

要約支援システムCOGITO —マンマシンインタフェース部—

加藤 安彦 小松 英二 安原 宏 椎野 努
(沖電気工業株式会社 総合システム研究所)

0. はじめに

要約支援システムCOGITO⁽¹⁾⁽²⁾⁽³⁾では、インタフェース部に要約文表示を初めとして、グラフィック(PSI-GRAPH)及びテーブル形式を用いた要約結果の表示、テキストの目次情報抽出、語彙頻度カウント、中間言語表示等を行ない、要約結果抽出のための情報の参照及びテキスト理解を支援するようなメニューを用意している。ここではそれらの概要と基本機能について述べる。

1. インタフェースの概要

インタフェース部は大別して要約エディタ、表層的な浅い要約処理部、深い要約処理部という3つの部分からなる。(図1. 参照)

要約エディタは、テーブルエディタとグラフィックエディタからなっており、要約結果をテーブル形式あるいはグラフ形式で表示するためのエディタである。表層的な浅い要約処理部とのインタフェースは、自動目次情報抽出、段落先頭情報抽出、文字列検索及び語彙頻度カウント、自動主述語情報抽出、自動数値情報抽出という5つの機能からなっている。深い要約処理部とのインタフェースは、中間言語表示、コンテキスト・レベルの意味表現抽出⁽²⁾⁽³⁾、要約テーブル表示、要約文抽出⁽³⁾という4つの機能からなる。

2. 要約エディタ

要約エディタは、要約テーブルエディタ及びグラフィックエディタの2種類のエディタを用意している。

要約テーブルは、あらかじめ決められたサイズや形式に従うのではなく、自由に設定できるような融通性に富んだものを目的としている。要約テーブルエディタの操作は簡単明瞭な使い勝手のよさを前提とし、文字列の入力以外はマウスを使用する。それによりテーブルのレイアウトなどユーザの思い通りのものをキーボード入力に比べて迅速かつ簡単に作成することが可能となっている。また、必要に応じて縦書き／横書きの使い分けが可能となるような書式設定機能を設け、漢字入力については、SIMPOSのカナ漢字入力を介しての入力となっている。

テーブルのフレームについては、サイズや形式が異なっても形状も併せて考慮し、フレームラインなどのデータをすべてキャラクタ単位で扱っている。

要約結果の受取りは、マウスでテーブル内に表示されている関連おいた位置に結果を返すようになっている。PSI-GRAPHと呼ばれるグラフィックエディタは、形式や形状を自由なものにできるため、例えば世界地図を描いておいて、各国の国勢情報についてのテキストを要約し、情報を得たい国の上にカーソルを合わせてクリックして要約結果を表示させるといった応用ができ、よりビジュアルな要約が可能である。

画面情報のデータは、要約テーブルエディタではキャラクタ単位で扱っているのに対し、グラフィックエディタではドット単位で扱っている。

エディタのウィンドウは、table_menuウィンドウ、canvasウィンドウ、informationウィンドウ、textウィンドウの4つから構成されており、このうちtable_menuウィンドウだけがテンポラリウィンドウであり、他の3つのウィンドウは階層ウィンドウとなっている。(図2. 及び図3. 参照)

本研究は、第5世代コンピュータプロジェクトの一環としてICOTからの委託で行われたものである

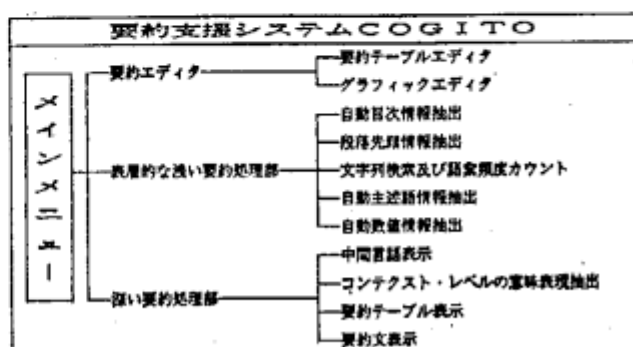


図1. インタフェース・メニュー

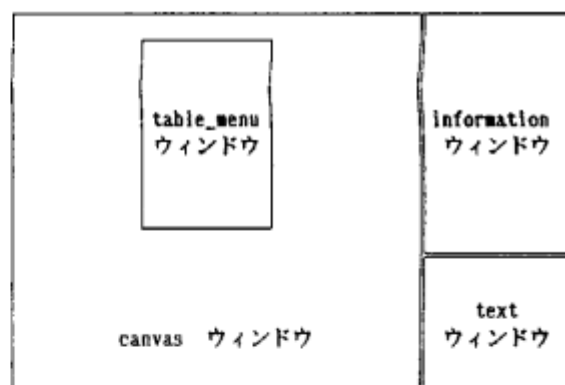


図2. スクリーン上のウィンドウ割

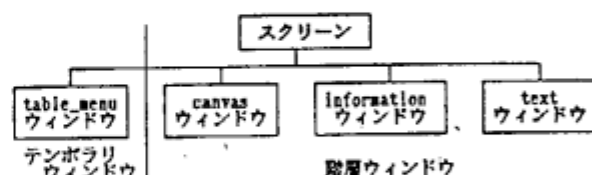


図3. ウィンドウ構成

Summarization Support System COGITO, —Man-Machine Interface—
Yasuhiko KATO, Eiji KOMATSU, Hiroshi YASUHARA, Tsutomu SHIINO
(OKI Electric Industry Co., Ltd.)

- 図4. に要約テーブルエディタの黒隠の画面を示しておく。

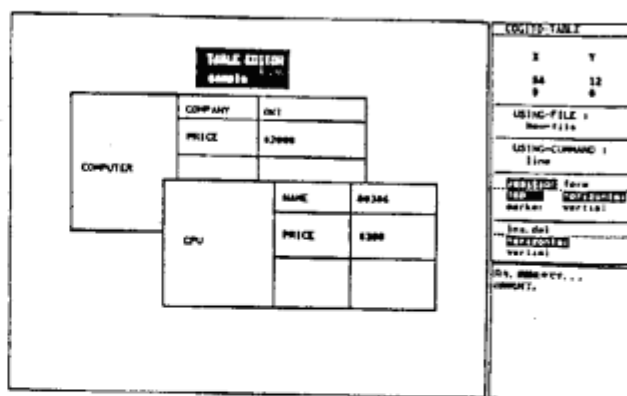


図4. 要約テーブルエディタ実画面

3. 表層的な浅い要約処理部とのインタフェース

以下に浅い要約処理部における各機能の説明をする

①自動目次情報抽出

与えられたテキストから、章、節等の番号を抽出して文中の目次内容を明示する。出力は内部表現と外部表現とで行なう。ここで内部表現とは、他のプログラムから使用できるもので、Prologでのホーン節を用いた表現を指し、外部表現とは、ユーザが直接読むもので、自然言語に近い表現を指す。

②段落先頭情報抽出

テキスト中の各段落における先頭の1文を取り出して表示する。これによりテキストのより具体的な概要を知ることができる。出力は内部表現及び外部表現による。

③文字列検索及び語彙頻度カウント

ユーザが入力した文字列がテキスト内にあるかどうかの検索及び該当文字列の出現頻度のカウントを行なう。検索結果はテキストを表示する時に反転表示で該当文字列を示す。

④自動主述語情報抽出

テキストの各センテンスの最後に位置する用言を抽出する。出力は内部表現と外部表現によって行なう。

⑤自動數值情報抽出

テキスト中に出現する数値を単位系、属性値系の2種類に分けて表示する。出力は内部表現と外部表現とで行なう。

4. 深い要約処理部とのインタフェース

次に深い要約処理部の概要について述べる。

①中間言語表示

テキスト中の各々の文が内部的にどのような形で捉えられているか、弊社独自の中間言語表現を用いて表示する。これにより、修飾-被修飾関係、格関係等の文の構造の把握が容易になる。

②コンテキスト・レベルの意味表現抽出

テキスト中の各センテンスが相互に条件文となるか並列となるかといったセンテンス同士の関係を抽出表示する。

③要約テーブル表示

要約テーブルエディタで作成したテーブルに、対象、属性、属性値をそれぞれマウスをクリックすることにより、要約結果から読み込んでくる。対象は要約結果として得られた話題となる対象のことであり、属性はそれに付随する内容、例えばコンピュータを対象として選んだ場合、価格、開発会社名、販売会社名等が属性として挙げられる。そしてその属性値には具体的な名称、数値が要約結果から導かれてくる。

④要約文表示

7つある文の重要性評価ルールによって、テキスト文中から選ばれた重要と思われる文を抽出して要約文とする。

5. おわりに

現在COGITOのマンマシンインタフェース部分では、要約テーブルエディタ、グラフィックエディタと、要約テーブル表示機能が稼働しており、他の機能についてもインプリメント中である。

要約テーブル表示は、ユーザの意向に沿った対象、属性を要約結果内容から選択できるように考えられた方式になっているが、要約結果内容を自動的にテーブルにて表示することも可能であり、現在その両立ができていないが、そのどちらかをユーザがあらかじめ選択できるようにすれば更に使い勝手のよいシステムとなる。

また、要約テーブルをエディットする場合、Y軸上で位置の異なる2点間に垂直線を引くためには、2点のX軸上での位置の差が0となるように位置設定を行わねばならないが、informationウィンドウ上に2点間の差異が表示されているとはいえ、マウスを用いての操作はやや煩瑣な感じを与えるので、今後はそうした点の改良も必要である。

参考文献

- [1] 北、小松、安原：要約支援システムCOGITO。自然言語処理研究会58-7, 1986
[2] 小松、加藤、安原、椎野：要約支援システムCOGITO-文書の構造解析-。自然言語処理研究会64-11, 1987
[3] 小松、加藤、安原、椎野：要約支援システムCOGITO-要約処理部-。情報処理学会第36回全国大会, 1988