

科学技術文献の要約システムについて (2)

5J-10

上窪真一 鈴木康広 栃内香次 永田邦一
北海道大学

1. はじめに

文章間の接続関係が保存される自動要約手法として¹⁾、本稿では文の接続関係を利用した自動要約システムについて述べる。科学技術文献等の論理的な文章は、その文章の主題を示すために、各文が論理的に展開されており、これら各文の接続関係を求めることで、文章の要約を生成することが可能である。

2. 要約システムの構成

本システムの処理の流れを図1に示す。

- ① 要約対象文献を入力する。
- ② 入力された文献を後述の要約単位に分割し、各要約単位内の文について、その接続関係を接続詞辞書の情報を用いて決定する。
- ③ 各要約単位の中心文を文の接続関係を利用して決定する。
- ④ 代名詞、接続詞を考慮して中心文を組み合わせ、各章の要約を生成する。
- ⑤ 各章の要約から最終的な要約を生成する。

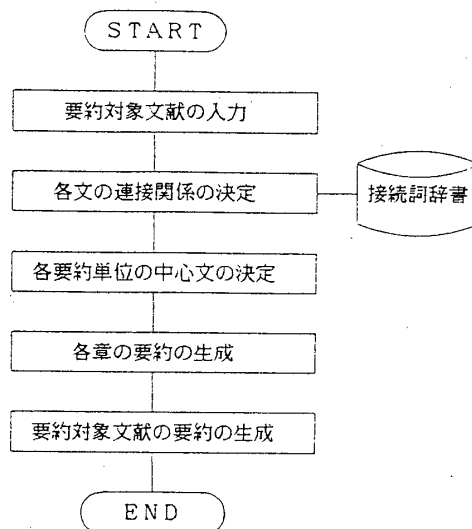


図1 処理の流れ

3. 要約方法

3.1 文の接続関係の種類

文の接続関係は、文と文との論理的なつながりを表わすもので、一般に表1のような8種類の型に分類され、その意味に対応した接続詞がある。

3.2 接続詞辞書

接続関係の種類や接続詞によって、接続詞の前後のどちらの文がより重要かが決っている。そこで、接続詞ごとに前後の文について重要度を示す情報を付加し、接続詞辞書に登録している。

要約対象文献の各文は、必ずしも接続詞で結ばれているとは限らないので、接続詞と同等の働きをする語句を接続詞相当句として、接続詞と同様に接続詞辞書に登録している。

3.3 要約単位

本システムは、要約対象文献の中で意味的にまとまっている部分を要約単位とし、各要約単位ごとに文の接続関係を利用して中心文を決定し、これを中心に要約文を生成する。ここで、要約対象を科学技術文献とすると、関連する事項が一つの章にまとめられている場合が多いので、原則として各章を要約単位とした。

型	内容
順接	前の事柄を受けて、それから順当の結果を生ずる型 例：こうして、したがって、すると
逆接	前の事柄に反する内容を述べる型 例：けれども、しかし、だが
選択	前後の事柄を比べ、どちらかを選択させる型 例：あるいは、それとも、または
並列	前後の事柄を同列に述べ、要約したり詳しく述べる型 例：すなわち、たとえば、つまり
転換	前の事柄とは違う別の事柄を述べる型 例：さて、それでは、ところで
添加	前の事柄に対して、さらに付け加える型 例：さらに、しかも、そして
補足	前の事柄に対して、その原因、理由などを補う型 例：ただし、しかも、そして
連続	前後の事柄を密着させて続けたり、同じ内容を反復する型

表1 文と文とのつながり

3.4 要約単位の中心文の決定

文の接続関係から各文の重要度を求め、各要約単位の中で重要度の高い数文を中心文とする。

一般に、段落の最初の文で重要な事項が述べられている場合が多いので、ある程度の重みを持たせることにした。

3.5 指示代名詞の処理

要約単位の中心文を組み合わせて各章の要約を生成するとき、中心文に指示代名詞が含まれていると、前後の脈絡が欠け、意味が通じない場合があるので、指示代名詞は中心文から削除するかあるいは、直前の文の主語と置換することにした。

3.6 要約の生成

各章の要約をまとめ、その各文に2文間の要約規則¹⁾を適用して、要約対象文献全体の最終的な要約を生成する。

4. 要約例

本自動要約システムによる要約の例を以下に示す。

要約単位

1. はじめに

著者らはCOBOLプログラムを高速に実行する構造を備えたCOBOLマシンを開発したが、本論文ではCOBOLマシンのハードウェア構成について論じている。

これまでに、COBOLマシンの開発は行われているが、これらは、COBOLの言語仕様が低く(ANSI 74仕様を満たしていない)、また、中間言語仕様やハードウェア構成の機能レベルも低い。このために、高い機能、性能が達成されていないと考えられる。

本COBOLマシンはANSI 74 COBOL言語仕様を満たし、COBOLの言語機能に直接対応する高い中間言語仕様を持ち、かつ、中間言語を効率良く処理する高度なハードウェア構成を備えており、これまでのCOBOLマシンと比べて、アーキテクチャ、ハードウェア構成の両面から高級言語指向度が高いCOBOLマシンを実現している。

高機能なCOBOLマシンのアーキテクチャを高速に実現するために、本COBOLマシンは3つのプロセッサから成る複合体で構成される。そして、各プロセッサはそこで要求される処理を高速に行うために、ファームウェアとハードウェアによって専用化されている。また、COBOLマシンは汎用計算機に接続されて動作し、入出力処理などの実行を汎用計算機に依頼するための高性能なインタフェースを備えている。

なお、COBOLマシンのアーキテクチャについては文献4)を参照のこと。

山本 他：COBOLマシンとその設計思想 情報処理学会論文誌 Vol.23 No.1

要約

著者らはCOBOLプログラムを高速に実行する構造を備えたCOBOLマシンを開発したが、本論文ではCOBOLマシンのハードウェア構成について論じている。本COBOLマシンはANSI 74 COBOL言語仕様を満たし、COBOLの言語機能に直接対応する高い中間言語仕様を持ち、かつ、中間言語を効率良く処理する高度なハードウェア構成を備えており、これまでのCOBOLマシンと比べて、アーキテクチャ、ハードウェア構成の両面から高級言語指向度が高いCOBOLマシンを実現している。

5. おわりに

本稿では、文の接続関係を用いた自動要約の手法について述べた。今後、前論文¹⁾の手法を組み合わせるなどして、より質の高い自動要約システムを開発していく予定である。

参考文献

- 1) 鈴木, 上窪, 栃内, 永田: 科学技術文献の要約システムについて (1)