(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3832613号
(P3832613)
(45)発行日　平成18年10月11日(2006.10.11)　(24)登録日　平成18年7月28日(2006.7.28)

(51)Int.Cl.　F I
G06F 17/30 (2006.01)　G06F 17/30 220A
G06F 17/30 170A

請求項の数 8　(全 14 頁)

(21)出願番号　特願平10-270147
(22)出願日　平成10年9月24日(1998.9.24)
(65)公開番号　特開2000-99536(P2000-99536A)
(43)公開日　平成12年4月7日(2000.4.7)
審査請求日　平成16年4月8日(2004.4.8)

特許権者において、実施許諾の用意がある。

(73)特許権者　000004352
日本放送協会
東京都渋谷区神南２丁目２番１号
(74)代理人　100077481
弁理士　谷　義一
(74)代理人　100088915
弁理士　阿部　和夫
(74)代理人　100105371
弁理士　加古　進
(72)発明者　加藤　直人
東京都世田谷区砧一丁目１０番１１号　日本放送協会　放送技術研究所内

審査官　辻本　泰隆

最終頁に続く

(54)【発明の名称】自動要約装置および自動要約プログラムを記録した記録媒体

(57)【特許請求の範囲】
【請求項１】
入力された原文を形態素解析し、原文に含まれる文字列を単語に分割しその品詞を付与した形態素情報を出力する形態素解析手段と、
入力された原文の要約率と入力された該原文の文字数とから削減すべき文字数の最小値を計算する文字削減数最小値計算手段と、
前記形態素解析手段から出力される前記形態素情報に基づいて、前記原文の単語列に対して先頭の単語列から順に、あらかじめ得られている置換知識との照合を行い、照合に成功した場合には置換する単語列を前記形態素情報に追加し、かつその置換コストを付与した単語ラティス構造を出力する置換知識検索手段と、
該置換知識検索手段から出力される前記単語ラティス構造を文末から文頭に向けて探索し、該単語ラティス構造の文末から各ノードまでの最大可能な文字削減数を計算する後向き文字削減数計算手段と、
前記置換知識検索手段から出力される前記単語ラティス構造を文頭から文末に向けて探索し、各パスに対して文頭からの文字削減数を計算し、その計算結果と前記後向き文字削減数計算手段から出力された文字削減数との和を求め、その和が所望の文字削減数よりも小さい場合には、そのパスを枝刈りし、その和が所望の文字削減数よりも小さくない場合には、そのパスを出力する前向き文字削減数計算手段と、
該前向き文字削減数計算手段から出力されるパスに対してその置換コストの和を計算し、文末に達したときには文頭から文末までのパスを出力する置換コスト計算手段と、

該置換コスト計算手段から出力されたパスの中で、文頭から文末までの置換コストの和が最小となるパスを求め、求めた最小パスの文字列を要約文として出力する置換コスト最小パス計算手段と
を有することを特徴とする自動要約装置。
【請求項２】
前記置換知識検索手段は、パスごとに必須適用置換知識リストをもっておき、置換知識を適用する時にそのリストを参照することを特徴とする請求項１に記載の自動要約装置。
【請求項３】
前記後向き文字削減数計算手段は、前記単語ラティス構造の文末ノードｎ（ｎは単語数）から文頭ノード０に向けてノード番号ｉを１ずつ減少させて、各ノードｉにおいて、文末ノードｎからその現ノードｉまでの最大可能な文字削減数すなわち文字削減数の最大値m_b（ｉ）をダイナミックプログラミングにより求めることを特徴とする請求項１または２記載の自動要約装置。
【請求項４】
前記前向き文字削減数計算手段は、文末ノードから文頭ノードまでの文字削減数の最大値m_b（０）が、削減すべき所望の文字数の最小値ｍよりも小さい場合には、入力された要約率では要約できない旨を出力することを特徴とする請求項１ないし３のいずれかに記載の自動要約装置。
【請求項５】
前記前向き文字削減数計算手段は、文末ノードから文頭ノードまでの文字削減数の最大値m_b（０）が、削減すべき文字数の最小値ｍよりも小さくない場合には、前記単語ラティス構造の文頭ノード０から文末ノードｎに向けてノード番号ｉを１ずつ増加させて、各ノードｉまでのすべてのパスにおいて、そのパスの文字削減数m_f（ｉ）を計算し、前記後向き文字削減数計算手段で得られた前記ノードｉにおける文字削減数m_b（ｉ）との和をとり、その和m_f（ｉ）＋m_b（ｉ）が削減すべき文字数の最小値ｍよりも小さい場合には，そのパスを枝刈りし、その和m_f（ｉ）＋m_b（ｉ）が削減すべき文字数の最小値ｍよりも小さくない場合には、そのパスを出力することを特徴とする請求項４に記載の自動要約装置。
【請求項６】
前記前向き文字削減数計算手段は、前向き計算の処理中に、正例がない単語に達したときに、条件（前向き文字削減数、必須適用置換リスト等）の同じ候補は、置換コストが最小でないパスも枝刈りすることを特徴とする請求項５に記載の自動要約装置。
【請求項７】
コンピュータによって原文と所望の要約率から該要約率の中で最適な要約文を自動的に求めるための自動要約プログラムを記録した記録媒体であって、該自動要約プログラムはコンピュータに、
入力された原文を形態素解析させることで原文に含まれる文字列を単語に分割させ、かつその品詞を付与した形態素情報を生成させ、
入力された原文の要約率と入力された原文の文字数とから削減すべき文字数の最小値を計算させ、
前記形態素情報に基づいて、前記原文の単語列に対して先頭の単語列から順に、あらかじめ得られている置換知識との照合を行わせ、照合に成功した場合には置換する単語列を前記形態素情報に追加して、かつその置換コストを付与した単語ラティス構造を求めさせ、
前記単語ラティス構造を文末から文頭に向けて探索させて、該単語ラティス構造の文末から各ノードまでの最大可能な文字削減数（後向き文字削減数）を計算させ、
前記単語ラティス構造を文頭から文末に向けて探索させて、各パスに対して文頭からの文字削減数を計算させ、その計算結果と前記後向き文字削減数との和を求め、その和が所望の文字削減数よりも小さい場合には、そのパスを枝刈りさせ、その和が所望の文字削減数よりも小さくない場合には、そのパスを選出させ、
該選出されたパスに対してその置換コストの和を計算させ、文末に達したときには文頭か

ら文末までのパスを出力させ、
該出力されたパスの中で、文頭から文末までの置換コストの和が最小となるパスを求め、求めた最小パスの文字列を要約文として出力させることを特徴とする自動要約プログラムを記録した記録媒体。
【請求項８】
前記自動要約プログラムはコンピュータに、
文末ノードから文頭ノードまでの文字削減数の最大値が、削減すべき所望の文字数の最小値よりも小さい場合には、入力された要約率では要約できない旨を出力させることを特徴とする請求項７に記載の自動要約プログラムを記録した記録媒体。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】
本発明は、原文に対して所望の要約率が与えられたときに、その所望の要約率の中で最適な要約文を自動的に求める自動要約を行う自動要約装置および自動要約プログラムを記録した記録媒体に関する。
【０００２】
本発明は、ＴＶニュース等の文章を自動的に要約するのに好適であり、また字幕作成への応用なども考えられる。
【０００３】
【従来の技術】
自動要約とは、原文の単語列を短い単語列に置換することにより、原文を自動的に縮約することである。自動要約を実現するためには、原文のどの単語列をどのような単語列に置換するのかという置換知識が必要となる。例えば、次が置換知識の一例である。
【０００４】
［置換知識１］　（原文単語列→要約文単語列）
［置換知識１ａ］　明らかにしました→表明，置換コスト＝０．６
［置換知識Ｉｂ］　明らかに→表明，置換コスト＝０．３
［置換知識１ｃ］　まし→φ（φは空、省略を表す記号），置換コスト＝０．２
ここで、置換知識に付属している「 置換コスト」 は、その置換知識を使う際のペナルティーを表しており、置換コストが０に近いほどその置換知識は使いやすいとしている。
【０００５】
上述の置換知識１の例を使って、下記の原文１を次の要約率で要約することを考えてみよう。なお、要約率＝要約文の文字数／原文の文字数×１００とする。
［原文１］　「明らかにしました」　（８文字）
［要約率１］　要約率６０％以下。
【０００６】
要約率を６０％以下にするということは、削除する文字数でいうと次のようになる。
【０００７】
８文字×（１００－６０）％＝３．２文字
すなわち、原文中の３．２文字以上の文字数を削減しなければならない。
【０００８】
従来の自動要約では、置換知識を出現順に順次適用することにより要約文を得ていた。また、そこで使われる置換知識は人手で収集し作成していたために、その数は非常に少なかった。（例えば、文献１：『山本ほか：「文章内構造を複合的に利用した論説文要約システムＧＲＥＥＮ」自然言語処理、Ｖｏ１．２，Ｎｏ．２，ｐｐ．３９‐５５，Ｉ９９４』は、人手で作成した置換知識をはじめとする要約知識を順次適用することにより自動要約している。）。
【０００９】
要約率６０％以下の例では、上記文献１と同様に、削減文字数３．２文字以上を置換知識１の出現順に順次適用すると、置換知識１ａが適用され、下記の要約文１ａが得られる。

【００１０】
［要約文１ａ］　「表明」　（削減文字数＝６文字、置換コストの和＝０．６）
【００１１】
【発明が解決しようとする課題】
しかし、最近、置換知識を自動的に作成する手法が開発され、大量の置換知識を簡単に得ることができるようになった。（例えば、文献２：『加藤直人：「ニュース文要約のための局所的要約知識獲得とその評価」電子情報通信学会言語理解とコミュニケーション研究会，ＮＬＣ９８－１６，ｐｐ．７－１４，１９９８』）。
【００１２】
このようにして置換知識が大幅に増えると、今度は１つの原文の単語列に対して複数の置換知識を適用できる場合があり、そのため自動要約の際に、適用する置換知識間で競合が生じるようになって、最適な要約文が求められないという場合があった。
【００１３】
上述の例でも、置換知識１ａを適用して得られる要約文１ａの場合と、置換知識１ｂと１ｃを適用して得られる要約文１ｂの場合が競合するが、下記の要約文１ｂのほうが置換コストの和が０に近いので適切な要約となる。
【００１４】
［要約文１ｂ］　「表明した」　（削減文字数＝４文字，置換コストの和＝０．５）
そこで、本発明の目的は、上述のような点に鑑みて、原文と所望の要約率が与えられたときに、置換コストと文字削減数を利用して、置換知識を適切にかつ効率的に選択することにより、原文の最適な要約文を自動的に得ることを可能にすることにある。
【００１５】
【課題を解決するための手段】
上記目的を達成するため、請求項１の自動要約装置の発明は、入力された原文を形態素解析し、原文に含まれる文字列を単語に分割しその品詞を付与した形態素情報を出力する形態素解析手段と、入力された原文の要約率と入力された該原文の文字数とから削減すべき文字数の最小値を計算する文字削減数最小値計算手段と、前記形態素解析手段から出力される前記形態素情報に基づいて、前記原文の単語列に対して先頭の単語列から順に、あらかじめ得られている置換知識との照合を行い、照合に成功した場合には置換する単語列を前記形態素情報に追加し、かつその置換コストを付与した単語ラティス構造を出力する置換知識検索手段と、該置換知識検索手段から出力される前記単語ラティス構造を文末から文頭に向けて探索し、該単語ラティス構造の文末から各ノードまでの最大可能な文字削減数を計算する後向き文字削減数計算手段と、前記置換知識検索手段から出力される前記単語ラティス構造を文頭から文末に向けて探索し、各パスに対して文頭からの文字削減数を計算し、その計算結果と前記後向き文字削減数計算手段から出力された文字削減数との和を求め、その和が所望の文字削減数よりも小さい場合には、そのパスを枝刈りし、その和が所望の文字削減数よりも小さくない場合には、そのパスを出力する前向き文字削減数計算手段と、該前向き文字削減数計算手段から出力されるパスに対してその置換コストの和を計算し、文末に達したときには文頭から文末までのパスを出力する置換コスト計算手段と、該置換コスト計算手段から出力されたパスの中で、文頭から文末までの置換コストの和が最小となるパスを求め、求めた最小パスの文字列を要約文として出力する置換コスト最小パス計算手段とを有することを特徴とする。
【００１６】
ここで、好ましくは、前記置換知識検索手段は、パスごとに必須適用置換知識リストをもっておき、置換知識を適用する時にそのリストを参照する。
【００１７】
また、好ましくは、前記後向き文字削減数計算手段は、前記単語ラティス構造の文末ノードｎ（ｎは単語数）から文頭ノード０に向けてノード番号ｉを１ずつ減少させて、各ノードｉにおいて、文末ノードｎからその現ノードｉまでの最大可能な文字削減数すなわち文字削減数の最大値m_b（ｉ）をダイナミックプログラミングにより求める。

【００１８】
また、好ましくは、前記前向き文字削減数計算手段は、文末ノードから文頭ノードまでの文字削減数の最大値m_b（０）が、削減すべき所望の文字数の最小値ｍよりも小さい場合には、入力された要約率では要約できない旨を出力する。
【００１９】
また、好ましくは、前記前向き文字削減数計算手段は、文末ノードから文頭ノードまでの文字削減数の最大値m_b（０）が、削減すべき文字数の最小値ｍよりも小さくない場合には、前記単語ラティス構造の文頭ノード０から文末ノードｎに向けてノード番号ｉを１ずつ増加させて、各ノードｉまでのすべてのパスにおいて、そのパスの文字削減数m_f（ｉ）を計算し、前記後向き文字削減数計算手段で得られた前記ノードｉにおける文字削減数m_b（ｉ）との和をとり、その和m_f（ｉ）＋m_b（ｉ）が削減すべき文字数の最小値ｍよりも小さい場合には，そのパスを枝刈りし、その和m_f（ｉ）＋m_b（ｉ）が削減すべき文字数の最小値ｍよりも小さくない場合には、そのパスを出力する。
【００２０】
また、好ましくは、前記前向き文字削減数計算手段は、前向き計算の処理中に、正例がない単語に達したときに、条件（前向き文字削減数、必須適用置換リスト等）の同じ候補は、置換コストが最小でないパスも枝刈りする。
【００２１】
上記目的を達成するため、請求項７の記録媒体の発明は、コンピュータによって原文と所望の要約率から該要約率の中で最適な要約文を自動的に求めるための自動要約プログラムを記録した記録媒体であって、該自動要約プログラムはコンピュータに、入力された原文を形態素解析させることで原文に含まれる文字列を単語に分割させ、かつその品詞を付与した形態素情報を生成させ、入力された原文の要約率と入力された原文の文字数とから削減すべき文字数の最小値を計算させ、前記形態素情報に基づいて、前記原文の単語列に対して先頭の単語列から順に、あらかじめ得られている置換知識との照合を行わせ、照合に成功した場合には置換する単語列を前記形態素情報に追加して、かつその置換コストを付与した単語ラティス構造を求めさせ、前記単語ラティス構造を文末から文頭に向けて探索させて、該単語ラティス構造の文末から各ノードまでの最大可能な文字削減数（後向き文字削減数）を計算させ、前記単語ラティス構造を文頭から文末に向けて探索させて、各パスに対して文頭からの文字削減数を計算させ、その計算結果と前記後向き文字削減数との和を求め、その和が所望の文字削減数よりも小さい場合には、そのパスを枝刈りさせ、その和が所望の文字削減数よりも小さくない場合には、そのパスを選出させ、該選出されたパスに対してその置換コストの和を計算させ、文末に達したときには文頭から文末までのパスを出力させ、該出力されたパスの中で、文頭から文末までの置換コストの和が最小となるパスを求め、求めた最小パスの文字列を要約文として出力させることを特徴とする。
【００２２】
ここで、好ましくは、前記自動要約プログラムはコンピュータに、文末ノードから文頭ノードまでの文字削減数の最大値が、削減すべき所望の文字数の最小値よりも小さい場合には、入力された要約率では要約できない旨を出力させる。
【００２３】
【発明の実施の形態】
本発明の実施の形態を説明するに先立ち、本発明に係る自動要約に必要な上述の置換コストについて説明する。
【００２４】
自動要約に必要な要約知識は、置換知識と置換条件の２つから構成されている。置換知識は上述のように原文の単語列をどのような単語列に置換するかを規定する知識である。例えば、連体助詞の「 の」 を省略するという知識である。一方、置換条件とは置換知識の適用の良否を数値化したもの、すなわち上述の置換コストである。置換知識はその前後の単語列によって適用の良否が決まる。例えば、「 日本の銀行」 の「 の／体助」 を省略することはできない。

【００２５】
そこで、置換コストは、置換知識の前後の単語列と、あらかじめ獲得しておいた置換条件との距離を計算している。すなわち、ｉ番目からｊ番目までの単語列w_{ij}を、単語列x_{ij}に置換するという置換コストをdistsub（$w_{ij}\rightarrow x_{ij}$）と表すと、（１）式で定義される（さらに詳しくは、上記文献２を参照。）。
【００２６】
【数１】

$$\mathrm{distsub}\ (w_{ij}\rightarrow x_{ij}) = \begin{cases} g'\ (w_{ij}\rightarrow x_{ij}) & \text{if正例があるとき} \\ 0.0 & \text{otherwise} \end{cases} \quad \cdots (1)$$

ただし、

$$g'\ (w_{ij}\rightarrow x_{ij}) = (1.0 - g_{low}) \times g\ (w_{ij}\rightarrow x_{ij}、\text{正例}) + g_{low} \quad \cdots (1a)$$

$$g_{low} = 0.01 \quad \cdots (1b)$$

【００２７】
上記(1) 式は、正例がある場合には、
g_{low}（＝0.01）～1.0 の値（0.0 ≦ｇ（$w_{ij}\rightarrow x_{ij}$1.0 、正例）≦1.0 ））を取り、0.0 に近いほど置換することが可能であると定義されている。また、正例がない（適用される置換知識がない）場合には0.0 を取る。
【００２８】
また、本発明による自動要約アルゴリズムの概要を説明する。説明を簡単にするために、以下では１文を要約する場合を考える。複数の文にわたる場合には単純に連結すればよい。
【００２９】
今、原文をある要約率以下に要約したいとする。このとき、ｍ（＝原文の文字数×要約率）文字以上の文字を削除しなければならない。さらに、最適な要約であってほしい。ここで、「 最適な要約」 とは、適用した置換知識のコストの和（置換コスト）が最小となる場合であると定義する。したがって、自動要約とは、ｍ文字以上の文字数を削除し、文頭から文末までの置換コストが最小のパス（最適パス）を求めることである。定式化すると、（２）式のようになる。
【００３０】
【数２】

$$\underset{x\in X}{\mathrm{argmin}} \sum_{x} \mathrm{distsub}\ (w_{ij}\rightarrow x_{ij}) \quad (2)$$

$$X=\{(x_0, \cdots, x_{ij}, \cdots, x_n) \mid \Sigma\ (|w_{ij}|-|x_{ij}|\ ) \geqq m)\ \}$$

（２）式の解を求めるアルゴリズムについては図１、図２を用いて後述する。なお、本発明では、文字削減数と置換コストという２つの評価関数を用いているが、前者を計算する際にヒューリスティック関数（現在のノードからゴールまでの評価関数の予測値）を用いている。
【００３１】
以下、図面を参照して本発明の実施形態を詳細に説明する。

【００３２】
図１は、本発明の一実施形態の装置構成を示す。図１において、ｉｌ は原文を入力する端子であり、ｉ２は要約率を入力する端子である。ｏｌ は要約文を出力する端子である。
【００３３】
形態素解析装置１は、ｉ１の端子に入力された原文を形態素解析し、原文に含まれる文字列を単語に分割しその品詞を付与した形態素情報を出力する。
【００３４】
文字削減数最小値計算装置２は、ｉ２の端子に入力された原文の要約率とｉ１の端子に入力された原文の文字数とから削減すべき文字数の最小値を計算する。置換知識検索装置３は、形態素解析装置１から出力される形態素情報に基づいて、原文の単語列に対して先頭の単語列から順に、あらかじめ得られている置換知識との照合を行い、照合に成功した場合には置換する単語列を上記形態素情報に追加し、かつその置換コストを付与した単語ラティス構造を出力する。本例では、その置換コストは、各置換知識に付随して置換知識と共に内部メモリ（図示しない）にあらかじめ格納されているものとする。
【００３５】
後向き文字削減数計算装置４は、置換知識検索装置３から出力される単語ラティス構造を文末から文頭に向けて探索し、その単語ラティス構造の文末から各ノードまでの最大可能な文字削減数を計算する。
【００３６】
前向き文字削減数計算装置５は、置換知識検索装置３から出力される単語ラティス構造を文頭から文末に向けて探索し、各パスに対して文頭からの文字削減数を計算し、その計算結果と後向き文字削減数計算装置４から出力された文字削減数との和を求め、その和が所望の文字削減数（入力された上記原文と要約率から算出）よりも小さい場合には、そのパスを枝刈りし、その和が所望の文字削減数よりも小さくない場合には、そのパスを出力する。
【００３７】
置換コスト計算装置６は、前向き文字削減数計算装置５から出力されるパスに対してその置換コストの和を計算し、文末に達したときには文頭から文末までのパスを出力する。
【００３８】
置換コスト最小パス計算装置７は、置換コスト計算装置６から出力されたパスの中で、文頭から文末までの置換コストの和が最小となるパスを求め、求めた最小パスの文字列を要約文としてｏ１の端子から出力する。
【００３９】
図２は、図１の装置構成により自動要約の処理を行う手順の一例を示すフローチャートである。図２に従って、以下、本発明による自動要約の手順を説明する。
【００４０】
まず、ｉ１の端子に原文が入力されると、ステップＳ１では形態素解析装置１によって原文が形態素解析され、その単語分割と品詞が出力される。ｉ２の端子に要約率が入力されると、次のステップＳ２では原文の文字数と要約率から削減すべき文字数の最小値が計算される。
【００４１】
続くステップＳ３では、上記ステップＳ１で得られた形態素解析結果に基づいて、先頭の単語列から順に、あらかじめ得られている置換知識との照合を行い、照合に成功した場合には、形態素解析結果に置換する単語列を追加し、その置換コストを付与して単語ラティス構造を作成する。
【００４２】
ステップＳ４～Ｓ７では、後向きの計算を行う。すなわち、ステップＳ４、Ｓ６、Ｓ７で単語ラティス構造の文末ノードｎ（ｎは単語数）から文頭ノード０に向けてノード番号ｉを１ずつ減少させて、各ノードｉにおいてステップＳ５の処理を実行する。ステップＳ５では、文末ノードｎからその現ノードｉまでの最大可能な文字削減数（後ろ向き文字削減

数）すなわち文字削減数の最大値m_b（ｉ）をダイナミックプログラミング（ＤＰ）により求め、保存する。

【００４３】

次のステップＳ８では、文末ノードから文頭ノードまでの文字削減数の最大値m_b（０）が、削減すべき所望（指定の）の文字数の最小値ｍよりも小さい場合には、ステップＳ９を実行した後、本自動要約処理を終了する。ステップＳ９では入力された要約率では要約できない旨を出力する。一方、文末ノードから文頭ノードまでの文字削減数の最大値m_b（０）が、削減すべき文字数の最小値ｍよりも小さくない場合には、ステップＳ１０を実行する。

【００４４】

ステップＳ１０～Ｓ１６では前向きの計算を行う。ステップＳ１０、Ｓ１５、Ｓ１６において単語ラティス構造の文頭ノード０から文末ノードｎに向けてノード番号ｉを１ずつ増加させて、各ノードｉまでのすべてのパスにおいてステップＳ１１～Ｓ１４の処理を実行する。まず、ステップＳ１１では、そのパスの文字削減数（前向き文字削減数）m_f（ｉ）を計算し、上述のステップＳ５で求めたノードｉにおける文字削減数m_b（ｉ）との和をとり、ステップＳ１２でその和m_f（ｉ）＋m_b（ｉ）が削減すべき文字数の最小値ｍよりも小さい場合には，このパスは最終的な解となり得ないので、ステップＳ１３でそのパスを枝刈りする。文末にしたがい可能なパスの候補が増加していくが、このような枝刈りにより候補数を抑えることができる。一方、その和m_f（ｉ）＋m_b（ｉ）が削減すべき文字数の最小値ｍよりも小さくない場合には、ステップＳ１４を実行する。ステップＳ１４ではそのパスの置換コストcost（ｉ）の和を求め、文末に達しているときには文頭から文末までのパスを出力する。

【００４５】

次のステップＳ１７では、上記のステップＳ１４で求められたパスの中で、置換コストの和が最小となるパスを求め、要約文をｏ１の端子から出力する。

【００４６】

さらに、図２の処理の具体的な一例を、下記の原文２を次のように要約する場合を用いて説明する。

【００４７】

［原文２］

「福沢総理大臣は特別委員会で方針を明らかにしました」　（２４文字）

［要約率２］

要約率７０％以下。

【００４８】

また、置換知識として、次があらかじめ得られているとする。

【００４９】

［置換知識２］

［置換知識２ａ］　総理大臣／普通名詞→首相／普通名詞，
置換コスト＝０．０１

［置換知識２ｂ］　委員会／普通名詞→委／普通名詞，
置換コスト＝０．０２

［置換知識２ｃ］　明らか／形容名詞　に／格助詞に　し／サ変連用
まし／助動丁寧　た／助動過去
→表明／サ変名詞，置換コスト＝０．６

［置換知識２ｄ］　明らか／形容名詞　に／格助詞
→表明／サ変名詞，置換コスト＝０．３

［置換知識２ｃ］　まし→／φ（φは空を表す記号），置換コスト＝０．２

ここで，単語は「表層表現／品詞」と表している。

【００５０】

ｉ１の端子に原文が入力されると、ステップＳ１では原文を形態素解析する。すると、原文２はその形態素解析結果として、図３（ａ）に示すように、単語数（これをｎと表す）１３個の単語に分割され、品詞が付与される。図３（ａ）において、上段の数字は単語間に文頭から順につけたノード番号である。このとき、文頭ノードの番号は０であり、文末ノードの番号は単語数ｎ（＝１３）である。

【００５１】

ｉ２の端子に要約率７０％が入力されると、ステップＳ２では、原文の文字数が２４文字であることから、削減すべき文字数の最小値（これをｍと表し、所望の文字削減数と呼ぶ）が次のように計算される。

【００５２】

［所望の文字削減数］

ｍ＝２４文字×（１００－７０）％＝７．２文字

ステップＳ３では、上記形態素解析結果に基づいて、先頭の単語列から順に置換知識２との照合を行い、照合に成功した場合には置換する単語列を形態素解析結果に追加し、その置換を付与した単語ラティス構造を作成する。すると、図３（ｂ）に示すような単語ラティス構造が得られる。なお、図３（ｂ）で、例えば、置換候補の「 首相」 の脇に記載した「 ０．０１」 は置換コストを表す。

【００５３】

ステップＳ４～Ｓ７では後向きの計算を行う。ステップＳ４，Ｓ６，Ｓ７で単語ラティス構造の文末ノード１３から文頭ノード０に向けてノード番号ｉをｉ＝１３，１２，…，０と１ずつ減少させて、各ノードｉにおいてステップＳ４を実行する。ステップＳ４では文末ノード１３から現在着目しているノードｉまでに最大可能な文字削減数（これをm_b（ｉ）と表し、後向き文字削減数と呼ぶ）を計算する。この際に、ダイナミックプログラミングにより効率的に処理を行う。すると、図３(c) に示すように、各ノードにおける後向き文字削減数が計算される。

【００５４】

図３（ｃ）において、例えば、ノード８での後向き文字削減数m_b（８）は、

ケース８ａ　置換知識２ｄ「まし→φ」と置換知識２ｅ「明らかに→表明」を適用．（文字削減数４）

ケース８ｂ　置換知識２ｃ「明らかにしました→表明」を適用．（文字削減数６）

と２つの場合が考えられるが、ケース８ｂが文字削減数が最も大きいので、

m_b（８）＝６と求められる。

【００５５】

ステップＳ８では文末ノードから文頭ノードまでの文字削減数の最大値m_b（０）＝１０が、所望の文字削減数ｍ＝７．２も小さくない（m_b（０）＝１０≧ｍ＝７．２）ので、ステップＳ９は実行せずに、ステップＳ１０を実行する。

【００５６】

ステップＳ１０～Ｓ１６では前向きの計算を行う。ステップＳ１０，Ｓ１５，Ｓ１６で単語ラティス構造の文頭ノード０から文末ノード１３に向けてノード番号ｉをｉ＝０，１，・・・，１３と１ずつ増加させて，各ノードｉにおけるすべてのパスにおいてステップＳ１１～Ｓ１４を実行する。ステップＳ１１では文頭から現在着目しているノードｉまでのそれぞれのパスにおける文字削減数（これをm_f（ｉ）と表し，前向き文字削減数と呼ぶ）を計算する。ノード５（ｉ＝５）の場合を例にとると、図４（ａ）に示すように、４つのパスが求められる。

【００５７】

次に、ステップＳ５で求めた後向き文字削減数m_b（ｉ）との和をとる。ノード５の例では、後向き文字削減数m_b（５）＝６との和をとると、図４（ｂ）に示すようになる。

【００５８】

図４（ｂ）に示すその和（m_f（ｉ）＋m_b（ｉ））が所望の文字削減数ｍよりも小さい場合（m_f（ｉ）＋m_b（ｉ）＜ｍ）には、ステップＳ１３を実行し、小さくない場合（m_f（ｉ）＋m_b（ｉ）≧ｍ）には、ステップＳ１４を実行する。ステップＳ１３ではそのパスを枝刈りする。ステップＳ１４ではそのパスの置換コストの和を求める。ノード５の例では、パス５ａは、前向き文字削減数と後向き文字削減数との和が所望の文字削減数（ｍ＝７．２）よりも小さいパスであるので、ステップＳ１３で枝刈りされ、以降の処理では使われない。パス５ｂ、５ｃ、５ｄはその和が所望の文字削減数（ｍ＝７．２）よりも小さくないので、ステップＳ１４が実行される。ステップＳ１４では、それぞれのパスにおける置換コストの和を求める。ノード５の例では図４（ｃ）に示すようになる。

【００５９】

同様にして、ステップＳ１４において文頭から文末までの各パスにおける置換コストの和が求められる。文末まで達すると、文頭から文末までのパスである、図５（ａ）に示すような、４つの要約候補が得られる。

【００６０】

ステップＳ１７ではその要約候補の中で、置換コストの和が最小であるパスｌ３ｄが選択され、単語列をつないで得られる次の要約文がｏ１の端子から出力される（図５（ｂ）参照）。

【００６１】

［要約文２］　「福沢首相は特別委で方針を表明した」

（他の実施の形態）

以上の説明では適用される置換知識がそれぞれ独立であるとした。しかし、「 総理大臣→首相」 のように、１度適用したら次回にも必ず適用しなければならない置換知識もある。これに対応するためには、パスごとに必須適用置換知識リストをもっておき、置換知識を適用する時にそのリストを参照する処理を図２のアルゴリズムに追加すればよい。

【００６２】

また、上述したステップＳ１０～Ｓ１６の前向き計算の処理中には、正例がない単語（例えば、図４（ａ）の「 で」 ）に達したときに、条件（前向き文字削減数、必須適用置換リスト等）の同じ候補は、置換コストが最小でないパス（例えば、図４（ａ）のパス５ｂ）も枝刈りするという改善を加えることも可能である。

【００６３】

なお、図１の形態素解析装置１、文字削減数最小値計算装置２等はＲＯＭ等を利用したモジュール回路（デバイス）のようなハードウェアで構成しても、形態素解析ルーチン、文字削減数最小値計算ルーチンのようにコンピュータ制御で動作するソフトウェアで構成し

てもよい。また、本発明は、複数の機器（例えば、ホストコンピュータ、インターフェース機器、リーダ、プリンタなど）から構成されるシステムに適用しても、１つの機器からなる専用装置（例えば、自動要約装置、自動字幕作成装置など）に適用してもよい。

【００６４】

また、本発明の目的は、前述した実施の形態の機能を実現するソフトウエアのプログラムコードを記録した記録媒体（記憶媒体）を、システムあるいは装置に供給し、そのシステムあるいは装置のコンピュータ（またはＣＰＵやＭＰＵ）が記録媒体に格納されたプログラムコードを読み出し、実行することによっても、達成されることは言うまでもない。この場合、記録媒体から読み出されたプログラムコード自体が前述した実施の形態の機能を実現することになり、そのプログラムコードを記録した記録媒体（例えば、ＣＤ－ＲＯＭ，ＭＤ，フロッピーなど）は本発明を構成することになる。

【００６５】

【発明の効果】

以上の説明から明らかなように、本発明によれば、文字削減数と置換コストという２つの評価関数を用いて、与えられた要約率以下で、原文を最適に要約するので、原文と要約率を入力するだけで、最適な要約を自動的に求めることが可能となる。

【図面の簡単な説明】

【図１】本発明の一実施形態の自動要約装置の構成を示すブロック図である。

【図２】図１の装置の自動要約の処理手順を示すフローチャートである。

【図３】具体的な原文を入力した場合の本発明の一実施形態の各段階の処理内容と結果を順次に説明する図であり、（ａ）形態素解析結果、（ｂ）単語ラティス構造、（ｃ）後向き文字削減数計算を具体例で示す。

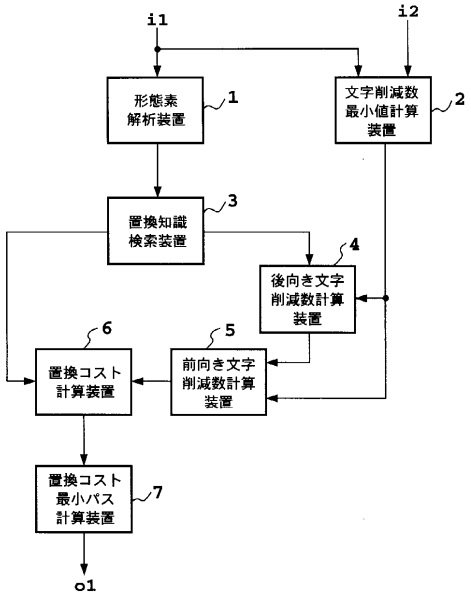
【図４】図３に連続する説明図であり、（ａ）前向き文字削減数の計算、（ｂ）前向き文字削減数と後向き文字削減数との和の計算、（ｃ）置換コストの和の計算を具体例で示す。

【図５】図４に連続する説明図であり、(a) 要約候補、（ｂ）要約文を具体例で示す。

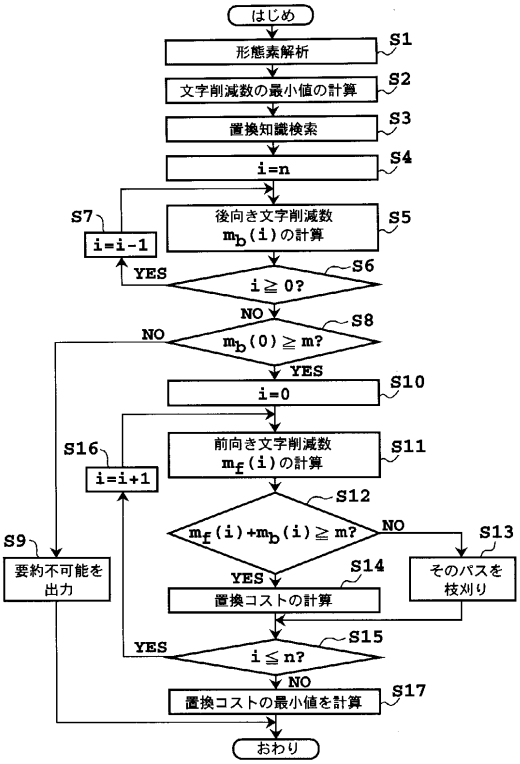
【符号の説明】

１　形態素解析装置
２　文字削減数最小値計算装置
３　置換知識検索装置
４　後向き文字削減数計算装置
５　前向き文字削減数計算装置
６　置換コスト計算装置
７　置換コスト最小パス計算装置

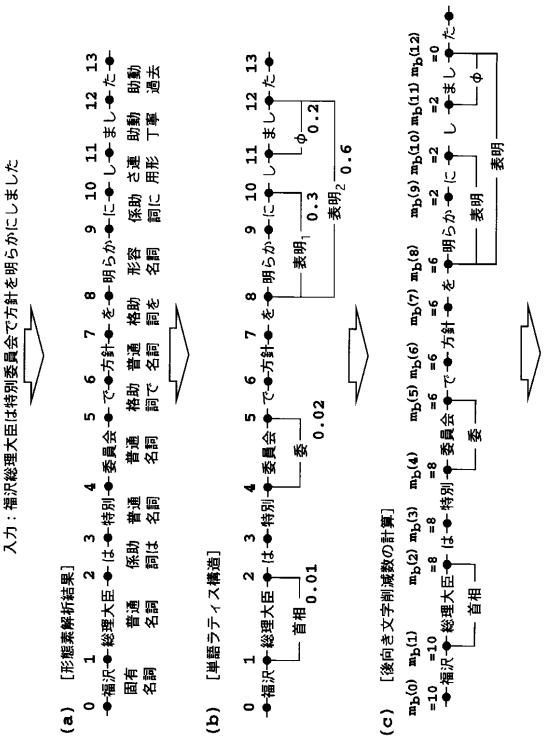
【図１】



【図２】



【図３】



【図４】

(a) ［前向き文字削減数の計算］

パス5a: 福沢→総理大臣→は→特別→委員会 $m_f(10)=0$
パス5b: 福沢→ 首相 →は→特別→委員会 $m_f(10)=2$
パス5c: 福沢→総理大臣→は→特別→ 委 $m_f(10)=2$
パス5d: 福沢→ 首相 →は→特別→ 委 $m_f(10)=4$

(b) ［前向き文字削減数と後向き文字削減数との和の計算］

パス5a $m_f(5)+m_b(5)=0+6=6<7.2(=m)$ ×
パス5b $m_f(5)+m_b(5)=2+6=8>7.2(=m)$ ○
パス5c $m_f(5)+m_b(5)=2+6=8>7.2(=m)$ ○
パス5d $m_f(5)+m_b(5)=4+6=10>7.2(=m)$ ○

(c) ［置換コストの和の計算］

パス5b 置換コストの和=0.01
パス5c 置換コストの和=0.02
パス5d 置換コストの和=0.01+0.02=0.03

【図５】

(a) ［要約候補］

パス13a 福沢→総理大臣→は→特別→委→で→方針→を表明$_2$
（文字削減数=8，置換コストの和=0.62）

パス13b 福沢→首相→は→特別→委員会→で→方針→を表明$_2$
（文字削減数=8，置換コストの和=0.61）

パス13c 福沢→首相→は→特別→委→で→方針→を表明$_2$
（文字削減数=10，置換コストの和=0.63）

パス13d 福沢→首相→は→特別→委→で→方針→を表明$_1$→し→た
（文字削減数=8，置換コストの和=0.53）

(b) ［要約文］

出力：福沢首相は特別委で方針を表明した

フロントページの続き

(56)参考文献 特開平８－２１２２２８（ＪＰ，Ａ）
特開平１０－６３６５８（ＪＰ，Ａ）
加藤 直人，ニュース文要約のための局所的要約知識獲得とその評価，情報処理学会研究報告，
日本，社団法人 情報処理学会，１９９８年 7月２４日，第９８巻 第６３号，６９～７６

(58)調査した分野(Int.Cl.，ＤＢ名)
G06F 17/30，
G06F 17/27