

# ICOT Technical Memorandum: TM-0428他

---

TM-0428他

学会発表論文(日本語処理)

第2研究室

January, 1988

- |         |                                 |       |
|---------|---------------------------------|-------|
| TM 0428 | 要約支援システムCO GITO—要約処理部—          | 小松英二他 |
| TM 0429 | 要約支援システムCO GITO—マンマシンインターフェース部— | 加藤安彦他 |
| TM 0433 | 柔軟な検索機能を持つKWIC検索システムの一方式        | 三吉秀夫他 |
| TM 0434 | 規格化日本文作成支援の一方式                  | 小淵保司他 |
| TM 0436 | 意味談話理解のための知識処理                  | 木下 聰他 |
| TM 0440 | 手段目的解析に基づく対話の解釈                 | 今村 誠他 |
| TM 0485 | 語構成に基づく形態素解析                    | 佐野 洋他 |

©1988, ICOT

**ICOT**

Mita Kokusai Bldg. 21F  
4-28 Mita 1-Chome  
Minato-ku Tokyo 108 Japan

(03) 456-3191-5  
Telex ICOT J32964

---

**Institute for New Generation Computer Technology**

## 要約支援システム COGITO 一要約処理部

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

### 1.はじめに

要約支援システムCOGITOは、言語理解に基づいて要約を行うことをめざしたシステムである。本システムは、言語解析部、文脈処理部、知識ベース部、及び、要約処理部からなっている。本稿では、言語解析部及び文脈処理部について簡単に述べた後、要約処理部について述べる。

### 2.言語解析部・文脈処理部

要約処理のための前処理については、文献[1]、[2]においても述べたが、概略を以下で説明する。

#### 2.1 単文への分割ルーチン

入力：日本語文、出力：分割した日本語文  
文書の構造を決定するために、入力文を単文に分割する。  
単文の定義は、1つの単文に1つの用言があることを基本とするが、サ変名詞の入っている名詞句についてサ変名詞を動詞化して、単文とする処理も行う。

ex. サ変名詞の動詞化の例：

MS-DOSの採用で他のマシンとの互換性が増した。

---> MS-DOSを採用し、他のマシンとの互換性が増した。

#### 2.2 文書解析ルーチン

入力：分割した日本語、出力：格フレーム1

meaning述語と呼ぶ意味表現で表された格フレームを生成する。詳細は文献[1]参照。

ex. meaning述語生成例：

A社は18日32ビットコンピュータAAAを販売した、と発表した。

---> meaning(発表#7, tense, past, []),

meaning(発表#7, agent, A社#1, []),

meaning(発表#7, obj, 販売#1, []),

meaning(販売#6, tense, past, []),

meaning(販売#6, time, 18日#2, []),

meaning(販売#6, obj, AAA#5, []),

meaning(AAA#5, mod, コンピュータ#4, []),

meaning(コンピュータ#4, mod, 32ビット, []).

#### 2.3 言い換えルーチン

入力：格フレーム1、出力：格フレーム2

入力文を標準化することにより、後の処理を容易にする。

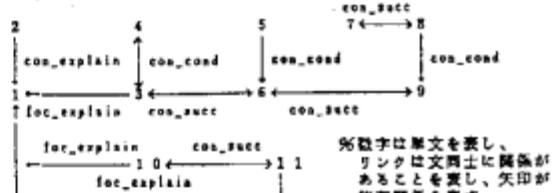
文：A社はスーパーパーソナルコンピュータB800を開発した。

会社	#1
会社名	: A社
製品	: スーパーパーソナルコンピュータ
スーパーパーソナル・コンピュータ	#2
製作会社	: A社
製品名	: B800

図1 フレームの例

入力文：

1. A社、B社は32ビットMPUでオリジナル路線の方針を固めた。
2. 32ビットMPUは今後の主力マーケットとして期待されている。
3. 32ビットハイエンドプロセッサーでは、
4. 開発投資、ソフトウェア実装とアプライケーションの拡張性や影響力などから
5. 従来のような安易な対応には問題があるとし、
6. 国内半導体メーカーでは、セカンダリソース、オリジナル路線と平行して
7. 優良な選択を行ってきたが、MPUも多様化、細分化の時代を迎え、
8. 独自開発品を持つことが重要との判断から、
9. オリジナル路線でいくことを決めた。
10. 国内メーカーでは、C社がオリジナルMPU「Dシリーズ」で32ビットMPUの発売を開始しており、
11. 両社の32ビットMPUのオリジナル路線はこれにつづくもの。



%数字は單文を表し、  
リンクは文書間に関係があることを表し、矢印が依存関係を表す。  
\* --- b : bがaに依存  
\* --- b : aとbが対等

図2 テキスト・トリーの例

ex. サ変名詞の動詞化の例：

A社はB B Bの販売を開始した。

---> A社はB B Bを販売し始めた。

#### 2.4 単文の分類ルーチン

入力：格フレーム2、出力：単文カテゴリ

以下の処理を効率化するため、文をいくつかのカテゴリに分類する。製品記事に関しては、数種類に分類する。

ex. 概要：A社はこのほど、BファミリーにCCCを搭載したDDDシリーズを発表した。

詳細：処理速度が1.5～2.0倍向上させた機種となっている。

作者の見解：何が1台3役なのか。

登場人物の見解：C社のD専務は「OSを含め基盤もできた」と語った。

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。

Summarization Support System COGITO, -Summarizing processing-

Eiji KOMATSU, Yasuhiko KATO, Hiroshi YASUHARA, Tsutomu SHIBANO

OKI Electric Industry Co., Ltd

## 2.5 入力文のフレーム化ルーチン

入力: 格フレーム 2、出力: フレーム 1

格要素をフレーム化する。詳細は、文献[2]参照。

## 2.6 要約処理ルーチン

入力: フレーム 1、出力: フレーム 2

フレーム同士のユニフィケーションにより、同一の事物を表すフレームに同一の識別子をつける。図 1 に出力例を示す。

## 2.7 単文と單文の関係決定ルーチン

入力: 格フレーム + フレーム 2、出力: テキスト・トリー

単文と単文の関係を示すネットワークを生成する。図 2 に出力例を示す。アークには、関係名と、依存関係を表す矢印がつく。

以上のような処理を要約のための前処理として行った結果、要約処理部には、文脈情報としてのテキスト・トリー、文情報としての格フレーム (meaning述語)、対象情報としてのフレームが渡される。要約処理部が、直接、要約の対象とするのは、フレームとテキスト・トリーである。

## 3. 要約処理

本システムは、現在、要約テーブルと要約文を生成する。

### 3.1 要約テーブル

要約テーブルは、フレームをテーブル形式に表示したものである。図 3 に要約テーブルの例を示す。テーブルの項目を予め入れておくことにより、情報検索的に用いることもできる。なお、総ての単文の名詞句をフレーム化して、テーブル表示するのは、無意味な情報が増えるため、単文への分割ルーチンにおいて事物の説明であるカテゴリ（「概要」または「詳細」）に分類された単文だけを選んでテーブル化する。

### 3.2 要約文

要約文は、文書中で重要と評価された単文を選びだしたものである。要約文の生成処理は、まず、テキスト・トリーに対して、重要性評価を行う。図 4 に、重要性評価ルールを示す。ルールは、各単文の重要性の評価値に加える値を示している。評価値は、初めは既に 0 とする。図 2 のトリーに対する評価結果を表 1 に示す。単文を評価値の大きいものから適当な個数だけ選び表示する。連用止めを終止形に直す等の処理を行い、要約文として表示する。図 5 に図 2 の入力文の要約文を示す。

### 4. おわりに

本システムは、現在試作中であるが、実用性を高めるためには、言語解析及び文脈解析の解析率を上げることが必要である。また、重要性評価の各項目に重みをつけるのも課題である。

### 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

販売者	A社	
販売者	販売者と同じ	
販売物	3.2 ビットスーパーパソコン	
ハード	CPU	プロセッサ BBB
	クロック	1.0 MHz
	メイン	最大 8 Mバイト

図 3 要約テーブルの例

ルール 1: パラグラフの最初の単文ならば +1  
/\*パラグラフの最初の単文は重要\*/

ルール 2: + [そのノードを通り、かつ、アーカーを矢印の方向にたどったときの最長のアーカー列の長さ]  
/\*長いリンクは重要\*/

ルール 3: + [そのノードにつながるアーカーの数]  
/\*たくさんのアーカーの数でいる単文は重要\*/

ルール 4: + [そのノードに向かっているアーカーの数]  
/\*たくさんの単文から依存されている単文は重要\*/

ルール 5: 総も大きいフレームを含む単文ならば +1  
/\*重要な対象を含む単文は重要\*/

ルール 6: 断定の法情報を持つ単文ならば +1  
/\*断定の法情報をもつ単文は重要\*/

ルール 7: ユーザが指定した項目を含む単文ならば +3.9.9  
/\*ユーザが指定した項目を含む単文は重要\*/

図 4 単文の重要性評価ルール

表 1 重要性評価の例

ルール 単文 番号	1	2	3	4	5	6	7	総計
1	1	6	4	4	3	0	0	16
2	0	2	1	0	1	0	0	4
3	0	6	2	1	1	0	0	10
4	0	3	1	0	0	0	0	4
5	0	4	1	0	0	0	0	5
6	0	6	3	2	0	0	0	11
7	0	6	1	1	1	0	0	9
8	0	6	2	1	0	0	0	9
9	0	6	2	1	0	0	0	9
10	0	3	2	1	1	0	0	7
11	0	3	2	0	1	0	0	6

### 要約文:

A社、B社は 3.2 ビット MPU でオリジナル基盤の方針を固めた。  
3.2 ビットハイエンドプロセッサーでは、従来のような安易な対応には問題があるとし、慎重な選択を行ってきた。

図 5 要約文の例