章を入力している際に実時間で高精度の構文解析を行なうことは、現在の自然言語処理技術ではまだ不十分である。しかし、日本語では「格要素」が表層に「格助詞」として現れるため、これを中心にパターンの「名詞-格助詞-用言」の組を入力文字列からある程度正確に抽出可能である。また、もともとユーザに発想の糸口を与える情報提示が目的なので100%の再現率は必要としない。

この格の抽出パターンを用いて検索要素を抽出するためには、入力しつつある文字列から、品詞や語切り区間の把握を行なう必要がある。これについては、言語構造を保持するエディタの技術[3]を用いた。検索単位抽出部では、この言語構造保持エディタより、文節として確定した文字列とその品詞や単語/文節構造を受けとり、パターンの検索要素を取り出している。

実際の例として「自民党に離党届を提出した議員は、」という文章が入力された点で、「自民党-に-提出」「離党届-を-提出」の2つの検索要素が抽出される。

2.2. 検索情報の提示

検索部での検索処理は、情報源中の文で入力文から抽出された検索要素と同じ検索要素が抽出できる文を「ヒットした」とする。この際に、用言の活用形および名詞の不完全一致検索を行なう。また、文を検索の単位とすることで、同一検索要素で、同一文書の複数箇所が検索されることもあり得る。文を検索の単位とする理由は、以下の2点である。

1. 検索された箇所の特定/提示が容易
2. 情報源の持つ情報を誤りなく伝達可能

このようにして検索された文は、現在のシステムでは画面の許す限りの1文単位ウィンドウで表示される。この際に大量の文が検索される場合を想定し、優先度計算部で名詞の完全一致および動詞の活用形までの一致で順序付けを行なう。更に、検索された文の検索単位同士で一致するものをまとめ、検索単位ウィンドウに別に表示し、画面の制限で表示できない文をこのウィンドウより呼び出せるようにしている。また更に、優先度計算部では、同一用言で複数の検索単位にヒットした文を優先して画面に表示する。

前節で述べた例で、2つの検索単位が抽出された場合の情報の提示を、実際の文で説明する。「自民党に椎名素夫参院議員は二十三日午前、離党届を提出した」「武村正義代議士ら11人が離党届を

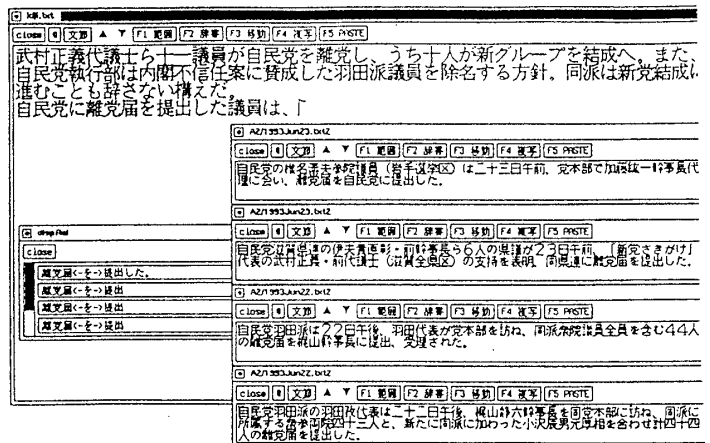


図2. システムの動作画面

提出し、」「羽田派衆議院議員全員を含む44人の離党届を梶山幹事長に提出した。」という3つの文が、情報源内に含まれていたとする。これらの文から抽出される関係として「自民党に提出」「離党届を提出」という要素が検索要素に一致し、これらの文を「ヒットした」とする。更に、「離党届を提出した」という活用形まで含めた一致を見て、1番目および3番目の文を2番目の文よりも優先し、さらに「自民党に提出」と「離党届を提出」の両方の関連要素の一致を見て、1番目の文を3番目の文よりも優先することを決定する。この情報提示が行われた様子の画面を図2に示す。

3. おわりに

本稿で提案した環境により、ユーザが明示的な検索要求を持たなかった場合でも、未知情報に遭遇することが可能で、その後の検索意図やアイデアの発展が期待できることになる。今後の課題としては、アイデアをまとめていく作業を計算機で支援するために、既に提案したテキスト資産を容易に再利用させるエピソードのインデックス化手法等と統合し、既知情報/未知情報の全てを活用していく環境を構築していく必要がある。

参考文献

- [1] 佐藤、村木「パーソナルテキストベース管理」、AI学会第6回全国大会 7-9,1992
- [2] 佐藤、村木「Detection of Relevant Facts without Retrieving」、AI学会第7回全国大会 11-5,1993
- [3] 野村、村木「言語の構造単位を保持した文書執筆支援システム」、情処自然言語処理研究会,82-10,1991