

英文要約システム『DIET』

6D-9

石橋弘義, 重永信一, 新納浩幸, 福重貴雄, 安川秀樹
松下電器産業(株) 東京研究所

1 はじめに

大量の文書の中から必要な文書を検索する時、全体を読まなくても要点のみを見てそれが必要かどうか判断できれば、望む文書を迅速に取り出すことが可能となる。このため、文書の要点を取り出したり、要約文を生成するシステムが望まれている。

本システムは、英語の文章の中から重要と思われる文を文単位で選択し入力文章上に示す、オンライン的使用を目的とした要約システムである。また、文脈理解などの深い意味理解は行わずに、主に語彙的、構文的な情報や文の接続関係を手掛かりにして、要点を含む文を選択する新しい簡易な要約手法を提供している。

2 特徴

- (1) 記述内容の分野を限定しない新しい、簡易な要約手法である。
- (2) 対象文章は論説調の文章(著者の主張が展開されている文章)とする。
- (3) 重要と思われる文を文単位で選択し、入力文章上に示す。
- (4) 問題提起、結論、理由などを選ぶという観点により、文を選択する。
- (5) 文脈理解などの深い意味理解は行わずに、主に語彙的、構文的な情報や文の接続関係を手掛かりにして、要点を含む文を選択する。

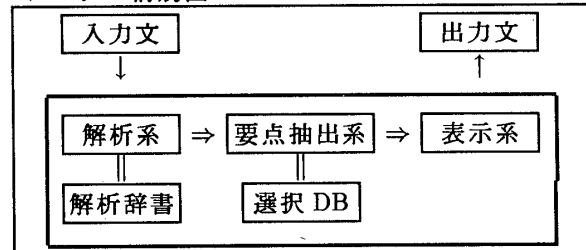
3 システム構成、機能

本システムは、解析系、要点抽出系、表示系から成る。

解析系は、文章を読み込み、解析用辞書を使って構文意味解析し、さらに、代名詞処理などの照応解析を行なう。要点抽出系は、選択データベースを基にして要点抽出アルゴリズムの適用を行ない、文を選択するかどうかを決定する。表示系は、選択された文章を入力文章上でカラー表示する。

ユーザは、まず、示された選択文を読み、文章の

システム構成図



全体像を掴む。そして、必要なら選択された文章の前後の文章を読むことにより、詳細を理解することができる。このように、本システムは、オンライン的使用を目的とした、いわゆる斜め読みの機能を実現するためのシステムである。

4 対象文章

対象文章は著者の主張が展開されている文章、つまり、論説調の文章とし、記述分野については限定しない。開発用には、英字新聞「The Japan Times」の社説を用いている。

文章理解という立場で考えると、著者の意思の入った文は扱いにくい。しかし、要約の観点から考えると、論説調の文章は重要な点が明確であるというのが長所である。つまり、重要なのは著者が何を言いたいかである。こうした文章では結論や論理の展開を把握するのに効果的な語句や文型が要所要所に使用されているので、それらを検出することにより、要点を抽出できる。

5 解析処理

動詞の格を中心に解析するデモンベースのパーザ[例えば、McDYPAR]に文法規則を導入したパーザを用いる。これにより、構文が複雑で完全な解析ができないような場合でも、ある程度の解析が可能となる。

DIET: Digest Extracting System

Hiroyoshi ISHIBASHI, Shin-ichi SHIGENAGA, Hiroyuki SHINNOU, Yoshio FUKUSHIGE, Hideki YASUKAWA
Tokyo Research Laboratory, Matsushita Electric Industrial Co., Ltd.

Internationalization in crime

As though in anticipating of its first anniversary this week, crime fans have been indulged these past weeks with daily new developments - speculations, if nothing else - in the investigation unraveling the spectacular ¥333 million robbery of a cash delivery van in Yurakucho, the heart of Tokyo, on November 25, 1986.

The biggest cash robbery in Japanese history contains all the elements of the fictional genre except, so far, sex. It is as though perpetrators deliberately left that plethora of clues to thrill the law abiding public... and, it must be remembered, to thumb their nose at us.

From the inception, of course, there was the fundamental element of the foreign identity of the suspected criminals, which was guaranteed to fascinate the Japanese especially. **How exotic that they turn out to be French!**

But the central point of mystery has been whether all those carefully laid plans could have been worked out without the participation of Japanese. No, says almost everyone, and indeed one Japanese suspect, who had met the two alleged leaders of the Yurakucho heist in a French prison, has been arrested. His apparent complicity has expanded the case to take in the theft of famous paintings from French museums and attempts to sell them in this country.

The National Police Agency made internationalization of crime the theme of its recently issued annual white paper, pointing out a fivefold increase in foreigners arrested and, as well, the growing incidence of Japanese involved in crimes abroad, such as murder for insurance and trafficking by yakuza in guns and drugs. But the ¥333 million robbery is different- international in the true sense, i.e., of cooperation.

Crime has never known borders, except to labor under the same language and custom handicaps as legitimate enterprises. These are fading in the criminal world, too.

But if international cooperation is growing among criminals, so must it also among law and peace enforcers. We are pleased to see that this is not lacking in this case, in which the resources of Interpol, the world police agency headquarters in Paris, are being well applied.

So the more the world changes, the more it remains the same. **That is, the more its needs are reiterated and, we hope, the responses reinforced.**

解析用辞書には、意味として概念カテゴリーが記述されている。さらに、動詞については必須格や自由格に対する制約条件、形容詞については被修飾語のカテゴリーについても書かれている。

6 要点抽出処理

論説調の文章は「前置き、問題提起、論理の展開、結論、補足文」などからなる。このうち、問題提起、結論、理由などを選ぶという観点に立って文を選択する。これらは、比較的表層的な情報、つまり、主に語彙的、構文的な情報や文の接続関係を手掛かりにして選ぶことができる。

語彙的情報としては、「should」や「must」などの主張を表わす助動詞、「deserve」などの評価を表わす動詞、「to solve the problem」など解決策を示す言い回しなどの出現を用いる。

構文的情報としては、問題提起を表わす疑問文、評価を表す感嘆文、主張を表わす強調構文などの使用に注目する。

文の接続関係については、接続詞や段落、代名詞の指しているものなどの情報を用いる。

選択データベースはこれらの情報を記述したものである。

7 実行例

太文字の部分は選択された文を示し、下線の部分は選択の手掛かりとなった語句や文型を示す（本来

問題提起: 疑問文, The news that, The central point of, 結論の対象を示した文(代名詞の利用)

問い掛けの答え(結論): 疑問文の答え

主張(結論): must, should, We hope

評価(結論): deserve, We are pleased to, 感嘆文

解決策(結論): to solve the problem

理由: because, The reason that

選択理由と選択語句、文型の例

はカラー表示される)。この例の場合、入力文16文に対して、6文が選択される。

8 おわりに

今後、論説調の文章以外への文章についても適用できるように、対象分野を拡大する必要がある。また、意味理解技術が実用に達した場合、それらを本システムに組み込むことにより、より高度な選択基準が確立されるであろう。さらに、長期的には言い換えを含む要約システムを考えている。これらは、データベースシステムと組み合わせた、知的文書検索システムなどへの応用が可能である。

参考文献

[McDYPER] M. G. Dyper : In-Depth Understanding, the MIT Press, 1983