

科学技術文献の要約システムについて (1)

5J-9

鈴木康広 上窪真一 栃内香次 永田邦一
北海道大学

1. はじめに

Luhnのキーワード密度法¹⁾,あるいはさきに我々が提案した高頻度隣接語法²⁾など,文献から文章を抜き出して並べるといふ自動抄録法には,作成された抄録において前後の脈絡が欠け抄録全体としてのまとまりを失いやすいという問題点³⁾がある.この問題点を解決するために,我々は文章間の接続関係が保存される自動要約の手法を開発中である.本稿では意味的に関連のあるブロック内における2文間の要約規則を用いる手法について述べる.

2. 要約手順

図-1は本手法の要約手順を示したものである.以下,図に従って説明する.

A. 要約対象文献を入力し,高頻度隣接語を抽出する.

B. 文献を意味的に関連のあるブロックに分割する.一つのブロックの開始点は,段落の初めおよび文頭が継続語以外の語で始まっている文である.ここで,継続語とは「この」「これらの」など前文とのつながりを示す語のことで,現在39語が継続語辞書に登録されている.

C. 各ブロックの高頻度隣接語の密度を算出する.高頻度隣接語とは,文献の内容に関係の深い語の対のことである²⁾.

D. 高頻度隣接語の密度の大きい順に数ブロック(各章ごとに各章のブロック数の20%程度)を抽出する.

E. 抽出された各ブロックごとに2文間の要約規則を用いて要約を行う.なお,要約規則の当てはまらない文章はそのまま出力する.要約規則は現在41種が辞書に登録されている.

F. 各ブロックごとに作成された要約文を並べて文献の要約とする.

3. 2文間の要約規則と要約例

2文間の要約規則は,大別すると4種類に分けられる.以下に,各々の代表的な要約規則と要約例を示す.

1. 一般文の要約規則

rule) .これに対し,.

→ + が, + .

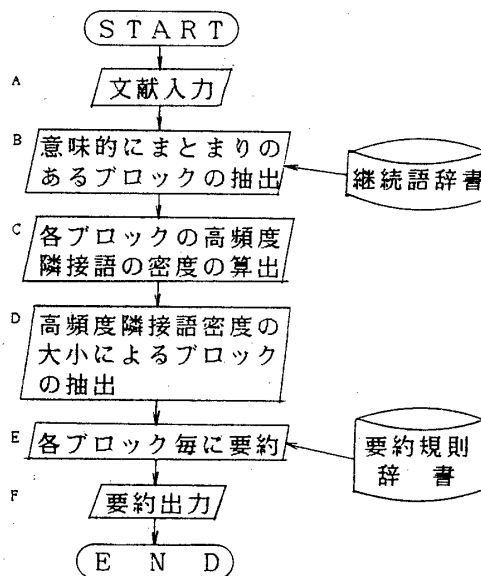


図-1 自動要約の手順

- ex) 命令機能の増強やファミリ展開に対しても同様に膨大な作業量を必要とする。これに対し、ランダム論理をマイクロコード化し、これをROMに集約するマイクロプログラム制御方式は論理を規則化するとともにプロセッサに柔軟性を与える特徴がある。
- 命令機能の増強やファミリ展開に対しても同様に膨大な作業量を必要とするが、マイクロプログラム制御方式は論理を規則化するとともに柔軟性を与える特徴がある。

2. 1文でブロックを構成している文の要約規則

rule) このような .

→ .

- ex) このようなマクロ命令からマイクロプログラムへの直接マッピングを実施するプロセッサはすでに存在している。
- マクロ命令からマイクロプログラムへの直接マッピングを実施するプロセッサはすでに存在している。

3. 箇条書き文に関する要約規則

rule) は、第1ににある。第2ににある。

→ は、

1.

2.

にある。

- ex) マイクロプログラム制御構造によって得られる利点は、第1に μ -ROM中のマイクロ命令の追加・変更によって論理修正を容易にする柔軟性にある。第2にマクロ命令デコーダでマクロ命令解釈の自由度を維持するため、より柔軟性を高めることにある。
- マイクロプログラム制御構造によって得られる利点は、
1. μ -ROM中のマイクロ命令の追加・変更によって論理修正を容易にする柔軟性
 2. マクロ命令デコーダでマクロ命令解釈の自由度を維持するため、より柔軟性を高めること
- にある。

4. 図、表の参照部分を含む文の要約規則

rule) 図 n に示したように、.

→ .

- ex) 図 7 に示したように、マイクロ命令は 1 ビットの PS フィールドを持つ。
- マイクロ命令は 1 ビットの PS フィールドを持つ。

4. おわりに

本稿では、2文間の要約規則を用いる自動要約の手法について述べた。今後、この手法を組み込んだ自動要約システムを作成し性能評価を行う予定である。

参考文献

- (1) Luhn, P.H.: "The automatic creation of literature abstract." IBM Journal Vol.2 (1958)
- (2) 鈴木, 上窪, 栃内, 永田: 高頻度隣接語を利用した科学技術文献の自動抄録
情報処理学会第32回全国大会講演論文集 4T-11 1986
- (3) 水谷: 統計的自動抄録法の問題点 「計量国語学」 27