

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 聡他 |
| 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

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 照応処理ルーチン

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 単文と単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

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、文脈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、初めは総て 0 とする。図 2 のトリ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り表示する。適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 一文書の構造解析, 自然言語処理研究会 64-11, 1987

販売者	A 社		
開発者	販売者と同じ		
販売物	32 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	10 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、アークを矢印の方向にたどったときの最長のアーク列の長さ)
/*長いリンクは重要*/
- ルール 3: + (そのノードにつながるアークの数)
/*たくさんアークの集まっている単文は重要*/
- ルール 4: + (そのノードに向かっているアークの数)
/*たくさん単文から依存されている単文は重要*/
- ルール 5: 最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: 特定の法情報を持つ単文ならば +1
/*特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 +999
/*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	0	4	4	1	0	0	10
2	0	2	1	0	1	0	0	4
3	0	0	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	0	3	2	0	0	0	11
7	0	0	1	1	1	0	0	3
8	0	0	2	1	0	0	0	3
9	0	0	2	1	0	0	0	3
10	0	0	3	2	1	0	0	7
11	0	3	2	0	1	0	0	6

要約文:
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。
32 ビットハイエンドプロセッサでは、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