

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4706227号

(P4706227)

(45) 発行日 平成23年6月22日 (2011.6.22)

(24) 登録日 平成23年3月25日 (2011.3.25)

(51) Int. Cl. F I

G 0 6 F	17/30	(2006.01)	G 0 6 F	17/30	2 2 O A
G 0 6 F	17/21	(2006.01)	G 0 6 F	17/30	1 7 O A
G 0 6 F	17/27	(2006.01)	G 0 6 F	17/21	5 5 O A
			G 0 6 F	17/27	Z

請求項の数 12 (全 35 頁)

(21) 出願番号	特願2004-299833 (P2004-299833)	(73) 特許権者	000005496
(22) 出願日	平成16年10月14日 (2004.10.14)		富士ゼロックス株式会社
(65) 公開番号	特開2005-122743 (P2005-122743A)		東京都港区赤坂九丁目7番3号
(43) 公開日	平成17年5月12日 (2005.5.12)	(74) 代理人	100079049
審査請求日	平成19年10月11日 (2007.10.11)		弁理士 中島 淳
(31) 優先権主張番号	684508	(74) 代理人	100084995
(32) 優先日	平成15年10月15日 (2003.10.15)		弁理士 加藤 和詳
(33) 優先権主張国	米国 (US)	(72) 発明者	リビア ポラーニー
			アメリカ合衆国 94306 カリフォルニア州 パロ アルト ヴィラ ヴィスタ 4037

最終頁に続く

(54) 【発明の名称】 ハイブリッドテキスト要約を決定する方法、システム、及び、制御プログラム、談話構文解析方法、システム、及び、該システムにおける方法、談話構成要素にテキストを分割する方法及びシ

(57) 【特許請求の範囲】

【請求項 1】

分割規則を使用してテキストを談話構成要素に分割する分割手段と、

前記分割手段により分割された談話構成要素を、結合規則を使用して、談話構成要素を談話構成要素ノードとして談話構成要素ノードをエッジでリンクするツリー型の談話の構造表現に結合する談話構造表現決定手段と、

統計、キーワード、知識ベースの少なくとも一つを含む非構造方法による関連性基準に基づいて談話構成要素ノードの関連性スコアを決定する関連性スコア決定手段と、

前記談話構造表現決定手段により結合された談話の構造表現に従って前記関連性スコア決定手段により決定された関連性スコアを談話構成要素ノードに浸透する浸透手段であって、

関連性スコアに基づいて談話構成要素ノードの関連性が親の談話構成要素ノードの関連性よりある場合に、談話構成要素ノードの関連性スコアを親の談話構成要素ノードに割り当て、

関連性スコアに基づいて談話構成要素ノードの関連性が、該談話構成要素ノードを被従属談話構成要素ノードとしたとき従属する従属談話構成要素ノードの関連性よりある場合に、従属談話構成要素ノードに被従属談話構成要素ノードの関連性スコアを割り当て、

関連性スコアに基づいて等位の談話構成要素ノードの子の談話構成要素ノードのうち最も関連性がある談話構成要素ノードの関連性スコアを他の等位の談話構成要素ノードの子の談話構成要素ノードに割り当てる、

10

20

ことにより前記関連性スコアを談話構成要素ノードに浸透する、前記浸透手段と、
前記浸透手段により浸透された関連性スコアが閾値重要性スコアを超える談話構成要素
ノードを決定し、該決定された談話構成要素に基づいてテキストの要約を表示する表示手
段と、

を有するテキストを要約するシステム。

【請求項 2】

分割規則を使用してテキストを談話構成要素に分割する分割手段と、

前記分割手段により分割された談話構成要素を、結合規則を使用して、談話構成要素を
談話構成要素ノードとして談話構成要素ノードをエッジでリンクするツリー型の談話の構
造表現に結合する談話構造表現決定手段と、

統計、キーワード、知識ベースの少なくとも一つを含む非構造方法による関連性基準に
基づいて談話構成要素ノードの関連性スコアを決定する関連性スコア決定手段と、

前記談話構造表現決定手段により結合された談話の構造表現に従って前記関連性スコア
決定手段により決定された関連性スコアを談話構成要素ノードに浸透する浸透手段であっ
て、

関連性スコアに基づいて談話構成要素ノードの関連性が親の談話構成要素ノードの関連
性よりある場合に、談話構成要素ノードの関連性スコアを親の談話構成要素ノードに割り
当て、

関連性スコアに基づいて談話構成要素ノードの関連性が、該談話構成要素ノードを被従
属談話構成要素ノードとしたとき従属し且つ先行する従属談話構成要素ノードの関連性よ
りある場合に、該従属談話構成要素ノードに被従属談話構成要素ノードの関連性スコアを
割り当て、

関連性スコアに基づいて、等位の談話構成要素ノードの子の談話構成要素ノードのうち
最も関連性がある子の談話構成要素ノードの関連性スコアを、他の等位の談話構成要素ノ
ードの子の談話構成要素ノードのうちの該他の等位の談話構成要素ノードに先行する談話
構成要素ノードに割り当てる、

ことにより前記関連性スコアを談話構成要素ノードに浸透する、前記浸透手段と、
前記浸透手段により浸透された関連性スコアが閾値重要性スコアを超える談話構成要素
ノードを決定し、該決定された談話構成要素に基づいてテキストの要約を表示する表示手
段と、

を有するテキストを要約するシステム。

【請求項 3】

分割規則を使用してテキストを談話構成要素に分割する分割手段と、

前記分割手段により分割された談話構成要素を、結合規則を使用して、談話構成要素を
談話構成要素ノードとして談話構成要素ノードをエッジでリンクするツリー型の談話の構
造表現に結合する談話構造表現決定手段と、

統計、キーワード、知識ベースの少なくとも一つを含む非構造方法による関連性基準に
基づいて談話構成要素ノードの関連性スコアを決定する関連性スコア決定手段と、

前記談話構造表現決定手段により結合された談話の構造表現に従って前記関連性スコア
決定手段により決定された関連性スコアを談話構成要素ノードに浸透する浸透手段であっ
て、

談話構成要素ノードに、エッジ数 + 1 を従属レベルとして割り当て、

談話構成要素ノードのうち終端のノードにエッジ数 + 1 の逆数をスコアとして割り当て

、
関連性スコアに基づいて談話構成要素ノードの関連性が親の談話構成要素ノードの関連
性より少ない場合に、談話構成要素ノードのスコアを親の談話構成要素ノードに割り当て
、

該談話構成要素ノードを被従属談話構成要素ノードとしたとき従属する従属談話構成要
素ノードより被従属談話構成要素ノードのスコアが低い場合に、被従属談話構成要素ノ
ードのスコアを従属談話構成要素ノードに割り当て、

10

20

30

40

50

等位の談話構成要素ノードの子の談話構成要素ノードのスコアが等位の談話構成要素ノードのスコアより低くない場合に、等位の談話構成要素ノードのスコアを子の談話構成要素ノードに割り当て、

談話構成要素ノードのスコア及び従属レベルに基づいて、談話構成要素ノードの関連性スコアを調整する、

ことにより前記関連性スコアを談話構成要素ノードに浸透する、前記浸透手段と、

前記浸透手段により浸透された関連性スコアが閾値重要性スコアを超える談話構成要素ノードを決定し、該決定された談話構成要素に基づいてテキストの要約を表示する表示手段と、

を有するテキストを要約するシステム。

10

【請求項 4】

前記分割規則が、言語談話モデル、統一言語談話モデル、修辞構造理論、談話構造理論及び構造化談話表現理論の少なくとも一つである、請求項 1 ~ 3 のいずれかに記載のテキストを要約するシステム。

【請求項 5】

前記浸透手段が、さらに、

照応を含む談話構成要素の談話構成要素ノードを決定し、

前記決定された談話構成要素ノードに先行する談話構成要素ノードの談話構成要素から前記照応の対象を決定して、該照応の対象を前記照応に代入する、

ことにより前記関連性スコアを談話構成要素ノードに浸透する

請求項 1 ~ 3 のいずれかに記載のテキストを要約するシステム。

20

【請求項 6】

前記浸透手段が、さらに、

関連性スコアに基づいて重要談話構成要素ノードを識別し、

前記識別された重要談話構成要素ノードの談話構成要素に含まれる照応を決定し、

前記決定された照応の対象を含む可能性のある前記重要談話構成要素ノードより先行する先行談話構成要素ノードを所定の方法により決定し、

前記決定された先行談話構成要素ノードに前記重要談話構成要素ノードの関連性スコアを浸透し、

前記浸透された関連性スコアに基づいて、テキストの要約に使用される談話構成要素ノードのサブセットを決定する、

30

ことにより前記関連性スコアを談話構成要素ノードに浸透する

請求項 1 ~ 3 のいずれかに記載のテキストを要約するシステム。

【請求項 7】

コンピュータに、

テキストを談話構成要素に分割規則を使用して分割するステップと、

結合規則を使用して、前記分割された談話構成要素を、談話構成要素を談話構成要素ノードとして談話構成要素ノードをエッジでリンクするツリー型の談話の構造表現に結合するステップと、

統計、キーワード、知識ベースの少なくとも一つを含む非構造方法による関連性基準に基づいて談話構成要素ノードの関連性スコアを決定するステップと、

40

前記談話の構造表現に従って前記決定された関連性スコアを談話構成要素ノードに浸透するステップであって、

関連性スコアに基づいて談話構成要素ノードの関連性が親の談話構成要素ノードの関連性よりある場合に、談話構成要素ノードの関連性スコアを親の談話構成要素ノードに割り当てるステップと、

関連性スコアに基づいて談話構成要素ノードの関連性が、該談話構成要素ノードを被従属談話構成要素ノードとしたとき従属する従属談話構成要素ノードの関連性よりある場合に、従属談話構成要素ノードに被従属談話構成要素ノードの関連性スコアを割り当てるステップと、

50

関連性スコアに基づいて等位の談話構成要素ノードの子の談話構成要素ノードのうち最も関連性がある談話構成要素ノードの関連性スコアを他の等位の談話構成要素ノードの子の談話構成要素ノードに割り当てるステップと、

を有する、前記関連性スコアを談話構成要素ノードに浸透するステップと、

前記浸透された関連性スコアが閾値重要性スコアを超える談話構成要素ノードを決定し、該決定された談話構成要素に基づいてテキストの要約を表示するステップと、

を実行させるためのプログラム。

【請求項 8】

コンピュータに、

分割規則を使用してテキストを談話構成要素に分割するステップと、

前記分割された談話構成要素を、結合規則を使用して、談話構成要素を談話構成要素ノードとして談話構成要素ノードをエッジでリンクするツリー型の談話の構造表現に結合するステップと、

統計、キーワード、知識ベースの少なくとも一つを含む非構造方法による関連性基準に基づいて談話構成要素ノードの関連性スコアを決定するステップと、

前記談話の構造表現に従って前記決定された関連性スコアを談話構成要素ノードに浸透するステップであって、

関連性スコアに基づいて談話構成要素ノードの関連性が親の談話構成要素ノードの関連性よりある場合に、談話構成要素ノードの関連性スコアを親の談話構成要素ノードに割り当てるステップと、

関連性スコアに基づいて談話構成要素ノードの関連性が、該談話構成要素ノードを被従属談話構成要素ノードとしたとき従属し且つ先行する従属談話構成要素ノードの関連性よりある場合に、該従属談話構成要素ノードに被従属談話構成要素ノードの関連性スコアを割り当てるステップと、

関連性スコアに基づいて、等位の談話構成要素ノードの子の談話構成要素ノードのうち最も関連性がある子の談話構成要素ノードの関連性スコアを、他の等位の談話構成要素ノードの子の談話構成要素ノードのうちの該他の等位の談話構成要素ノードに先行する談話構成要素ノードに割り当てるステップと、

を有する、前記関連性スコアを談話構成要素ノードに浸透するステップと、

前記浸透された関連性スコアが閾値重要性スコアを超える談話構成要素ノードを決定し、該決定された談話構成要素に基づいてテキストの要約を表示するステップと、

を実行させるためのプログラム。

【請求項 9】

コンピュータに、

分割規則を使用してテキストを談話構成要素に分割するステップと、

前記分割された談話構成要素を、結合規則を使用して、談話構成要素を談話構成要素ノードとして談話構成要素ノードをエッジでリンクするツリー型の談話の構造表現に結合するステップと、

統計、キーワード、知識ベースの少なくとも一つを含む非構造方法による関連性基準に基づいて談話構成要素ノードの関連性スコアを決定するステップと、

前記談話の構造表現に従って前記決定された関連性スコアを談話構成要素ノードに浸透するステップであって、

談話構成要素ノードに、エッジ数 + 1 を従属レベルとして割り当てるステップと、

談話構成要素ノードのうち終端のノードにエッジ数 + 1 の逆数をスコアとして割り当てるステップと、

関連性スコアに基づいて談話構成要素ノードの関連性が親の談話構成要素ノードの関連性より少ない場合に、談話構成要素ノードのスコアを親の談話構成要素ノードに割り当てるステップと、

該談話構成要素ノードを被従属談話構成要素ノードとしたとき従属する従属談話構成要素ノードより被従属談話構成要素ノードのスコアが低い場合に、被従属談話構成要素ノード

10

20

30

40

50

ドのスコアを従属談話構成要素ノードに割り当てるステップと、

等位の談話構成要素ノードの子の談話構成要素ノードのスコアが等位の談話構成要素ノードのスコアより低くない場合に、等位の談話構成要素ノードのスコアを子の談話構成要素ノードに割り当てるステップと、

談話構成要素ノードのスコア及び従属レベルに基づいて、談話構成要素ノードの関連性スコアを調整するステップと、

を有する、前記関連性スコアを談話構成要素ノードに浸透するステップと、

前記浸透された関連性スコアが閾値重要性スコアを超える談話構成要素ノードを決定し、該決定された談話構成要素に基づいてテキストの要約を表示するステップと、

を実行させるためのプログラム。

10

【請求項 10】

前記分割規則が、言語談話モデル、統一言語談話モデル、修辞構造理論、談話構造理論及び構造化談話表現理論の少なくとも一つである、請求項 7～9 のいずれかに記載のプログラム。

【請求項 11】

前記関連性スコアを談話構成要素ノードに浸透するステップが、

照応を含む談話構成要素の談話構成要素ノードを決定するステップと、

前記決定された談話構成要素ノードに先行する談話構成要素ノードの談話構成要素から前記照応の対象を決定して、該照応の対象を前記照応に代入するステップと、

をさらに有する請求項 7～9 のいずれかに記載のプログラム。

20

【請求項 12】

前記関連性スコアを談話構成要素ノードに浸透するステップが、

関連性スコアに基づいて重要談話構成要素ノードを識別するステップと、

前記識別された重要談話構成要素ノードの談話構成要素に含まれる照応を決定するステップと、

前記決定された照応の対象を含む可能性のある前記重要談話構成要素ノードより先行する先行談話構成要素ノードを所定の方法により決定するステップと、

前記決定された先行談話構成要素ノードに前記重要談話構成要素ノードの関連性スコアを浸透するステップと、

前記浸透された関連性スコアに基づいて、テキストの要約に使用される談話構成要素ノードのサブセットを決定するステップと、

を有する請求項 7～9 のいずれかに記載のプログラム。

30

【発明の詳細な説明】

【技術分野】

【0001】

本発明はテキスト情報の要約を決定することに関する。詳細には、ハイブリッドテキスト要約を決定する方法、システム、及び、制御プログラム、談話構文解析方法、システム、及び、該システムにおける方法、談話構成要素にテキストを分割する方法及びシステム、談話の構造表現を決定する方法及びシステム、ハイブリッドテキスト要約システムに関する。

40

【背景技術】

【0002】

企業、政府政策者、情報分析者、他の研究者及び知識労働者は増え続ける大量の情報を検討し、理解することに直面している。最も関連性のある情報に基づき要約を作成することにより情報量の削減を試みた研究者もいた。関連性の基準は、多くの場合、キーワード及び/又は用語頻度、並びに情報リポジトリ内の文書の集合についての文書逆頻度統計等の非構造特徴に基づいている。しかしながら、これらの従来の非構造テキスト要約は頻繁に解体され、読むのが困難である。さらに、非構造テキスト要約のために選択される要素は統計学上の重要度の基準に基づいている。したがって、順次選択されたテキスト要約要素の代名詞は、これらの従来の要約の有用性を削減する省かれた先行する参考文献を参照

50

することがある。

【 0 0 0 3 】

構造に基づいた要約からあまり重要ではない情報を選択的に削除することによりテキスト要約の質の改善を試みた他の研究者もいた。これらの従来の構造に基づいたテキスト要約システムでは、テキストの構造表現が決定される。あまり重要ではない情報は、テキスト要約に含まれる所望される従属のレベルを選択することにより構造表現から取り除かれる。これらの従来の構造テキスト要約は対象情報を含むことによって文法性を失わないようにする。しかしながら、結果として生じるテキスト要約は多くの場合、前記対象を解決するために必要とされる以上の情報を含んでいる。

【 0 0 0 4 】

例えば、特許文献 1、2 及び非特許文献 1 は、本願に関連する技術について記載している。詳細には、特許文献 1 は、コンピュータ解析支援を用いた読解技能向上システムについて記載し、特許文献 2 は、ワード・プロセッサのための文書要約に関する技術について記載している。

【特許文献 1】米国特許出願公開第 2 0 0 4 / 0 0 6 7 4 7 2 A 1 号明細書

【特許文献 2】米国特許第 6、3 4 9、3 1 6 B 2 号明細書

【非特許文献 1】E. ホヴィ (E. Hovy)、「談話構造関係を用いた自動談話生成 (Automated Discourse Generation Using Discourse Structure Relations)」、人工知能 (Artificial Intelligence)、アムステルダム、1 9 9 3 年、巻 6 3、頁 3 4 1 ~ 3 8 5

【発明の開示】

【発明が解決しようとする課題】

【 0 0 0 5 】

ハイブリッドテキスト要約を決定する上で構造情報及び非構造情報の組み合わせを可能にするシステム及び方法が有用となるであろう。本発明のシステム及び方法による多様な例示的な実施形態は、文法的なハイブリッドテキスト要約を決定する際の構造情報及び非構造情報の組み合わせを可能にする。テキストは談話構成要素を決定するために談話分析の理論の分割規則に基づいて分割される。決定された談話構成要素は、談話分析の理論の結合規則に基づいて談話の構造表現に結合される。関連性スコアは、統計、キーワード、知識ベース、又は任意の既知の又は今後開発される、関連性を決定する非構造方法に基づいて決定される。関連性スコアは談話の構造表現中の構造的に関連する談話構成要素ノードに浸透されている。選択可能及び/又は調整可能な重要度スコアは、談話の構造表現から談話構成要素のサブツリーを選択するために使用される。浸透方法及び/又は関連性スコアに対する調整は、談話構成要素の前記サブツリーの圧縮を調整し、ハイブリッドテキスト要約を形成するために使用される。テキストを基本的な談話構成要素に分割し、談話構成要素を談話分析の理論に基づいて分類するシステム及び方法が提供される。各種の決定された談話構成要素を談話の構造表現に結合させるシステム及び方法も提供される。

【課題を解決するための手段】

【 0 0 0 6 】

本発明の第 1 の態様のハイブリッドテキスト要約を決定する方法は、テキストの談話構成要素を決定するステップと、テキストの談話の構造表現を決定するステップと、統計、キーワード、知識ベースの少なくとも一つに基づいて決定される少なくとも一つの関連性の非構造基準に基づいて談話構成要素の関連性スコアを決定するステップと、談話の構造表現に基づいて関連性スコアを浸透するステップと、閾値重要性スコアと比較した関連性スコアにより、談話構成要素に基づいてハイブリッドテキスト要約を決定するステップと、を有する。

【 0 0 0 7 】

本発明の第 2 の態様は、第 1 の態様のハイブリッドテキスト要約を決定する方法であって、前記談話分析の理論が、言語談話モデル、統一言語談話モデル、修辞構造理論、談話構造理論及び構造化談話表現理論の少なくとも一つである。

【 0 0 0 8 】

本発明の第3の態様は、第1の態様のハイブリッドテキスト要約を決定する方法であって、前記関連性スコアを浸透するステップが、子の談話構成要素ノードがより関連性のある場合に、構造表現中の子の談話構成要素ノードごとに、子の談話構成要素ノードの関連性スコアを親の談話構成要素ノードに割り当てるステップと、被従属談話構成要素がより関連性がある場合に、任意の従属ノードについて、従属談話構成要素に被従属談話構成要素の関連性スコアを割り当てるステップと、任意の等位ノードについて、最も関連性のある子の関連性スコアを他の子の談話構成要素ノードに割り当てるステップと、を有する。

【0009】

本発明の第4の態様は、第1の態様のハイブリッドテキスト要約を決定する方法であって、前記関連性スコアを浸透するステップが、子の談話構成要素ノードが親より関連性がある場合に、構造表現中の子の談話構成要素ノードごとに、親の談話構成要素ノードに子の談話構成要素ノードの関連性スコアを割り当てるステップと、等位にされた談話構成要素ノードごとに、等位にされた談話構成要素ノードの関連性スコアをそれぞれの先行より関連性がない同位のノードに割り当てるステップと、等位にされた談話構成要素ノードではなく、被従属談話構成要素ノードではない子の談話構成要素ノードごとに、親の談話構成要素ノードが子より関連性がある場合に、親の談話構成要素ノードの関連性スコアを子の談話構成要素ノードに割り当てるステップと、等位にされた談話構成要素ノードごとに、等位にされた談話ノード及びすべての同位が親のノードより関連性がない場合に、親の談話構成要素ノードの関連性スコアを等位にされた談話構成要素ノードに割り当てるステップと、被従属談話構成要素が従属ノードより関連性がある場合に、被従属談話構成要素ノードごとに、被従属談話構成要素ノードの関連性スコアを従属談話構成要素に割り当てるステップと、ノードごとに、その関連性スコアが別のノードの関連性スコアに変更されるノードが見つからなくなるまで前記ステップを繰り返すステップと、を有する。

【0010】

本発明の第5の態様は、ハイブリッドテキスト要約を決定する方法であって、テキストの談話構成要素を決定するステップと、テキストの談話の構造表現を決定するステップと、談話構成要素の関連性スコアを決定するステップと、談話の構造表現に基づいて関連性スコアを浸透するステップと、閾値重要性スコアに比較した関連性スコアにより談話構成要素に基づいてハイブリッドテキスト要約を決定するステップと、を有し、前記関連性スコアを浸透するステップは、談話構成要素リーフノードごとに、被従属エッジに1を加えた数を決定するステップと、被従属エッジ+1の逆数に基づいてスコアを決定するステップと、スコアの関連性がより少ない場合に、談話構成要素ノードごとに、子の談話構成要素ノードのスコアを親の談話構成要素ノードに割り当てるステップと、被従属談話構成要素スコアがより低い場合に、任意の従属談話構成要素ノードについて、被従属談話構成要素ノードのスコアを従属談話構成要素ノードに割り当てるステップと、任意の等位構造談話構成要素ノードの関連性スコアを、より低い場合に、子の談話構成要素の等位のそれぞれに割り当てるステップと、スコア及び従属レベルに基づいて、調整された関連性スコアを決定するステップと、を有する。

【0011】

本発明の第6の態様は、第1の態様のハイブリッドテキスト要約を決定する方法であって、照応を含むあらゆるリーフ談話構成要素を決定するステップと、照応ごとに、照応のユニークな先行する対象を決定するステップと、照応について、リーフ談話構成要素にユニークな先行する対象を代入するステップと、閾値重要性スコアよりさらに関連性のある関連性スコアにより、談話構成要素の集合からユニークな先行する対象を含む談話構成要素を削除するステップと、をさらに有する。

【0012】

本発明の第7の態様は、第1の態様のハイブリッドテキスト要約を決定する方法であって、関連性スコアを浸透するステップが、重要談話構成要素ノードを決定するステップと、未解決照応を決定するステップと、照応を解決する可能性のある潜在的な先行する対象を含む潜在的な解決する談話構成要素を決定するステップと、潜在的な解決する談話構成

10

20

30

40

50

要素の縮小された範囲を通して、重要談話構成要素の関連性スコアを浸透するステップと、関連性スコアに基づいて、談話構成要素の縮小された範囲を決定するステップと、を有する。

【0013】

本発明の第8の態様は、ハイブリッドテキスト要約を決定するシステムであって、テキストを取り出す入力/出力回路と、テキストの談話構成要素を決定し、談話の構造表現に談話構成要素を結合するプロセッサと、キーワード、知識ベースの少なくとも一つに基づいて決定される少なくとも一つの関連性の非構造基準に基づいて、談話構成要素の関連性スコアを決定する関連性スコア決定回路と、談話の構造表現に基づいて、談話構成要素関連性スコアを浸透する浸透回路と、を備え、前記プロセッサが、閾値重要性スコアを超え 10
る関連性スコアにより談話構成要素に基づいて、ハイブリッドテキスト要約を決定する。

【0014】

本発明の第9の態様は、第8の態様のハイブリッドテキスト要約を決定するシステムであって、前記談話分析の理論が、言語談話モデル、統一言語談話モデル、修辞構造理論、談話構造理論及び構造化談話表現理論の少なくとも一つである。

【0015】

本発明の第10の態様は、第8の態様のハイブリッドテキスト要約を決定するシステムであって、前記浸透回路が、子の談話構成要素ノードがより関連性がある場合に、構造表現中の子の談話構成要素ノードごとに、子の談話構成要素ノードの関連性スコアを親の談話構成要素ノードに割り当て、前記浸透回路が、被従属談話構成要素がより関連性がある 20
場合に、任意の従属ノードについて、従属談話構成要素に被従属談話構成要素の関連性スコアを割り当て、前記浸透回路が、任意の等位ノードについて、最も関連性のある子の関連性スコアを他の子の談話構成要素ノードに割り当てる。

【0016】

本発明の第11の態様は、第8の態様のハイブリッド要約を決定するシステムであって、前記浸透回路が、子の談話構成要素ノードが親より関連性がある場合に、構造表現中の子の談話構成要素ごとに、子の談話構成要素ノードの関連性スコアを親の談話構成要素ノードに割り当て、前記浸透回路が、等位にされた談話構成要素ノードごとに、それぞれの先行より関連性がない同位のノードに、等位にされた談話構成要素ノードの関連性スコアを割り当て、前記浸透回路が、等位にされた談話構成要素ノードではなく、被従属談話構成要素ノードではない子の談話構成要素ノードごとに、親の談話構成要素が子より関連性 30
がある場合に、子の談話構成要素ノードに親の談話構成要素ノードの関連性スコアを割り当て、前記浸透回路が、等位にされた談話ノード及びすべての同位が親のノードより関連性が少ない場合に、等位にされた談話構成要素ノードごとに、親の談話構成要素ノードの関連性スコアを等位にされた談話構成要素ノードに割り当て、前記浸透回路が、被従属談話構成要素が従属ノードより関連性がある場合に、被従属談話構成要素ノードごとに、被従属談話構成要素ノードの関連性スコアを従属談話構成要素に割り当て、前記浸透回路が、関連性スコアが別のノードの関連性スコアに変更されるノードを見つけられなくなるまで、ノードごとに、前記ステップを繰り返す。

【0017】

本発明の第12の態様は、ハイブリッドテキスト要約を決定するシステムであって、テキストを取り出す入力/出力回路と、テキストの談話構成要素を決定し、談話構成要素を談話の構造表現の中に結合するプロセッサと、少なくとも一つの関連性の非構造基準に基づいて、談話構成要素の関連性スコアを決定する関連性スコア決定回路と、談話の構造表現に基づいて、談話構成要素関連性スコアを浸透する浸透回路と、を有し、前記浸透回路が、談話構成要素リーフノードごとに、被従属エッジに1を加えた数の逆数に基づいて、被従属エッジに1を加えた数及びスコアを決定し、前記浸透回路が、スコアの関連性がより少ない場合に、談話構成要素ノードごとに、子の談話構成要素ノードのスコアを親の談話構成要素に割り当て、前記浸透回路が、被従属談話構成要素スコアがより低い場合に、 50
任意の従属談話構成要素ノードについて、被従属談話構成要素ノードのスコアを従属談話

構成要素ノードに割り当て、前記浸透回路が、より低い場合に、等位の子の談話構成要素のそれぞれに任意の等位談話構成要素ノードのスコアを割り当て、前記プロセッサが、スコア及び従属レベルに基づき調整された関連性スコアを決定し、ハイブリッドテキスト要約が閾値重要性スコアを超える関連性スコアにより談話構成要素に基づいて決定される。

【0018】

本発明の第13の態様は、第8の態様のハイブリッド要約を決定するシステムであって、前記プロセッサが、照応を含むあらゆるリーフ談話構成要素を決定し、前記プロセッサが、照応ごとに、照応の任意のユニークな先行する対象を決定し、前記プロセッサが、ユニークな先行する対象を、照応対象の先行する談話構成要素に代入し、前記プロセッサが、閾値重要性スコアを超える関連性スコアにより談話構成要素からユニークな対象を含む先行する談話を削除する。

10

【0019】

本発明の第14の態様は、第8の態様のハイブリッド要約を決定するシステムであって、前記浸透回路が、照応を含むあらゆるリーフ談話構成要素を決定し、前記浸透回路が、照応ごとに、照応の任意のユニークな先行する対象を決定し、前記浸透回路が、ユニークな先行する対象を照応のリーフ談話構成要素に代入し、前記浸透回路が、より関連性のある関連性スコアにより談話構成要素の集合からユニークな先行する対象を含む談話構成要素を削除する。

【0020】

本発明の第15の態様は、第8の態様のハイブリッド要約を決定するシステムであって、前記プロセッサが、関連性の非構造基準に基づいて重要談話構成要素ノードを決定し、重要談話構成要素の中の未解決対象を決定し、対象を解決する可能性がある潜在的な解決する談話構成要素を決定し、潜在的な解決する談話構成要素の縮小された範囲を通して重要談話構成要素の関連性スコアを浸透し、関連性スコアに基づいて談話構成要素の縮小された範囲を決定する。

20

【0021】

本発明の第16の態様は、ハイブリッドテキスト要約を決定するためにコンピュータをプログラミングするために使用可能な制御プログラムであって、テキストの談話構成要素を決定する命令と、テキストの談話の構造表現を決定する命令と、キーワード、知識ベースの少なくとも一つに基づいて決定される少なくとも一つの関連性の非構造基準に基づいて、談話構成要素の関連性スコアを決定する命令と、談話の構造表現に基づいて関連性スコアを浸透する命令と、閾値重要性スコアと比較した関連性スコアにより談話構成要素に基づいて、ハイブリッドテキスト要約を決定する命令と、を有する。

30

【0022】

本発明の第17の態様は、談話構成要素にテキストを分割する方法であって、談話分析の理論を決定するステップと、候補セグメントを決定するステップと、談話を続行することに関連付けられる候補セグメントの品詞 (p a r t - o f - s p e e c h) タグ、蓋然性構文解析ツール、統計構文解析ツール、有限状態構文解析ツール、記号構文解析ツール、語彙目録及びW o r d N e t 関係の少なくとも一つに基づいて決定される属性を決定するステップと、を有する。

40

【0023】

本発明の第18の態様は、構造表現を決定する方法であって、テキストの談話構成要素を決定するステップと、談話構成要素の談話分析分類の理論、及び統語的な制限、意味論的な制限及び語彙的・意味論的な制限の少なくとも一つに基づいて談話の構造表現に談話構成要素を結合するステップと、を有する。

【0024】

本発明の第19の態様は、ハイブリッドテキスト要約システムであって、テキストの談話構成要素を決定する手段と、テキストの談話の構造表現を決定する手段と、キーワード、知識ベースの少なくとも一つに基づいて決定される少なくとも一つの関連性の非構造基準に基づいて談話構成要素の関連性スコアを決定する手段と、談話の構造表現に基づいて

50

関連性スコアを浸透する手段と、閾値重要性スコアと比較した関連性スコアにより談話構成要素に基づいてハイブリッドテキスト要約を決定する手段と、を有する。

【0025】

本発明の第20の態様は、ハイブリッドテキスト要約システムであって、テキストの談話構成要素を決定する手段と、テキストの談話の構造表現を決定する手段と、談話構成要素の関連性スコアを決定する手段と、談話の構造表現に基づいて関連性スコアを浸透する手段と、閾値重要性スコアと比較した関連性スコアにより談話構成要素に基づいてハイブリッドテキスト要約を決定するための手段と、を有し、前記関連性スコアを浸透する手段は、談話構成要素リーフノードごとに、被従属エッジに1を加えた数を決定し、被従属エッジ+1の逆数に基づいてスコアを決定し、スコアの関連性がより少ない場合に、談話構成要素ノードごとに、親の談話構成要素ノードに子の談話構成要素ノードのスコアを割り当て、被従属談話構成要素スコアがより低い場合に、任意の従属談話構成要素ノードについて、従属談話構成要素ノードに被従属談話構成要素ノードのスコアを割り当て、等位の子の談話構成要素のそれぞれに任意の等位談話構成要素ノードの関連性スコアを、より低い場合に、割り当て、スコア及び従属レベルに基づいて調整された関連性スコアを決定する。

10

【発明を実施するための最良の形態】

【0026】

図1は、本発明による例示的なハイブリッド要約システム100の概要である。ハイブリッド要約システム100は、ウェブ使用可能なパーソナルコンピュータ300、ウェブ使用可能なタブレットコンピュータ400、及び文書10000~10002を含む情報リポジトリ200に通信リンク99を介して接続されている。

20

【0027】

本発明による多様な例示的な実施形態の一つでは、ウェブ使用可能なパーソナルコンピュータ300が情報リポジトリ200に含まれている文書10000のハイブリッドテキスト要約に対する要求を開始する。前記要求は通信リンク99を介してハイブリッド要約システム100に転送される。ハイブリッド要約システム100はハイブリッドテキスト要約に対する要求を取り次ぐ。次に、ハイブリッド要約システム100は情報リポジトリ200から文書10000を取り出す。それから、テキスト文書10000は談話分析の理論のための決定された分割規則に基づき談話構成要素に分割される。

30

【0028】

本発明による多様な例示的な実施形態では、分割規則はテキストを基本的な談話構成要素単位(BDU)に分割するために使用される。基本談話構成要素単位は、統一言語談話モデル(ULDM)、修辞構造理論(RST)、談話構造理論、構造化談話表現理論(SDRT)又は談話分析の他の既知の、あるいは今後開発される理論について作成される分割規則に基づいて決定される。例えば、言語談話モデルのそれぞれの基本談話構成要素は、1)多くとも一つの考えられる世界の情勢についての情報、及び、2)先に発生した言語学的なジェスチャ、あるいはおそらく次の言語学的なジェスチャが、談話又は談話が発生する文脈の中の他の単位に構造上、意味論上、インタラクティブに、あるいは修辞学的にどのように関係するのかについての情報を符号化する関数情報、を伝達する文脈を基準にして解釈される言語学上のジェスチャから構成されている。言語学的に実現された基本談話構成要素単位の文の統語的な反射作用は、それが見つけられる文の外部からの連続のために利用しやすいセグメントである。事象に対する参照は必要であるが、ある単語が大部分の談話構成要素の先頭になるためには十分ではない。したがって、助動詞及び法動詞はそれらの本動詞から別個の談話構成要素を形成しないであろう。それらは別個の事象を参照しないため、「しなければならぬ」、「すべきである」等の不定詞を含む他の法助動詞構造も同じ論法によって、それらの補語と単一のセグメントを構成する。

40

【0029】

このようにして、談話構成要素単位を決定するために、候補談話構成要素セグメントは事象を参照し、談話の中で後に事象を「続行する」ために十分な文脈を提供しなければな

50

らない。「試す」又は「説得する」などの同一動詞及び「と思われる」及び「信じる」などの繰り上げ動詞は、両方の事象とも続行できるためそれらの動詞補語から別個の文法項を形成する。対照的に、動名詞を含む事象名詞語句は、それらが文法項又は修飾語句である動詞とおそらく識別可能である事象を参照しても、それらの事象は容易に続行しない。

【0030】

統一言語談話モデルは決定された基本的な談話構成要素を結合規則に基づいて談話の構造表現に連結する。結合規則は、基本談話構成要素が談話の右に開いたツリー構造表現にどのように連結されるのかを突き止めるために使用される。

【0031】

例えば、統一言語談話モデルは、基本談話構成要素を右に開いたツリーに基づいて談話の構造表現に連結するために単独で、あるいは組み合わせて使用できる統語的、意味論的、語彙的 - 意味論的、構造上及びデフォルトの結合規則を提供する。統計的なキーワード及び/又は重要談話構成要素を決定する他の既知の、あるいは今後開発される方法は、連結された談話構成要素のサブツリーを定義するために使用される。追加の支持する談話構成要素は、談話分析の理論に関連付けられる結合規則に基づいて含まれる。本発明による多様な他の例示的な実施形態において、結合又は連結支持情報が統計的なキーワード及び/又は他の関連性情報と、各談話構成要素と関連付けられた単一の関連性スコアに合体されてよいことが明らかとなるであろう。ハイブリッド要約システムがハイブリッドテキスト要約を決定した後で、ハイブリッドテキスト要約はウェブ使用可能なパーソナルコンピュータ300に転送される。

【0032】

本発明による別の例示的な実施形態では、ウェブ使用可能なタブレットコンピュータ400のユーザが情報リポジトリ200の中の文書10001のハイブリッドテキスト要約を要求する。前記要求は通信リンク99を介してハイブリッド要約システム100に転送される。ハイブリッド要約システム100は通信リンク99を介して情報リポジトリ200から文書10001を取り出す。談話の構造表現が、決定される談話構成要素間のリンクを指定する談話分析の理論に基づき要求された文書10001について決定される。テキスト中の重要談話構成要素を突き止めるために、統計的な分析、キーワード、知識ベースの検索又は重要な情報を決定する任意の既知のあるいは今後開発される方法が使用されてよい。例えば、統計的な及び/又はキーワードに基づいた技法は、閾値重要度スコアを超える関連性スコアを有する談話構成要素を決定するために適用される。決定された重要談話構成要素及び談話の構造表現中の支持する先行談話構成要素が、談話構成要素のサブツリーを定義する。

【0033】

本発明による多様な例示的な実施形態では、談話構成要素リーフノードの関連性スコアが他の談話構成要素の関連性スコアを調整するために談話の構造表現を通して浸透される。関連性スコアの浸透は、変化する数の支持する利用しやすい談話構成要素を含むために談話分析の理論に基づいて調整されてよい。異なる浸透方法がテキストの異なるジャンル、テキストの異なる言語又は任意の属性、あるいは談話の構造表現に影響を及ぼす可能性のある文書に基づいて指定されてよい。

【0034】

談話分析の他の理論は、情報が構造にどのように符号化されるのかの詳細で異なる可能性があるが、先行の支持する談話構成要素の識別及び関連性スコアの構造に基づいた浸透は、文法性を保つ任意のテキスト要約機能とともに使用されてよい。例えば談話の構造表現が統一言語談話モデル(ULDM)に基づいている場合、リーフノードとして重要談話構成要素を有する談話の構造表現からの関連する談話構成要素の縮小されたサブツリーがハイブリッドテキスト要約として決定される。

【0035】

図2は、本発明の例示的な実施形態によるハイブリッドテキスト要約を決定する第1の方法である。プロセスはステップS100で開始し、談話分析の理論が決定されるステッ

10

20

30

40

50

プ S 2 0 0 に続行する。制御は次にステップ S 3 0 0 に続行する。

【 0 0 3 6 】

ステップ S 3 0 0 では、要約されるテキストが決定される。テキストは、ドロップダウンダイアログボックスから先に作成されたテキストを選択することにより決定され、エディタ又はワープロを使用して入力され、ウェブサーバ又はデジタルライブラリなどの情報リポジトリから取り出され、あるいは、既知のあるいは今後開発されるテキスト選択方法を使用して決定されてよい。テキストが決定された後、制御は、テキストが次に談話分析の理論の分割規則に基づいて談話構成要素に分割されるステップ S 4 0 0 に続行する。

【 0 0 3 7 】

テキストは、談話構成要素のグラフを決定する談話分析の任意の理論に基づいて分割されてよい。例えば、言語談話モデル及び統一言語談話モデルはそれぞれテキスト内の談話構成要素を決定するための分割規則を提供する。統一言語談話モデルの分割規則は、それぞれの識別された談話構成要素が、等位接続、従位接続、及び N 分 (N - a r y) ノード又は二分ノードに基づいて右に開いたツリーに結合できるようにテキストを分割する。しかしながら、本発明の多様な他の例示的な実施形態において、本発明の実践では談話構成要素を連結し、順序付けることができる他の談話分析の理論が使用されてもよい。テキストが分割された後、制御はステップ S 5 0 0 に続行する。

【 0 0 3 8 】

ステップ S 5 0 0 では、談話構成要素間のリンクが決定される。統一言語談話モデルでは、談話構成要素を談話の一つの構造表現に連結するための結合規則が、1) 統語的、2) 意味論的、3) 語彙的 - 意味論的、4) 構造上、及び、5) デフォルトの規則、として分類されてよい。統語的な規則は、時制、相、主語、目的語、又は、挿入句 (p a r e n t h e t i c a l s)、コンマで区切られた部分等の文の構文識別可能な特徴などの文の構成要素によって伝えられる統語的な属性の談話構成要素内での存在に基づいて談話構成要素間の関係を指定するために使用される。

【 0 0 3 9 】

統一言語談話モデルの意味論的な規則は、自然な言語談話構成要素に割り当てられる意味論上の表現における特定の属性、真の値、又は関係の発生により確立される関係性を指定する。例えば、付加的な文脈がない場合、第 2 の談話構成要素の特定の要素が後に続く第 1 の談話構成要素の一般的な要素の存在は、意味論上の従位接続規則に一致するパターンを定義するために使用される。意味論的な従位接続規則に一致すると、特定の要素を含む談話構成要素は、2 つの談話構成要素を連結する談話の構造表現に従位接続ノードを挿入することによって一般的な要素を含む談話構成要素に従属する。本発明による多様な他の例示的な実施形態では、特定の数量詞、「定冠詞 t h e」と「不定冠詞 a」などの限定詞、あるいは談話構成要素の一般的な解釈と特定の解釈を区別することができる他のパターン又は構造が、本発明に従って意味論的な従位接続を決定するために使用される。

【 0 0 4 0 】

統一言語談話モデルの語彙的 - 意味論的な規則は、単純な語彙項目又は複雑な語彙項目、語彙的な集合、又は構成要素の中のエンティティ又は特性を示す慣用語句が特定の種類の互いの関係性を示す談話構成要素間の関係性を指定する。例えば、「衣服」及び「ジャケット」は特定の種類の関係性を有する。同様に、玄関は建物に対する特定の関係の建築の特徴である。すなわち、建物は必ずしも玄関を含まない場合があるが、玄関はつねに建物の内部で見つけられる。別の例は文「ヨーロッパ人は異なる言語を話す。フランス人はフランス語を話す」によって提示される。第 1 の文の「ヨーロッパ人」と第 2 の文の「フランス人」の間には一般対特定の関係性が存在する。一般対特定の関係性は、ありうる従位接続を示している。句「例えば、しかし～に限られない」は、続く項目が互いになんらかの関係性があることを指定している。形式「の型である」、「の一部である」、「を有する」、「に同等である」、「と対照的である」等の句も、語彙的 - 意味論的な規則を使用して談話ツリーに談話構成要素が結合されるという情報を提供する。これらの句のどれか又はすべての発生は、語彙的 - 意味論的な規則の付加的な集合を使用可能にする及び /

10

20

30

40

50

又は使用不能にするために使用されてよい。

【 0 0 4 1 】

統一言語談話モデルの構造上の規則は、談話の構造表現とその表現に結合されている談話構成要素の間で当てはまる関係の型を指定する。構造上の規則は、その結合の性質だけではなく構造表現に対する談話構成要素の構造上の結合点も制限する。構造表現に結合される談話構成要素は単一の文から導出され、談話単位の主ノードに対応する単位に基づいて結合される一つ又は複数の単位から構成される。したがって、構造上の規則が、談話構成要素を談話の構造表現に追加するときに考慮する必要のある候補結合点の数を削減する上で役立つことが明らかになるであろう。

【 0 0 4 2 】

本発明による多様な他の例示的な実施形態では、結合点変数 (A P)、主要基本談話単位変数 (M - B D U)、現在の談話構成要素又は他の定義された変数などの変数が結合規則に関連付けられたパターンと一致すると、結合規則は始動 (f i r e d) されることが明らかとなるであろう。結合規則は随機的に相互に排他的な規則の集合、特定の規則又は規則の集合が始動した後にだけ有効にされる規則の集合、あるいは他のカテゴリに編成されてよい。本発明によるさらに他の実施形態では、複数の規則が適用されてよく、それぞれの規則は異なる重要性和関連付けられている。談話構成要素間のリンクが決定された後で、制御はステップ S 6 0 0 に移行する。

【 0 0 4 3 】

ステップ S 6 0 0 では、談話構成要素のための関連性スコアが決定される。談話構成要素の関連性スコアは、統計、キーワード検索、知識ベース又は既知のあるいは今後開発される関連性基準に基づいて決定される。談話構成要素の関連性スコアを決定した後、制御はステップ S 7 0 0 に移行する。

【 0 0 4 4 】

閾値重要性スコアがステップ S 7 0 0 で決定される。閾値重要性スコアはユーザプロファイルから取り出される、及び / 又は、動的に入力されてよい。閾値重要性スコアは重要であると見なされる談話構成要素に必要な重要性値を指定する。閾値重要性値が決定された後、制御はステップ S 8 0 0 に移行する。

【 0 0 4 5 】

ステップ S 8 0 0 では、閾値重要性スコアを超える関連性スコアを有する重要談話構成要素が決定される。関連性スコア及び閾値重要性スコアが、負のスコア、相互作用を表すスコア、又は既知のあるいは今後開発される単調な機能を使用して比較できる他の値で指定されてもよいことは明らかとなるであろう。次に、制御はステップ S 9 0 0 に移行する。

【 0 0 4 6 】

重要談話構成要素に連結された支持する先行談話構成要素はステップ S 9 0 0 で決定される。支持する先行談話構成要素は、照応に対する先行する対象を見つけることができる談話構成要素である。重要談話構成要素ノードとルートノードの間のノードが決定され、制御はステップ S 1 0 0 0 に移行する。

【 0 0 4 7 】

ステップ S 1 0 0 0 では、ハイブリッドテキスト要約が、リーフノード又はリーフノードとして重要談話構成要素を有する範囲又はサブツリーに基づいて決定される。支持する先行ノードの子ノードが談話分析の選択された理論のための結合規則に基づいてハイブリッドテキスト要約に選択的に含まれる。ハイブリッドテキスト要約が決定された後、制御はステップ S 1 1 0 0 に移行する。

【 0 0 4 8 】

ハイブリッドテキスト要約はステップ S 1 1 0 0 で表示される。本発明による多様な例示的な実施形態では、S 7 0 0 から S 1 0 0 0 が、異なる選択された閾値重要性値に基づいてハイブリッドテキスト要約を動的に調整するために反復されることが明らかとなるであろう。

10

20

30

40

50

【 0 0 4 9 】

図 3 は、本発明の例示的な実施形態による第 1 のハイブリッド要約システムである。ハイブリッドテキスト要約システム 1 0 0 は、メモリ 2 0、プロセッサ 3 0、談話構成要素分割ルーチン又は回路（手段） 4 0、談話構成要素リンク決定ルーチン又は回路（手段） 5 0、関連性スコア決定ルーチン又は回路（手段） 6 0、及び連結先行談話構成要素決定ルーチン又は回路（手段） 7 0 を備えている。これらの各々は、通信リンク 9 9 及び入力／出力回路（手段） 1 0 を介して、ウェブ使用可能なパーソナルコンピュータ 3 0 0、ウェブ使用可能なタブレットコンピュータ 4 0 0、及び、文書 1 0 0 0 0 から 1 0 0 0 2 を含む情報リポジトリ 2 0 0 に接続される

【 0 0 5 0 】

本発明による多様な例示的な実施形態の一つでは、ウェブ使用可能パーソナルコンピュータ 3 0 0 のユーザが、情報リポジトリ 2 0 0 に含まれる文書 1 0 0 0 0 のハイブリッド要約に対する要求を開始する。前記要求は通信リンク 9 0 0 を介してハイブリッド要約システム 1 0 0 の入力／出力回路 1 0 に転送される。プロセッサ 3 0 は、通信リンク 9 9 を介して情報リポジトリ 2 0 0 から要求された文書 1 0 0 0 0 を取り出し、文書 1 0 0 0 0 をメモリ 2 0 に記憶するために入力／出力回路 1 0 を起動する。

【 0 0 5 1 】

プロセッサ 3 0 は、テキストを談話構成要素に分割するために談話構成要素分割ルーチン又は回路 4 0 を起動する。分割は統一言語談話モデル、修辞構造理論又は任意の既知の又は今後開発される談話分析の理論を使用して実行されてよい。

【 0 0 5 2 】

テキストが談話構成要素に分割された後に、プロセッサ 3 0 は、談話構成要素リンク決定ルーチン又は回路 5 0 を活性化することによって談話分析の決定された理論に基づき、談話構成要素間のリンクを決定する。例えば、統一言語談話モデルの分割規則を使用して分割された談話構成要素は、結合規則に基づいて談話の右に開いたツリー構造表現に挿入される。前述したように、統一言語談話モデルは、統語的、意味論的、語彙的 - 意味論的、構造上、及びデフォルトの結合規則の型を指定する。一つ又は複数の異なる型の結合規則及び／又は制限を備える複合結合規則も提供される。

【 0 0 5 3 】

本発明による多様な他の例示的な実施形態では、始動する規則の条件及び／又はシーケンスを指定するために、異なる範囲の変数が使用されてよい。例えば、規則は、第 1 の集合での第 1 の規則の始動が第 2 のあるいは除外された集合での第 2 の規則の始動を一時的にあるいは恒久的に抑制できるように除外集合に合体されてよい。規則は第 1 の規則の始動が、第 3 の集合の第 3 の規則を始動のために検討できるように使用可能にする集合に合体されてもよい。しかしながら、規則が本発明の範囲から逸脱することなく、既知の又は今後開発されるシーケンス及び／又は組み合わせに基づいて起動できることが明らかでなければならない。

【 0 0 5 4 】

次にプロセッサ 3 0 が関連性スコア決定ルーチン又は回路 6 0 を起動することにより重要な又は関連する談話構成要素を決定する。重要談話構成要素は、統計、キーワード、知識ベース及び／又は他の関連性情報に基づいて決定される。関連性スコアは先に決定された閾値重要性スコアに比較される。閾値重要性スコアを超える関連性スコアの談話構成要素が重要談話構成要素と見なされる。しかしながら、関連性スコアの任意の順序付けが本発明の範囲から逸脱することなく使用されてよいことが明らかとなるであろう。

【 0 0 5 5 】

連結された先行する談話構成要素決定ルーチン又は回路 7 0 は、重要談話構成要素とルート談話構成要素間の支持する談話構成要素を決定するために起動される。例えば、統一言語談話モデルでは、談話構成要素は談話の正しいツリー構造表現に追加される。重要談話構成要素は、談話の構造表現のサブツリーのリーフノードとしての役割を果たす。決定された重要談話構成要素ノードの各親の談話構成要素、及び親の親が支持する先行談話構

10

20

30

40

50

成要素ノードとして決定される。重要な支持する先行談話構成要素ノードのサブツリーがハイブリッドテキスト要約を定義する。本発明による多様な例示的な実施形態では、サブツリー、ハイブリッドテキスト要約のために選択された先行する支持する談話構成要素を決定するために各談話構成要素と関連付けられる関連性スコアに対して調整が行われる。

【0056】

図4は、本発明の例示的な実施形態によるハイブリッドテキスト要約を決定する第2の方法である。プロセスはステップS1300で開始し、ステップS1400に続行する。

【0057】

ステップS1400では、要約されるテキストが決定される。テキストは、エディタ又はワープロを使用してテキストを手動で入力する、ウェブサーバ、デジタルライブラリ等の情報リポジトリからテキストを選択する、ボイス・ディクテーションを使用する及び/又は既知の又は今後開発される入力及び/又は選択方法を使用して決定されるテキストを入力することによって決定されてよい。要約されるテキストが決定された後、制御はステップS1500に続行する。

【0058】

談話分析の理論はステップS1500で決定される。談話分析の理論は、統一言語談話モデル、言語談話モデル、修辞構造理論あるいは談話の構造表現に談話構成要素を連結し、順序付けることができる他の又は今後開発される談話分析の理論を含んでよいが、それらに限られない。談話分析の理論が決定された後、制御はステップS1600に続行する。

【0059】

次にステップS1600でテキストを談話構成要素に分割するための規則が決定される。例えば、統一言語談話モデルに関連付けられる多様な分割規則は、品詞(part-of-speech)タグ、語彙目録及び決定された談話構成要素を従位接続、等位接続、及び二分ノード又はN分(N-ary)ノードに分割、分類するための他の属性又は特徴を使用する。分割規則は分割規則記憶構造から取り出されるか、あるいは談話分析の理論に基づいて動的に入力される。分割規則が決定された後、制御はステップS1700に続行する。

【0060】

ステップS1700では、結合規則が談話分析の理論に基づき決定される。統一言語談話モデルの結合規則は、統語的、意味論的、語彙的・意味論的、構造上、及びデフォルトの規則の型を含む。結合規則は談話構成要素が談話の構造表現にどのように結合されるのかを決定する規則シーケンス又は組み合わせに合体される。結合規則は結合規則記憶構造から取り出されるか、あるいは動的に入力されてよい。結合規則が決定された後、制御は分割されるテキストの部分が決定されるステップS1800に続行する。

【0061】

本発明による多様な例示的な実施形態では、テキストの部分は、強調表示、カーソルコマンド、音声制御、あるいは既知の又は今後開発されるテキスト選択方法によって決定されてよい。制御は、次に、決定されたテキスト部分が、決定された分割規則に基づいて談話構成要素に分割されるステップS1900に続行する。それから制御はステップS2000に続行する。

【0062】

次にステップS2000で談話の構造表現が決定される。談話の構造表現は談話分析の理論に関連付けられる結合規則に基づき決定された談話構成要素を連結する。統一言語談話モデルの統合的、意味論的、語彙的・意味論的、構造上及びデフォルトの規則が説明されるが、本発明の実践では談話構成要素間の順序付けられたリンクを決定する際に有効な既知の又は今後開発される結合規則も使用されてよいことが明らかとなるであろう。談話の構造表現が決定された後に、制御は次にステップS2100に続行する。

【0063】

ステップS2100では、談話構成要素のための関連性スコアが関連性の少なくとも一

10

20

30

40

50

つの基準に基づいて決定される。談話構成要素の関連性スコアは、統計的なスコア、キーワード等に基づいて決定されてよい。しかしながら、本発明の実践では既知の又は今後開発される関連性基準も使用されてよい。関連性スコアが決定された後、制御は、関連性スコアが談話の構造表現中の利用しやすい談話構成要素に浸透されるステップS 2 2 0 0に続行する。例えば、各談話構成要素の終端の子のノード又はリーフノードには、より関連性が高い関連性スコアの任意の直接的に又は間接的に従属する子の談話構成要素ノードの関連性スコアが割り当てられる。

【 0 0 6 4 】

多様な他の例示的な実施形態では、関連性スコアは、潜在的に照応を分割できる先行する談話構成要素に浸透される。さらに他の例示的な実施形態では、任意の一つの先行する対象が談話構成要素リーフノードの中の照応に代用される。関連性スコアが談話の構造表現を通して浸透された後で、制御はステップS 2 3 0 0に続行する。

【 0 0 6 5 】

ステップS 2 3 0 0では、閾値重要性スコアが決定される。閾値重要性スコアはユーザによってインタラクティブに入力され、メモリから取り出され、ユーザid等に基づき決定される。閾値重要性スコアは重要と見なされるために必要な関連性スコアを反映する。次に、制御は、談話構成要素のための関連性スコアが閾値重要性スコアと比較されるステップS 2 4 0 0に続行する。談話構成要素関連性スコアの比較は、テキストの圧縮を選択的に調整するために使用される。次に、制御は、照応が随意的に代用されるステップS 2 5 0 0に続行する。照応の任意の一つの先行する対象が決定され、随意的に照応に代用される。閾値重要性スコアを超える関連性スコアの談話構成要素がステップS 2 6 0 0で決定され、随意的に表示される。次に、制御はステップS 2 7 0 0に続行し、プロセスは終了する。

【 0 0 6 6 】

図5は、本発明の例示的な実施形態による第2のハイブリッド要約システムである。ハイブリッド要約システム110は、メモリ20、プロセッサ30、分割規則記憶装置40、結合規則記憶装置50、談話構造表現決定ルーチン又は回路(手段)60、関連性スコア決定ルーチン又は回路(手段)70、浸透ルーチン又は回路(手段)80、及び、表示ルーチン又は回路(手段)90を備える。これらは、入力/出力回路10に、及び、通信リンク99を介してウェブ使用可能パーソナルコンピュータ300、ウェブ使用可能タブレットコンピュータ400、及び、文書10000から10002を含む情報リポジトリ200に、接続される、

【 0 0 6 7 】

本発明による多様な例示的な実施形態の一つでは、ウェブ使用可能パーソナルコンピュータ300のユーザが情報リポジトリ200に含まれる文書10000のハイブリッドテキスト要約に対する要求を開始する。前記要求は通信リンク99を介してハイブリッド要約システム110の入力/出力回路10に転送される。プロセッサ30は、通信リンク99を介して情報リポジトリ200から文書10000を取り出し、それをメモリ20に記憶するために入力/出力回路10を起動する。

【 0 0 6 8 】

プロセッサ30は分割規則記憶装置40から取り出された分割規則を適用する。分割規則は先に決定された談話分析の理論に関連付けられている。分割規則は、テキストを談話構成要素にどのようにして分割するのかを指定する。本発明による多様な例示的な実施形態では、談話分析の理論が、本発明の範囲から逸脱することなく、ユーザによる入力、テキストのジャンル、要約されるテキストの一部に基づいて、又は他の方法によって決定されてよいことが明らかとなるであろう。

【 0 0 6 9 】

談話構造表現決定ルーチン又は回路60は、談話の構造表現に談話構成要素を結合するために起動される。例えば、統一言語談話モデル(ULD M)は、談話構成要素が談話の構造表現にどのようにして結合されるのかを定義する結合規則を指定する。統一言語談話

10

20

30

40

50

モデルは、5種類の結合規則、つまり1)統語的、2)意味論的、3)語彙的 - 意味論的、4)構造上、及び5)デフォルト、を指定する。結合規則の異なる種類は単独で使用されてよいし、あるいは、組み合わせ及び/又は規則シーケンスに合体されてよい。

【0070】

統語的結合規則は、時制、相、主語、目的語又は挿入句、コンマで区切られた部分等の文の構文識別可能特徴などの文の構成要素により運ばれる統語的な属性の談話構成要素での存在に基づいた談話構成要素間の関係を指定する。

【0071】

意味論的な結合規則は、自然な言語談話構成要素に割り当てられる意味論的な表現における特定の属性、真の値又は関係の発生により確立される関係性を指定する。例えば、付加的な文脈がない場合、第2の談話構成要素におけるより特定の要素が後に続く第1の談話構成要素のより一般的な要素の存在が、意味論的な従位接続規則に一致するために使用できるパターンを定義する。意味論的な従位接続規則は、より一般的な談話構成要素に特定の談話構成要素を従属させる。特定の数量詞、「定冠詞 the」及び「不定冠詞 a」などの限定詞、又は談話構成要素の一般的な解釈及び特定の解釈を区別することに関連付けられた他の決定されたパターン及び/又は構造も、特定の談話構成要素の一般的な談話構成要素に対する従属をトリガするために使用されてよい。

【0072】

統一言語談話モデルの語彙的 - 意味論的な規則は、単純な語彙項目又は複雑な語彙項目、語彙的な集合、又は構成要素の中のエンティティ又は特性を示す慣用語句が特定の種類の互いの関係性を示す談話構成要素間の関係性を指定する。例えば、「衣服」及び「ジャケット」は特定の種類の関係性を有する。同様に、玄関は建物に対する特定の関係の建築の特徴である。すなわち、建物は必ずしも玄関を含まない場合があるが、玄関はつねに建物の内部で見つけられる。別の例は文「ヨーロッパ人は異なる言語を話す。フランス人はフランス語を話す」によって提示される。第1の文の「ヨーロッパ人」と第2の文の「フランス人」の間には一般対特定の関係性が存在する。一般対特定の関係性は、ありうる従位接続を示している。句「例えば、しかし～に限られない」は、続く項目が互いになんらかの関係性があることを指定している。形式「の型である」、「の一部である」、「を有する」、「に同等である」、「と対照的である」等の句も、語彙的 - 意味論的な規則を使用して談話ツリーに談話構成要素が結合されるという情報を提供する。これらの句のどれか又はすべての発生は、語彙的 - 意味論的な規則の付加的な集合を使用可能にする及び/又は使用不能にするために使用されてよい。

【0073】

構造上の結合規則は、結合されている談話構成要素の一つ又は両方と、すでに談話構造の一部を形成している他の談話構成要素の間に当てはまる関係の種類を指定する。構造上の結合規則は、結合される一つ又は複数の談話構成要素のための結合を制限するために使用される。例えば、結合される2つの談話構成要素の場合、規則はどの談話結合物がツリーに直接的に結合されなければならないのか、及び、どこで結合が起こらなければならないのかを指定する。このようにして構造上の規則は、考慮しなければならない候補結合点の数を削減するために使用できる。デフォルトの結合規則は、他の規則が当てはまらないときに講じる処置を指定する。デフォルトの規則は他の種類の結合規則の組み合わせを含んでよい。

【0074】

本発明による多様な別の例示的な実施形態では、談話構成要素の属性が規則の属性に一致すると規則が始動することが明らかとなるであろう。規則は、相互に排他的な規則、他の規則が始動した後にだけ、あるいは他の既知の又は今後開発される規則シーケンスに基づいて有効にされる規則集合に編成されてよい。本発明による多様な実施形態では、複数の規則が適用されてよく、各規則はさまざまな重みと関連付けられる。

【0075】

プロセッサ30は、次に、各談話構成要素ごとに関連性スコアを決定するために関連性

10

20

30

40

50

スコア決定ルーチン又は回路70を起動する。関連性の基準は、統計、キーワード、知識ベース、あるいは既知の又は今後開発される非構造方法に基づいて決定されてよい。

【0076】

プロセッサ30は、談話構成要素リーフノードから先行する談話構成要素ノードに関連性スコアを浸透するために浸透ルーチン又は回路80を起動する。初期の関連性スコアは、談話の構造表現を通して上に浸透される。重要談話構成要素の文法性を保証するために必要な支持する利用しやすい談話構成要素は関連性スコアに基づいて選択される。多様な例示的な実施形態では、非構造関連性スコアは、ハイブリッドテキスト要約の文法性が保たれるように談話の構造表現を通して浸透される。支持する利用しやすい談話構成要素の数は、テキストジャンル、言語、要約の関数及び/又は既知のあるいは今後開発される属性又は特徴に基づいて動的に選択可能である。

10

【0077】

表示ルーチン又は回路90は、先に決定された閾値重要性スコアを超える関連性スコアの談話構成要素ノードを決定するために起動される。本発明による多様な別の例示的な実施形態では、支持する談話構成要素からの一つの識別された先行する対象が、ハイブリッドテキスト要約をさらに圧縮するために談話構成要素リーフノードへの照応に代用される。

【0078】

図6は、本発明の実施形態による分割規則600を記憶するための例示的なデータ構造である。分割規則600を記憶するための例示的なデータ構造は各規則を一意的に識別する識別子部分610、異なる種類の規則を分類する一つ又は複数の型部分620、規則を開始するために必要とされる制限を指定するパターン部分630、及び規則の制限が一致される場合に実行される動作を指定する動作部分640から構成されている。

20

【0079】

分割規則は、談話構成要素及び/又は文の単位の主要な談話構成要素として候補セグメントを分類するための条件を定義する。談話構文解析は、文の単位を談話構成要素に分割することと、一つ又は複数の主要な談話構成要素を決定することと、主要な談話構成要素に従属する文の談話構成要素のサブツリーを決定することと、談話の理論に基づいて結合規則を使用して談話の現存の構造表現に対するサブツリーの結合を決定することを含む。結合規則は、文の主要な談話構成要素及び/又は結合点談話構成要素に対する制限を指定する。例えば、異なる結合規則は、談話構成要素が修飾語句であるのか、関係詞節であるのか、文の主語等であるのかに応じてトリガされる。候補セグメントの属性又は特徴が品詞タグ、環境変数、ゼロックス言語環境(XLE)の出力、蓋然論的な、統計上の、有限状態の又は記号の、構文解析ツール、WordNetなどの語彙目録、あるいは他の既知の又は今後開発される候補談話構成要素の特徴を決定する方法から決定されることが明らかとなるであろう。

30

【0080】

分割規則情報を記憶するための例示的なデータ構造の第1列は識別子部分610に値「1」を、型部分620に値「SIMPLE__EVENTS(単純事象)」を含む。これは、規則「1」として識別される規則が単純事象型規則であることを示している。本発明による多様な例示的な実施形態では、規則は型部分620の値に基づいて規則の異なる集合に集められてよい。規則及び/又は型の異なる集合は、指定の規則の後に始動してよい規則のシーケンスを制限する多様な他の規則を除外する、及び/又は有効にしてよい。本発明によるさらに他の例示的な実施形態では、複数の規則が適用されてよく、それぞれの規則がさまざまな重要性和関連付けられている。

40

【0081】

パターン部分630は値「(SIMPLE__EVENT)」を含む。これは、単純事象が、規則が始動する前に候補セグメントで決定されなければならないことを示す。単純事象の決定が、蓋然論的な、統計的な、有限状態の又は記号の構文解析ツール、文の構文解析ツール、ゼロックス言語環境(XLE)などの言語処理環境あるいは既知の又は今後開

50

発される方法の出力に基づいてよいことは明らかとなるであろう。

【0082】

規則の動作部分640は、パターン部分630が一致されると起動される。動作部分640は、パターン部分630の制限が一致すると、規則の動作部分が、フラグ又は候補セグメントが基本的な談話単位である旨の他の表示を設定するために呼び出されることを示す値「BDU=TRUE」を含む。

【0083】

分割規則情報を記憶するための例示的なデータ構造の第2列は、識別子部分610に値「2」及び型部分620に値「DISCLOSURE_OPERATOR（開示演算子）」を含む。これは、規則「2」として識別される規則が開示演算子型規則であることを示している。

10

【0084】

パターン部分630は値「MODIFIER（修飾語句）」を含む。これは、候補談話構成要素セグメントが規則を始動するための修正語句でなければならないことを示している。前述されたように、修正語句の決定は、文の構文解析ツール、言語処理環境等の出力に基づいてよい。

【0085】

規則の動作部分640は、パターン部分630が一致すると起動される。動作部分640は、パターン部分630の制限が一致すると実行される動作を示す値「BDU=TRUE」を含む。

20

【0086】

分割規則情報を記憶するための例示的なデータ構造の第3列は、識別子部分610に値「3」を、型部分620に値「INTERPOLATION（補間）」を含む。これは、規則「3」として識別される規則が補間型規則であることを示している。

【0087】

パターン部分630は値「APPOSITIVE OR PARENTHETICAL OR INTERRUPTION OR NON_RESTRICTIVE_CLAUSE OR REFERENCE（同格又は挿入又は中断又は非制限節又は参照）」を含む。これは、候補セグメントが同格、挿入、中断、又は非制限節又は参照であると判断される場合、規則が始動することを示している。前述したように、制限は蓋然的な、統計的な、有限状態の、記号の、文の構文解析ツール、ゼロックス言語環境（XLE）などの言語処理環境あるいは既知の又は今後開発される方法の出力に基づいて決定される。

30

【0088】

規則の動作部分640は、パターン部分630が一致すると起動される。動作部分640は、パターン部分630の制限が一致すると、フラグ又は候補セグメントが基本的な談話単位である旨の他の表示を設定するために、規則の動作部分が呼び出されることを示す値「BDU=TRUE」を含む。

【0089】

分割規則情報を記憶するための例示的なデータ構造の第4列は、識別子部分610に値「4」を、型部分620に値「SUBORDINATE_CLAUSES（従属節）」を含む。これは、規則「4」として識別される規則が従属節型の規則であることを示している。

40

【0090】

パターン部分630は、候補セグメントが、規則が始動する前に補語節、文の主語、又は関係節でなければならないことを示す値「COMPLEMENT_CLAUSE OR

【0091】

SENTENTIAL_SUBJECT OR RELATIVE_CLAUSE（補語節又は文の主語又は関係節）」を含む。規則の動作部分640は、パターン部分630が一致すると起動される。動作部分640は、パターン部分630に含まれる制限が一致すると、規則の動作部分が、候補セグメントが基本的な談話単位であることを示すために

50

呼び出されることを示す値「BDU = TRUE」を含む。

【0092】

分割規則情報を記憶するための例示的なデータ構造の第5列は、識別子部分610に値「5」を、型部分620に値「POSTNOMINAL__MODIFIER（名詞後位修飾語）」を含む。これは、規則「5」と識別された規則が名詞後位修飾語型の規則であることを示している。

【0093】

パターン部分630は、値「(DISCRETE__EVENT AND (POSTNOMINAL__MODIFIER OR RELATIVE__CLAUSE OR PARTICIPIAL__MODIFIER))」(（非連続事象及び（名詞後位修飾語又は関係節又は分詞修飾語）））を含む。これは、これらの制限が、規則が始動する前に候補セグメントと関連付けられていると判断されなければならないことを示している。

10

【0094】

規則の動作部分640は、パターン部分630が一致すると起動される。動作部分640は、パターン部分630の制限が一致すると、規則の動作部分が、フラグ又は候補セグメントが基本的な談話単位である旨の他の表示を設定するために呼び出されることを示す値「BDU = TRUE」を含む。

【0095】

分割規則情報を記憶するための例示的なデータ構造の第6列は、識別子部分610に値「6」を、型部分620に値「MORPHOLOGICAL__FORM（形態的形）」を含む。これは、「6」と識別される規則が形態型規則であることを示している。パターン部分630は、候補セグメントが、規則が始動する前に非連続事象であり、不定詞又は動名詞のどちらかであると判断されなければならないことを示す値「(DISCRETE__EVENT AND (INFINITIVES OR GERUNDS))」(（非連続事象と（不定詞又は動名詞）））を含む。

20

【0096】

規則の動作部分640は、パターン部分630が一致すると起動される。動作部分640は、パターン部分630の制限が一致すると、規則の動作部分がフラグ又は候補セグメントが基本的な談話単位である旨の他の表示を設定するために呼び出されることを示す値「BDU = TRUE」を含む。

30

【0097】

分割規則情報を記憶するための例示的なデータ構造の第7列は、識別子部分610に値「7」を、型部分620に値「OTHER__PREDICTION（他の予測）」を含む。これは、規則が他の予測型の規則と識別されることを示している。パターン部分630は値「(COPULAR__PREDICTION OR SECONDARY PREDICTION OR DISCREAT__EVENT)」(（連結詞予測又は二次予測又は非連続事象））を含む。これは、連結詞予測、二次予測又は非連続事象が、規則が始動する前に決定されなければならないことを示している。

【0098】

規則の動作部分640は、パターン部分630が一致すると起動される。動作部分640は、パターン部分630の制限が一致すると、フラグ又は候補セグメントが基本的な談話単位である旨の他の表示を設定するために、規則の動作部分が呼び出されることを示す値「BDU = TRUE」を含む。

40

【0099】

分割規則情報を記憶するための例示的なデータ構造の第8列は、識別子部分610に値「8」を、型部分620に値「SUBORDINATING__CONJUNCTION（従属接続）」を含む。これは、規則「8」と識別される規則が従属接続詞型の規則であることを示している。パターン部分630は、値「(SUBORDINATING__CONJUNCTION AND DISCRETE__EVENT)」(（従属接続と非連続事象））を含む。これは、これらの制限が、規則が始動する前に候補セグメントで決定され

50

なければならないことを示している。

【 0 1 0 0 】

規則の動作部分 6 4 0 は、パターン部分 6 3 0 が一致すると起動される。動作部分 6 4 0 は、パターン部分 6 3 0 の制限が一致されると、フラグ又は候補セグメントが基本的な談話単位である旨の他の表示を設定するために、規則の動作部分が呼び出されることを示す値「BDU = TRUE」を含む。

【 0 1 0 1 】

分割規則情報を記憶するための例示的なデータ構造の第 9 列は、識別子部分 6 1 0 に値「9」を、型部分 6 2 0 に値「COORDINATING__CONJUNCTIONS (等位接続)」を含む。これは、規則「9」として識別される規則が等位接続詞型規則であることを示している。パターン部分 6 3 0 は、値「((COORDINATING__CONJUNCTION AND LENGTH = 1) OR ((X COORDINATING__CONJUNCTION Y) AND (X ISA BDU AND Y ISA BDU))((等位接続 及び 長さ = 1) 又は ((X 等位接続 Y)

【 0 1 0 2 】

及び (X は BDU 及び Y は BDU))」を含む。これは、規則が始動する前に、等位接続詞が 1) 長さ = 1 によって示されるように候補談話構成要素セグメント内の唯一の要素、又は、2) それぞれが BDU である要素 X と Y を連結する等位接続詞、のどちらかであることを判断されなければならないことを示している。候補セグメントが指定された制限のどちらかに一致すると、規則が始動する。

【 0 1 0 3 】

パターン部分 6 3 0 が一致すると、規則の動作部分 6 4 0 が起動される。動作部分 6 4 0 は、パターン部分 6 3 0 の制限が一致すると、フラグ又は候補セグメントが基本的な談話単位である旨の他の表示を設定するために、規則の動作部分が呼び出されることを示す値「BDU = TRUE」を含む。

【 0 1 0 4 】

分割規則情報を記憶するための例示的なデータ構造の第 10 列は、識別子部分 6 1 0 に値「10」を、型部分 6 2 0 に値「INFORMATION__STRUCTURE (情報構造)」を含む。これは、規則「10」として識別される規則が情報構造型の規則であることを示している。パターン部分 6 3 0 は値「(COPULA AND (CLEFT OR PSEUDO__CLEFT))((連結詞及び分裂又は擬似分裂))」を含む。これは、候補セグメントが、規則が始動する前に連結詞、及び分裂又は擬似分裂のどちらかでなければならないことを示している。

【 0 1 0 5 】

パターン部分 6 3 0 が一致すると規則の動作部分 6 4 0 が起動される。動作部分 6 4 0 は、パターン部分 6 3 0 の制限が一致すると、フラグ又は候補セグメントが基本的な談話単位ではない旨の他の表示を設定するために、規則の動作部分が呼び出されることを示す値「BDU = FALSE」を含んでいる。

【 0 1 0 6 】

分割規則情報を記憶するための例示的なデータ構造の第 11 列は、識別子部分 6 1 0 に値「11」を、型部分 6 2 0 に値「EVENT__MODIFIER (事象修飾語句)」を含む。これは、規則「11」と識別される規則が事象修飾語句型の規則であることを示している。パターン部分 6 3 0 は、候補セグメントが、規則が始動するための非連続事象の修飾語句であり、非連続事象と関連付けられてはならないことを示す値「(EVENT__MODIFIER AND NOT (DISCRETE__EVENT))((事象修飾語及び非(非連続事象)))」を含む。

【 0 1 0 7 】

パターン部分 6 3 0 が一致すると、規則の動作部分 6 4 0 は起動される。動作部分 6 4 0 は、パターン部分 6 3 0 の制限が一致すると、フラグ又は候補セグメントが基本的な談話単位ではない旨の他の表示を設定するために、規則の動作部分が呼び出されることを示

10

20

30

40

50

す値「BDU = FALSE」を含む。

【0108】

分割規則情報を記憶するための例示的なデータ構造の第14列は、識別子部分610に値「14」を、型部分620に値「INITIAL__MODIFIER（初期修飾子）」を含む。これは、第14の規則が初期修飾子型の規則であることを示している。パターン部分630は、値「（INITIAL__MODIFIER AND SCENE__SETTING）（初期修飾子及び状況設定）」を含む。これは、候補セグメントが初期修飾語句でなければならない、規則の始動をトリガするためにシーン設定値と関連付けられなければならないことを示している。

【0109】

パターン部分630が一致すると、規則の動作部分640が起動される。動作部分640は値「BDU = TRUE」を含む。パターン部の630の制限が一致すると、フラグ、若しくは、候補セグメントが基本的な談話単位でないことを示す他の表示を設定するために、規則の動作部分が呼び出されることを、該値は示す。

【0110】

図7は、本発明の実施形態による結合規則情報700を記憶するための例示的なデータ構造である。結合規則情報700を記憶するための例示的なデータ構造は、識別子部分710及び規則部分720から構成される。

【0111】

結合規則情報を記憶するための例示的なデータ構造の第1列は、規則識別子部分710に値「1」を含む。規則部分720は、複合規則「< Sx : (M - BDU / TYPE) = " Between Parenthesis " ; St : AP is Bottom Of Tree ; > Subordinate (AP , M - BDU) 」を含む。規則「1」は、主要なBDUが括弧によって区切られることを示す「 Sx : (M - BDU / TYPE) = " Between Parenthesis "」、及び、結合点が談話の構造表現の右に開いたツリーの底部になければならないことを示す第2の構造上の制限「 St : AP is Bottom Of Tree 」によって示される第1の統語制限から構成される複合規則である。第1の制限及び第2の制限は、両方の制限が、主要な基本談話単位（ M - BDU ）の、結合点変数（ AP ）により示される現在の結合点に対する従属をトリガするために一致しなければならないことを示す開いた制限「 < 」タグと閉じた制限「 > 」タグの間で結合される。

【0112】

結合規則情報を記憶するための例示的なデータ構造の第2列は、規則識別子部分710に「2」を含む。規則部分720は、複合規則「 < Sx : (AP / Tense) = " Present " ; Sx : (M - BDU / Tense) = " Past " ; > Subordinate (AP , M - BDU) 」を含む。規則「2」は、候補結合点談話構成要素が現在時制でなければならないことを示す「 Sx : (AP / Tense) = " Present " 」によって示される第1の統語制限、及び結合点（ AP ）に結合される主要な基本談話構成要素が規則をトリガするために過去時制でなければならないことを示す第2の統語制限「 Sx : (M - BDU / Tense) = " Past " ; 」から構成される複合規則である。第1の制限及び第2の制限は、両方の制限が現在の結合点（ AP ）に対する主要な基本談話単位（ M - BDU ）の従属をトリガするために一致しなければならないことを示す開いた制限「 < 」タグと閉じた制限「 > 」タグの間で結合される。

【0113】

結合規則情報を記憶するための例示的なデータ構造の第3列は、規則識別子部分710に「3」を含む。規則部分720は、複合規則「 < LS : (AP) isa (KB / Event) ; LS : (M - BDU) isa (KB / Event) ; Sx : (AP / Tense) = (M - BDU / Tense) > Coordinate (AP , M - BDU) 」を含む。規則「3」は、「 LS : (AP) isa (KB / Event) ; 」によって示される第1の語彙的 - 意味論的制限から構成される複合規則である。これは、候補結合点が知識

10

20

30

40

50

ベースKBに対する比較により決定されるように事象と関連付けられることを示す。第2の語彙的 - 意味論的制限「 $LS : (M - BDU) is a (KB / Event)$;」は、主要な基本談話単位も、知識ベースKBに対する比較により決定されるように事象と関連付けられることを示す。第3の統語制限「 $Sx : (AP / Tense) = (M - BDU / Tense)$ 」は、結合点AP及び主要な基本談話単位(M - BDU)の時制が一致しなければならないことを示す。第1の制限、第2の制限及び第3の制限は、すべて3つの制限が、主要な基本談話単位(M - BDU)及び現在の結合点(AP)の等位構造をトリガするために始動しなければならないことを示す開いた制限「 $<$ 」タグと閉じた制限「 $>$ 」タグの間で結合される。

【0114】

10

結合規則情報を記憶するための例示的なデータ構造の第4列は、規則識別子部分710に「4」を含む。規則部分720は複合規則「 $< LS : (AP / Lexeme) has a (M - BDU / Lexeme) : \$1 : \$2 ; Sx : (AP / * / COMP) is \$1 ; Sx : (M - BDU / * / SUBJ) is \$2 ; > Subordinate (AP , M - BDU)$ 」を含む。規則「4」は、2つの統語制限が後に続く語彙的 - 意味論的制限から構成される複合規則である。第1、第2、第3の制限は、3の全ての制限が主要な基本談話単位(M - BDU)の現在の結合点(AP)に対する従属をトリガするために始動しなければならないことを示す開いた制限「 $<$ 」タグと閉じた制限「 $>$ 」タグの間で結合される。

【0115】

20

結合規則情報を記憶するための例示的なデータの第5列は、規則識別子部分710に「5」を含む。規則部分720は、複合規則「 $< Sx : (M - BDU / Type) = "Temporal" ; St : M - BDU is Subordinated To AP ; > CreateNary (Cx , AP , M - BDU)$ 」を含む。規則「5」は、「 $Sx : (M - BDU / Type) = "Temporal" ;$ 」によって示される第1の統語制限から構成される複合規則である。これは、主要な基本談話単位(M - BDU)談話構成要素が、「時制(Temporal)」型の談話構成要素でなければならないことを示している。第2の統語制限「 $St : M - BDU is Subordinated To AP$ 」は、主要な基本談話単位M - BDUが結合点に従属することを示している。第1の制限及び第2の制限は、両方の制限が結合点(AP)及び主要な基本談話単位(M - BDU)の等位構造をトリガするために一致しなければならないことを示す、開いた制限「 $<$ 」タグと閉じた制限「 $>$ 」タグの間で結合される。項「Cx」は質疑応答型等位などの等位型を示す。

30

【0116】

結合規則情報を記憶するための例示的なデータ構造の第6列は規則識別子部分710に「6」を含む。規則部分720は、複合規則「 $< Sm : (AP) is "Generic" ; SM : (M - BDU) is "Specific" ; > Subordinate (AP , M - BDU)$ 」を含む。規則「6」は、「 $Sm : (AP) is "Generic" ;$ 」により示される第1の意味論的な制限から構成される複合規則である。これは、結合点が総称的な言語要素を含むことを示している。「 $Sm : (M - BDU) is "Specific" ;$ 」により示される第2の意味論的な制限は、主要な基本談話単位(M - BDU)が特定の言語要素を含むことを示している。第1の制限及び第2の制限は、両方の制限が、主要な基本談話単位(M - BDU)の、変数PAによって示される現在の結合点への従属をトリガするために一致しなければならないことを示す開いた制限「 $<$ 」タグと閉じた制限「 $>$ 」タグの間で結合される。

40

【0117】

結合規則情報を記憶するための例示的なデータ構造の第7列は規則識別子部分710に「7」を含む。規則部分720は、複合規則「 $< Sm : (AP) is "Irrealis" ; SM : (M - BDU) is "Realis" ; Sx : if AP then M - BDU ; > CreateNary (HYP , AP , M - BDU)$ 」を含む。規則「7

50

」はAPが「非現実相 (I r r e a l i s) 」であること、及び主要な基本談話構成要素が「現実相 (R e a l i s) 」であること、並びに結合点 (A P) が、結合点と主要な基本談話構成要素の間のN分関係を作成するために規則が始動される前に一致するという構造制限を必要とする。同様に、8から12の列は、談話の構造表現を決定する上で有用な他の統一言語談話モデル結合規則を記述する。

【 0 1 1 8 】

結合規則情報を記憶するための例示的なデータ構造の第99列は規則識別子部分710に「99」を含む。規則部分720は複合規則「< S t : A P i s B o t t o m O f T r e e ; P x : I f n o t A n y r u l e a p p l i e s ; > S u b o r d i n a t e (A P , M - B D U) 」を含む。規則「7」は、候補結合点がツリーの底部で発生しなければならないことを示す第1の構造上の制限「S t : A P i s B o t t o m O f T r e e 」から構成される複合規則である。先手 (p r e e m p t o r) 制限「P x : I f n o t A n y r u l e a p p l i e s ; 」は、他の制限が適用しない場合にだけ一致される。第1の制限及び第2の制限は、両方の制限が、主要な基本談話単位 (M - B D U) の現在の結合点 (A P) に対する従属をトリガするために一致しなければならないことを示す開いた制限「< 」タグと閉じた制限「> 」タグの間で結合される。

【 0 1 1 9 】

図8は、本発明の実施形態による関連性スコアを浸透する第1の例示的な方法の展開図である。プロセスはステップS2200で開始し、ステップS2202に続行する。

【 0 1 2 0 】

ステップS2202では、各談話構成要素の子のノードの関連性スコアが親の談話構成要素ノードに割り当てられる。これが、重要談話構成要素の関連性スコアを支持する先行する談話構成要素ノードに浸透し、結果として生じるハイブリッドテキスト要約の文法性を保つのに役立つ。制御は次にステップS2204に続行する。

【 0 1 2 1 】

ステップS2204では、被従属談話構成要素ノードの関連性スコアがより関連性がある場合には、各被従属ノードの関連性スコアが従属談話構成要素に割り当てられる。これにより、支持する従属談話構成要素ノードがハイブリッドテキスト要約に含まれることが確実になる。次に制御は、各等位構造又はN分 (N - a r y) ノードの関連性スコアが、関連性スコアがより関連性がある場合に、すべての等位とされた子の談話構成要素ノードに割り当てられるステップS2206に続行する。これにより、すべての等位構造情報が捕捉されることが確実になる。次に、制御はステップS2208に続行し、ただちに図4のステップS2300にジャンプする。

【 0 1 2 2 】

図9は、本発明の実施形態に従って関連性スコアを浸透する第2の例示的な方法の展開図である。プロセスはステップS2210で開始し、ステップS2212に続行する。

【 0 1 2 3 】

談話構成要素の子のノードのそれぞれの関連性スコアが、ステップS2212で、親の談話構成要素ノードに割り当てられる。これは、重要談話構成要素の関連性スコアを、支持する先行する談話構成要素ノードに浸透し、結果として生じるハイブリッドテキスト要約の文法性を保つのに役立つ。次に制御はステップS2214に続行する。

【 0 1 2 4 】

ステップS2214では、被従属談話構成要素ノードの関連性スコアがより関連性があり、ノードが被従属ノードの左側にあると思われる場合には、各被従属ノードの関連性スコアが従属談話構成要素に割り当てられる。これにより、支持する従属談話構成要素ノードがハイブリッドテキスト要約に含まれるが、支持する談話構成要素の、左側にある談話構成要素への追加をさらに制限することが確実になる。言語談話モデルの構造表現では、従属ノードの左側の子のノードは従属ノードである。しかしながら、多様な他の談話分析の理論と関連付けられる談話の構造表現での差異のために、支持する先行する談話構成要素情報を識別する他の方法が、本発明の範囲から逸脱することなく使用されてよいことが

10

20

30

40

50

明らかになるであろう。次に、制御は、各等位構造又はN分（n - a r y）談話構成要素ノード又は二分談話構成要素ノードでは、子の談話構成要素のさらに関連性のある関連性スコアが左側のすべての子の談話構成要素に割り当てられるステップS 2 2 1 6に続行する。次に、制御は、ステップS 2 2 1 8に続行し、図4のステップS 2 3 0 0にジャンプする。

【0125】

図10は、本発明の例示的な実施形態に従って関連性スコアを浸透する第3の例示的な方法の展開図である。プロセスはステップS 2 2 3 4で開始し、ステップS 2 2 3 6に続行する。

【0126】

ステップS 2 2 3 6では、各談話構成要素ノードは、談話構成要素ノードとルートノードに1を加えたものの間の被従属エッジの数でマーキングされる。これは、談話の構造表現のノードレベルを示している。各談話構成要素ノードのレベルがマーキングされてから、制御はステップS 2 2 3 8に続行する。

【0127】

ステップS 2 2 3 8では、各終端の子又はリーフノード談話構成要素が談話の構造表現での被従属エッジの数の逆数に等しいスコアを割り当てられる。スコアがノードごとに決定された後に、制御はステップS 2 2 4 0に続行する。

【0128】

ステップS 2 2 4 0では、低い方のスコアが談話の構造表現を通して浸透される。したがって、親のノードのスコアが子のノードより関連性がある場合には、スコアは親のノードに割り当てられる。次に制御はステップS 2 2 4 2に続行する。

【0129】

任意の被従属ノードのスコアは、被従属スコアがステップS 2 2 4 2でより低い場合には従属ノードに割り当てられる。次に制御はステップS 2 2 4 4に続行する。ステップS 2 2 4 4では、子のノードがより低いスコアを有さない限り、任意の等位構造又はN分（N - a r y）ノードのスコアが各子のノードに割り当てられる。次に制御はステップS 2 2 4 6に続行する。

【0130】

ステップS 2 2 4 6では、談話構成要素ノードのそれぞれについて調整された関連性スコアが従属レベル及びスコアに基づいて決定される。例示的な関連性スコアは、以下の式（1）に基づいて決定される。

【0131】

$$r = 1 - (L - 1) \cdot i \quad (1)$$

【0132】

ここで、Lはノードの従属レベルであり、iは各ノードに割り当てられるスコアである。項「- 1」は、より低い関連性と関連するより低い数値を有する実際の値に、関連性スコアをするために追加される。ルートノードは関連性「1」を有し、リーフノードは1 / Lという初期関連性スコアを有する。関連性スコアを決定した後、制御はステップS 2 2 4 8に続行し、図4のステップS 2 3 0 0にジャンプする。

【0133】

図11は、本発明の実施形態による照応置換の第1の例示的な方法の展開図である。プロセスはステップS 2 5 0 0で開始し、代名詞などの照応を含む各談話構成要素の子のノードが決定されるステップS 2 5 0 5に続行する。制御は次にステップS 2 5 1 0に続行する。

【0134】

ステップS 2 5 1 0では、照応の任意のユニークな先行する対象が決定される。代名詞などの照応が通常それらの対象に続くことが明らかとなるであろう。このようにして、照応又は代名詞の、任意のユニークな先行する対象は、通常、談話の構造表現中の照応又は代名詞の前に発生する。これにより、調べられなければならない先行する談話構成要素の

10

20

30

40

50

数が削減される。次に、任意の決定されたユニークな先行する対象が照応又は代名詞に代用される。次に制御はステップ S 2 5 1 5 に続行し、それから図 4 のステップ S 2 6 0 0 にジャンプする。

【 0 1 3 5 】

図 1 2 は、本発明の実施形態による照応置換の第 2 の例示的な方法の展開図である。プロセスはステップ S 2 5 2 0 で開始し、ステップ S 2 5 2 5 に続行する。

【 0 1 3 6 】

重要談話構成要素はステップ S 2 5 2 5 で決定される。重要談話構成要素は、通常、前のステップで決定され、取り出される。しかしながら、重要談話構成要素は、統計、キーワード、知識ベース等を使用して動的に決定されてもよい。重要談話構成要素が決定されてから、制御はステップ S 2 5 3 0 に続行する。

10

【 0 1 3 7 】

次に重要談話構成要素の解決されていない照応がステップ S 2 5 3 0 で決定される。例えば、重要談話構成要素の中の任意の代名詞は、解決されていない照応としてフラグが立てられる。次に制御はステップ S 2 5 3 5 に続行する。

【 0 1 3 8 】

ステップ S 2 5 3 5 では、潜在的な解決先行談話構成要素が決定される。代名詞又は他の未解決照応の潜在的な先行する参照を含む任意の談話構成要素は潜在的な解決先行談話構成要素である。本発明による多様な例示的な実施形態では、潜在的に解決する先行談話構成要素は Word Net 関係、存在論又は他の任意の関係連鎖の方法に基づいて決定される。潜在的な解決先行談話構成要素が決定されてから、制御はステップ S 2 5 4 0 に続行する。

20

【 0 1 3 9 】

重要談話構成要素の関連性スコアは、ステップ S 2 5 4 0 で決定された潜在的な解決先行談話構成要素を通して浸透される。潜在的な解決先行談話構成要素を通して浸透する関連性スコアは、考慮される分岐の範囲を削減してよい。これは、頻繁に、より縮約されたハイブリッドテキスト要約を生じさせる。次に制御はステップ S 2 5 4 5 に続行する。

【 0 1 4 0 】

ステップ S 2 5 4 5 では、談話構成要素の縮約された集合が重要談話構成要素、談話の構造表現及び関連性スコアに基づいて決定される。浸透チャネルは重要談話構成要素の中で潜在的に照応を解決できる談話構成要素に制限される。談話構成要素の縮約された集合がハイブリッドテキスト要約として選択されてから、制御はステップ S 2 5 5 0 に続行し、図 4 のステップ S 2 6 0 0 にジャンプする。

30

【 0 1 4 1 】

図 1 3 は要約されるテキストである。テキストは 1 2 の文から構成されている。

【 0 1 4 2 】

図 1 4 は、要約されるテキストの談話の例示的な構造表現である。談話の例示的な構造表現は結合されている談話構成要素から構成されている。例示的なテキストの行 1 0 は、重要談話構成要素 9 1 0 として示されている。談話構成要素は、統計、キーワード、知識ベース又は他の任意の重要性の非構造基準に基づいて重要であると決定されてよい。

40

【 0 1 4 3 】

図 1 5 は、本発明による関連性スコアを浸透する第 1 の例示的な方法を示す談話の例示的な構造表現である。重要談話構成要素 1 0 1 8 は、なんらかの関連性の非構造基準に基づいて識別される。矢印は R 値「1」により示される関連性スコア情報の談話の構造表現を通る浸透を示している。重要談話構成要素 1 0 1 8 の関連性スコアは子であるため、より関連性のある子の談話構成要素の関連性スコアは親の談話構成要素ノードに割り当てられる。選択された関連性の非構造基準により決定されるように、重要な談話 1 0 1 8 は最も関連性のある談話構成要素ノードであるため、重要談話構成要素 1 0 1 8 の関連性スコアは親の従属ノード 1 0 1 6 に割り当てられる。親のノードの関連性スコアは、順に、同じ理論的根拠に基づいて等位ノード 1 0 1 2、従属ノード 1 0 1 0、等位ノード 1 0 0 2

50

及びルート従属ノード 1 0 1 6 に割り当てられる。

【 0 1 4 4 】

従属談話構成要素ノード 1 0 0 1、1 0 0 4、1 0 1 1、1 0 1 4 及び 1 0 1 7 のそれぞれには、矢印により示されている被従属談話構成要素の関連性スコアが割り当てられる。最も関連性のある子のノードの関連性スコアは、等位ノードのそれぞれに割り当てられる。等位ノード 1 0 0 2 と 1 0 1 2 は、重要談話構成要素の親のノードとしてのそれらのステータスに基づいて含まれる。次に関連性スコアは識別された等位ノード 1 0 0 2 と 1 0 1 2 のそれぞれの子達に浸透される。このようにして、従属ノード 1 0 1 3 と 1 0 1 1 及びリーフノード 1 0 1 9 と 1 0 2 0 にも関連性スコアが割り当てられる。最も関連性のある関連性スコアの各談話構成要素がハイブリッドテキスト要約に選択される。

10

【 0 1 4 5 】

図 1 6 は、本発明に従って関連性スコアを浸透する第 2 の例示的な方法を示す談話の例示的な構造表現である。重要談話構成要素 1 0 1 6 は、一つ又は複数の関連性の非構造基準に基づいて決定される。浸透は等位構造及び二分の子のノードの処理において前述した方法とは異なる。等位ノード又は二分ノードの子達のより関連性のある関連性スコアはそれぞれの先行する同位に割り当てられる。これにより、ハイブリッドテキスト要約 9 2 0 に含まれる談話構成要素の数が削減する。

【 0 1 4 6 】

図 1 7 は、本発明に従って逆数スコアを浸透する第 3 の例示的な方法を示す談話の例示的な構造表現の第 1 の図である。例示的なテキストの行 4 と 5 は例示的な目的のために省略されている。談話構成要素のそれぞれは、被従属エッジに 1 を加えたものの数に基づいて従属レベル L でマーキングされる。各リーフノードには、従属レベルの逆数に等しいスコア「 i 」が割り当てられる。

20

【 0 1 4 7 】

$$i = 1 / L \quad (2)$$

【 0 1 4 8 】

次に、スコアの低い方の値が、以下の規則に基づき談話の構造表現を通して浸透される。スコア「 i 」の任意のノードでは、親がさらに低いスコアを有さない場合にはスコアを親に割り当てる。スコア「 i 」の任意の従属ノードでは、従属の子のノードのスコアがさらに低くない場合には従属子のノードにスコアを割り当てる。スコア「 i 」の任意の等位ノード又は二分ノードでは、子のノードがさらに低いスコアを有さない場合、各子のノードにスコア「 i 」を割り当てる。関連性スコアは、次に以下のようなスコア及び従属レベルに基づいて決定される。

30

【 0 1 4 9 】

$$r = 1 - (L - 1) * i \quad (3)$$

【 0 1 5 0 】

ここで、 r は関連性スコアであり、 L は従属レベルであり、「 i 」はスコア $1 / L$ である。

【 0 1 5 1 】

図 1 8 は、本発明に従って逆数スコアを浸透する第 3 の例示的な方法を示す談話の例示的な構造表現の第 2 の図である。次に関連付けられる関連性スコアは、式 (3) 又は他の関数に基づいて決定される。

40

【 0 1 5 2 】

図 1 9 は、本発明による閾値重要性レベルに基づいてハイブリッドテキスト要約に談話構成要素を選択する例示的な方法を示す談話の例示的な構造表現の第 3 の図である。例示的なテキストの行 4 と 5 は、例示的な目的のために省略されている。「 $1 / 2$ 」の重要性閾値関連性レベルが指定される。ハイブリッドテキスト要約 9 3 0 は、「 $1 / 2$ 」閾値関連性レベルを超えて関連性スコアと関連付けられる談話構成要素から構成される。

【 0 1 5 3 】

図 2 0 は、本発明に従った照応解決に基づいて関連性スコアの浸透を示す談話の例示的

50

な構造表現の第1図である。行5は、重要談話構成要素として識別される。この方法では、関連性スコアは、談話構成要素関連性スコアの調整が、談話構成要素が選択された談話構成要素リーフノードの対象を潜在的に解決できる場合にだけ発生するように限定的に浸透される。このようにして、語「温かいスープ(a hot soup)」が行5の「そのスープ(the

【0154】

soup)」に対する参照を潜在的に解決するために、行4が追加される。しかしながら、行4を加えると、対象「それらは(they)」を導く。語「それらは(they)」は、行3に関連性スコアを浸透することにより解決される。行3は、対象を一意的に識別する語「麺類(Noodles)」を含む。本発明による多様な他の例示的な実施形態では、語「麺類(Noodles)」はさらに圧縮されたハイブリッドテキスト要約923を形成するために行4に代入されてよい。

10

【0155】

図3及び図5に説明されるハイブリッド要約システム100、110の回路10~90のそれぞれは、適切にプログラミングされた汎用コンピュータの部分として実現できる。代わりに、前記に概略されたハイブリッド要約システム100、110の10~90は、ASIC内の物理的に別個のハードウェア回路として、つまりFPGA、PDL、PLA又はPALを使用して、あるいは離散論理要素又は離散回路要素を使用して実現できる。前記に概略されたハイブリッド要約システム100、110の回路10~90のそれぞれが採る特定の形式は設計選択肢であり、当業者にとって明白であり予測できるであろう。

【0156】

20

さらに、ハイブリッド要約システム100、110及び/又は前述された多様な回路のそれぞれは、それぞれソフトウェアルーチン、プログラミングされた汎用コンピュータで実行するマネージャ又はオブジェクト、特殊目的コンピュータ、マイクロプロセッサ等として実現できる。このケースでは、ハイブリッド要約システム100、110及び/又は前述された多様な回路のそれぞれは、それぞれ、サーバ等上に常駐するリソースとして、通信網に埋め込まれる一つ又は複数のルーチンとして実現できる。ハイブリッド要約システム100、110及び前述された多様な回路は、ウェブサーバ又はクライアントデバイスのハードウェアシステム及びソフトウェアシステムなどのソフトウェア及び/又はハードウェアシステムの中にハイブリッド要約システム100、110を物理的に組み込むことによって実現できる。

30

【0157】

図3及び図5に図示されるように、メモリ20は、改変可能な揮発性又は不揮発性のメモリ又は改変不可つまり固定されたメモリの任意の適切な組み合わせを使用して実現できる。改変可能なメモリは、揮発性であるか、不揮発性であるかに関係なく、スタティック又はダイナミックRAM、フロッピー(登録商標)ディスクとディスクドライブ、書き込み可能又は再書き込み可能の光ディスクとディスクドライブ、ハードディスクドライブ、フラッシュメモリ等の一つ又は複数を使用して実現できる。同様に、改変不可つまり固定されたメモリはROM、PROM、EPROM、EEPROM、CD-ROM又はDVD-ROMディスクなどの光ROMディスク、及びディスクドライブ等の任意の一つ又は複数を使用して実現できる。

40

【0158】

図1、図3及び図5に示される通信リンク99は、それぞれ直接ケーブル接続、ワイドエリアネットワーク又はローカルエリアネットワーク上の接続、イントラネット上の接続、インターネット上の接続、又は他の分散処理ネットワーク又はシステム上の接続を含む、ハイブリッド要約システム100、110に通信装置を接続するための既知の又は今後開発される装置又はシステムである場合がある。一般的には、通信リンク99は装置を接続し、通信を容易にするために有用な任意の既知の又は今後開発される接続システム又は構造である場合がある。

【0159】

さらに、通信リンク99がネットワークに対する有線リンク又は無線リンクである場合

50

があることが理解される必要がある。ネットワークはローカルエリアネットワーク、ワイドエリアネットワーク、イントラネット、インターネット、又は他の分散処理記憶ネットワークである場合がある。

【 0 1 6 0 】

本発明は上記概略された例示的な実施形態とともに説明されてきたが、多くの代替策、変型及び変形が当業者に明らかになるであろうことは確かである。したがって、上述されたような本発明の例示的な実施形態は、制限的ではなく例証となることが意図される。多様な変更は本発明の精神及び範囲を逸脱することなく加えられてよい。

【図面の簡単な説明】

【 0 1 6 1 】

【図 1】本発明による例示的なハイブリッド要約システムの概要である。

【図 2】本発明の例示的な実施形態によるハイブリッドテキスト要約を決定する第 1 の方法である。

【図 3】本発明の例示的な実施形態による第 1 のハイブリッド要約システムである。

【図 4】本発明の例示的な実施形態によるハイブリッドテキスト要約を決定する第 2 の方法である。

【図 5】本発明の例示的な実施形態による第 2 のハイブリッド要約システムである。

【図 6】本発明の実施形態による分割規則を記憶するための例示的なデータ構造である。

【図 7】本発明の実施形態による結合規則情報を記憶するための例示的なデータ構造である。

【図 8】本発明の実施形態による関連性スコアを浸透する第 1 の例示的な方法の展開図である。

【図 9】本発明の実施形態による関連性スコアを浸透する第 2 の例示的な方法の展開図である。

【図 10】本発明の例示的な実施形態による関連性スコアを浸透する第 3 の例示的な方法の展開図である。

【図 11】本発明の実施形態による照応置換の第 1 の例示的な方法の展開図である。

【図 12】本発明の実施形態による照応置換の第 2 の例示的な方法の展開図である。

【図 13】要約されるテキストである。

【図 14】要約されるテキストのための談話の例示的な構造表現である。

【図 15】本発明による関連性スコアを浸透する第 1 の例示的な方法を示す談話の例示的な構造表現である。

【図 16】本発明による関連性スコアを浸透する第 2 の例示的な方法を示す談話の例示的な構造表現である。

【図 17】本発明による逆スコアを浸透する第 3 の例示的な方法を示す談話の例示的な構造表現の第 1 の図である。

【図 18】本発明による逆スコアを浸透する第 3 の例示的な方法を示す談話の例示的な構造表現の第 2 の図である。

【図 19】本発明による閾値重要度レベルに基づいたハイブリッドテキスト要約のために談話構成要素を選択する例示的な方法を示す談話の例示的な構造表現である第 3 の図である。

【図 20】本発明による照応解決に基づいた関連性スコアの浸透を示す談話の例示的な構造表現の第 1 の図である。

【符号の説明】

【 0 1 6 2 】

1 0 0 ハイブリッド要約システム

2 0 0 情報リポジトリ

3 0 0 ウェブ使用可能パーソナル・コンピュータ

4 0 0 ウェブ使用可能タブレット・コンピュータ

1 0 0 0 0、1 0 0 0 1、1 0 0 0 2 文書

10

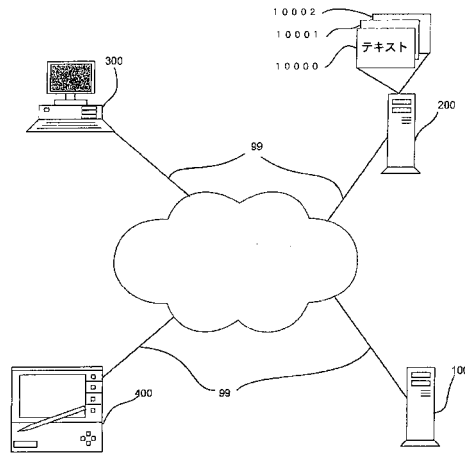
20

30

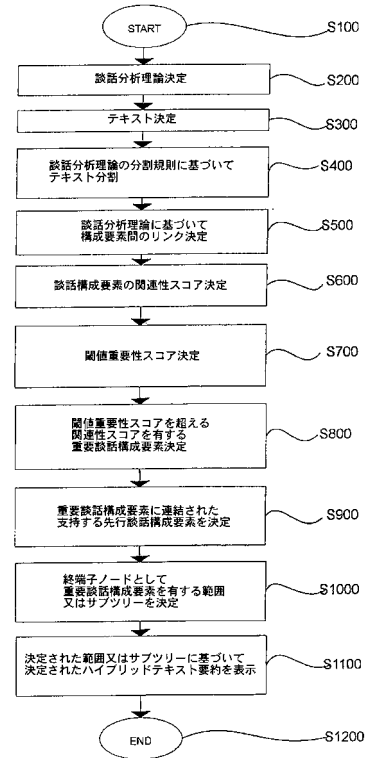
40

50

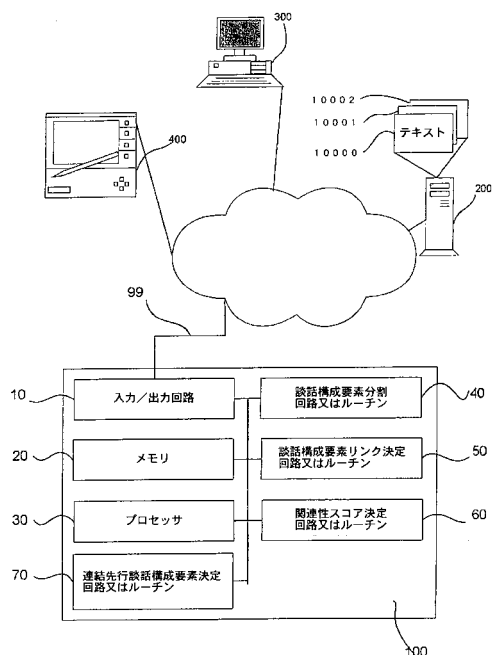
【図 1】



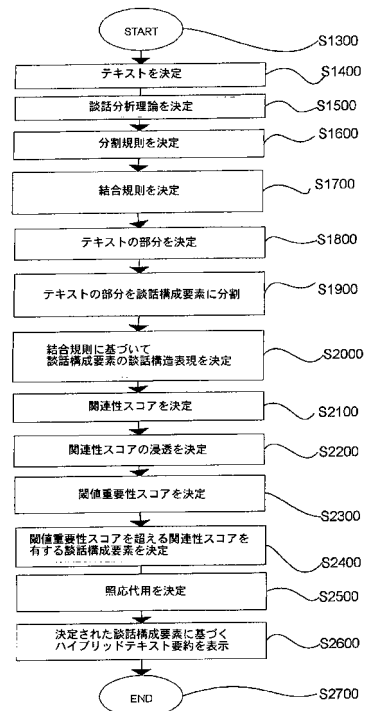
【図 2】



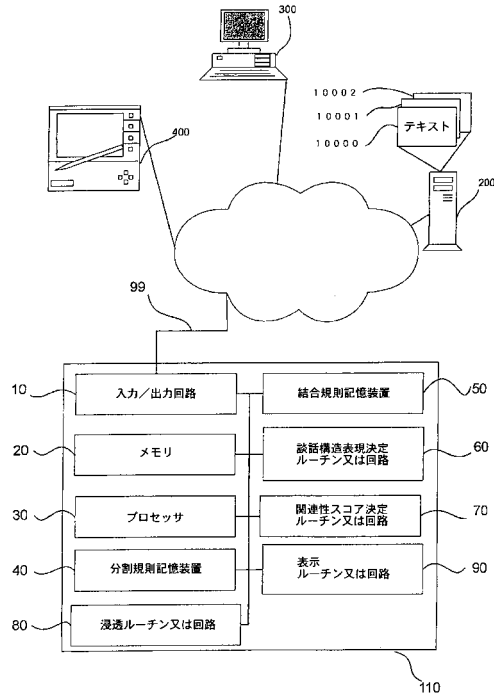
【図 3】



【図 4】



【図 5】



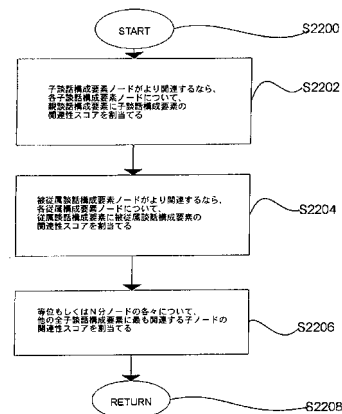
【図 6】

番号 (ID)	型 (TYPE)	パターン (PATTERN)	動作 (ACTION)
1	SIMPLE_EVENTS	SIMPLE_EVENT	BDU=TRUE
2	DISCLOSURE_OPERATOR	MODIFIER	BDU=TRUE
3	INTERPOLATIONS	APPOSITIVE OR PARENTHETICAL OR INTERRUPTION OR NON RESTRICTIVE_CLAUSE OR REFERENCE	BDU=TRUE
4	SUBORDINATE_CLAUSES	COMPLEMENT_CLAUSE OR SENTENTIAL SUBJECT OR RELATIVE_CLAUSE	BDU=TRUE
5	POSTNOMINAL_MODIFIER	(DISCRETE_EVENT AND (POSTNOMINAL_MODIFIER OR RELATIVE_CLAUSE OR PARTCIPAL_MODIFIER))	BDU=TRUE
6	MORPHOLOGICAL_FORM	(DISCRETE_EVENT AND (INFINITIVES OR GERUNDS))	BDU=TRUE
7	OTHER_PREDICATION	(COPULAR_PREDICATION OR SECONDARY_PREDICATION OR DISCRETE_EVENT)	BDU=TRUE
8	SUBORDINATING_CONJUNCTIONS	(SUBORDINATING_CONJUNCTION AND DISCRETE_EVENT)	BDU=TRUE
9	COORDINATING_CONJUNCTIONS	((COORDINATING_CONJUNCTION AND (LENGTH = 1) OR (X COORDINATING_CONJUNCTION Y) AND (X ISA BDU AND Y ISA BDU))	BDU=TRUE
10	INFORMATION_STRUCTURE	(COPULA AND (CLEFT OR PSEUDO_CLEFT))	BDU=FALSE
11	EVENT_MODIFIER	(EVENT_MODIFIER AND NOT(DISCRETE_EVENT))	BDU=FALSE
14	INITIAL_MODIFIER	(INITIAL_MODIFIER AND SCENE_SETTING)	BDU=TRUE

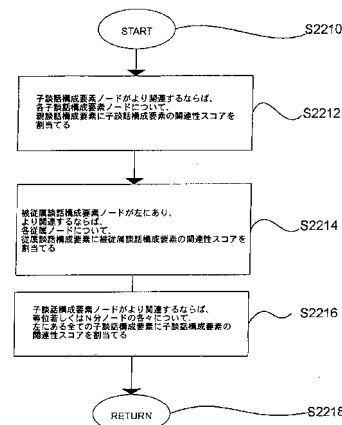
【図 7】

識別子 (ID)	規則 (RULE)
1	< Sx: (M-BDU/TYPE) = "BetweenParenthesis"; St: AP is BottomOfTree; >→ Subordinate(AP, M-BDU)
2	< Sx: (AP/Tense) = "Present"; Sx: (M-BDU/Tense) = "Past"; >→ Subordinate(AP, M-BDU)
3	< LS: (AP) isa (KB/Event); LS: (M-BDU) isa (KB/Event); Sx: (AP/Tense) = (M-BDU/Tense); >→ Coordinate(AP, M-BDU)
4	< LS: (AP/Lexeme) hasa (M-BDU/Lexeme): \$1: \$2; Sx: AP/COMP is \$1; Sx: (M-BDU/COMP) is \$2; >→ Subordinate(AP, M-BDU)
5	< Sx: (M-BDU/Type) = "Temporal"; St: M-BDU isSubordinatedTo AP; >→ CreateNary(Cx, AP, M-BDU)
6	< Sm: Σ(AP) is "Generic"; Sm: Σ(M-BDU) is "Specific"; >→ Subordinate(AP, M-BDU)
7	< Sm: Σ(AP) is "Irealis"; Sm: Σ(M-BDU) is "Realis"; Sx: if AP then M-BDU; >→ CreateNary(HYP, AP, M-BDU)
8	< Sm: Σ(AP) is "Irealis"; Sm: Σ(M-BDU) is "Realis"; Sx: if AP but instead M-BDU; >→ CreateNary(HYP, AP, M-BDU)
9	< Sx: AP but M-BDU >→ Coordinate(AP, M-BDU);
10	< Sm: Σ(AP) is "Generic": \$1; Sm: Σ(M-BDU) is "Generic": \$2; LS: \$1 and \$2; >→ Subordinate(AP, M-BDU);
11	< Sm: PointOfView(Σ(AP)) != PointOfView(Σ(M-BDU)); >→ Subordinate(AP, M-BDU)
12	< Sm: Σ(AP) is "Narrative"; Sm: Σ(M-BDU) is "Narrative"; Px: ifnot AnyRule(LS) applies; >→ Coordinate(AP, M-BDU);
99	< St: AP is BottomOfTree; Px: ifnot AnyRule applies; >→ Subordinate(AP, M-BDU)

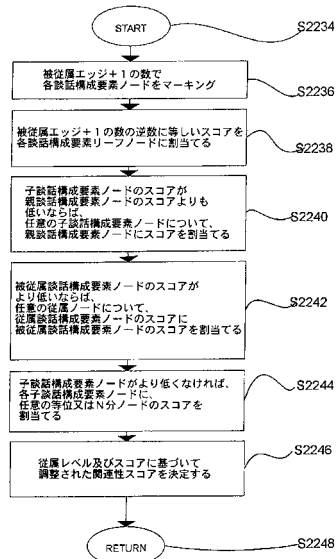
【図 8】



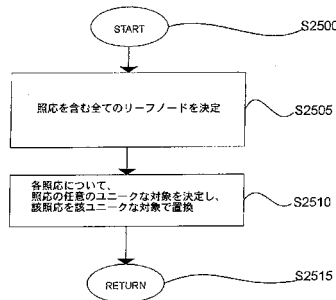
【図 9】



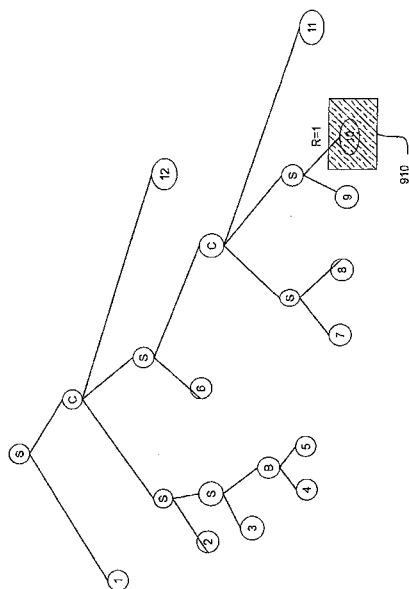
【 図 1 0 】



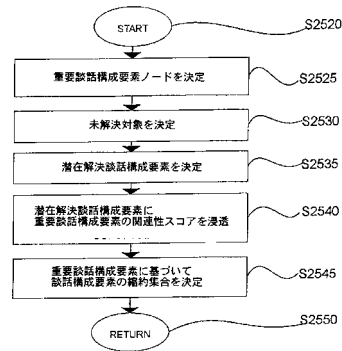
【 図 1 1 】



【 図 1 4 】



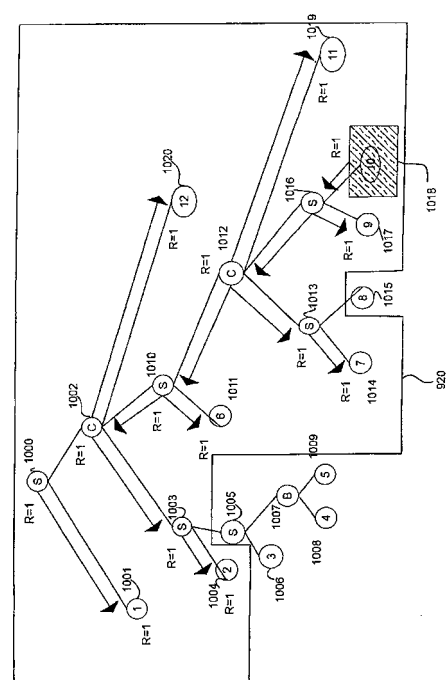
【 図 1 2 】



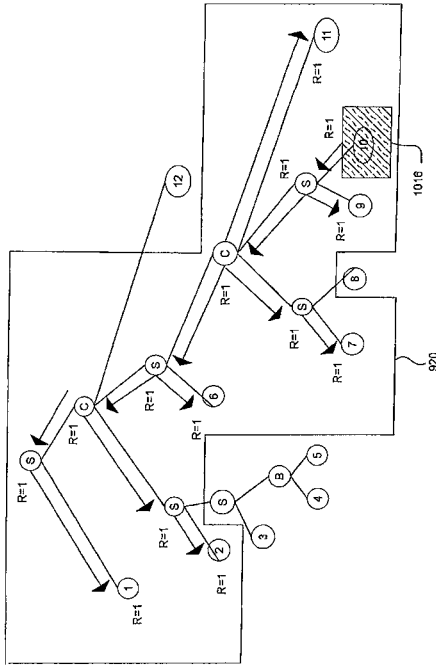
【 図 1 3 】

810	820
830	840
850	860
870	880
890	900
910	920
930	940
950	960
970	980
990	1000
1010	1020
1030	1040
1050	1060
1070	1080
1090	1100
1110	1120
1130	1140
1150	1160
1170	1180
1190	1200
1210	1220
1230	1240
1250	1260
1270	1280
1290	1300
1310	1320
1330	1340
1350	1360
1370	1380
1390	1400
1410	1420
1430	1440
1450	1460
1470	1480
1490	1500
1510	1520
1530	1540
1550	1560
1570	1580
1590	1600
1610	1620
1630	1640
1650	1660
1670	1680
1690	1700
1710	1720
1730	1740
1750	1760
1770	1780
1790	1800
1810	1820
1830	1840
1850	1860
1870	1880
1890	1900
1910	1920
1930	1940
1950	1960
1970	1980
1990	2000
2010	2020
2030	2040
2050	2060
2070	2080
2090	2100
2110	2120
2130	2140
2150	2160
2170	2180
2190	2200
2210	2220
2230	2240
2250	2260
2270	2280
2290	2300
2310	2320
2330	2340
2350	2360
2370	2380
2390	2400
2410	2420
2430	2440
2450	2460
2470	2480
2490	2500
2510	2520
2530	2540
2550	2560
2570	2580
2590	2600
2610	2620
2630	2640
2650	2660
2670	2680
2690	2700
2710	2720
2730	2740
2750	2760
2770	2780
2790	2800
2810	2820
2830	2840
2850	2860
2870	2880
2890	2900
2910	2920
2930	2940
2950	2960
2970	2980
2990	3000
3010	3020
3030	3040
3050	3060
3070	3080
3090	3100
3110	3120
3130	3140
3150	3160
3170	3180
3190	3200
3210	3220
3230	3240
3250	3260
3270	3280
3290	3300
3310	3320
3330	3340
3350	3360
3370	3380
3390	3400
3410	3420
3430	3440
3450	3460
3470	3480
3490	3500
3510	3520
3530	3540
3550	3560
3570	3580
3590	3600
3610	3620
3630	3640
3650	3660
3670	3680
3690	3700
3710	3720
3730	3740
3750	3760
3770	3780
3790	3800
3810	3820
3830	3840
3850	3860
3870	3880
3890	3900
3910	3920
3930	3940
3950	3960
3970	3980
3990	4000
4010	4020
4030	4040
4050	4060
4070	4080
4090	4100
4110	4120
4130	4140
4150	4160
4170	4180
4190	4200

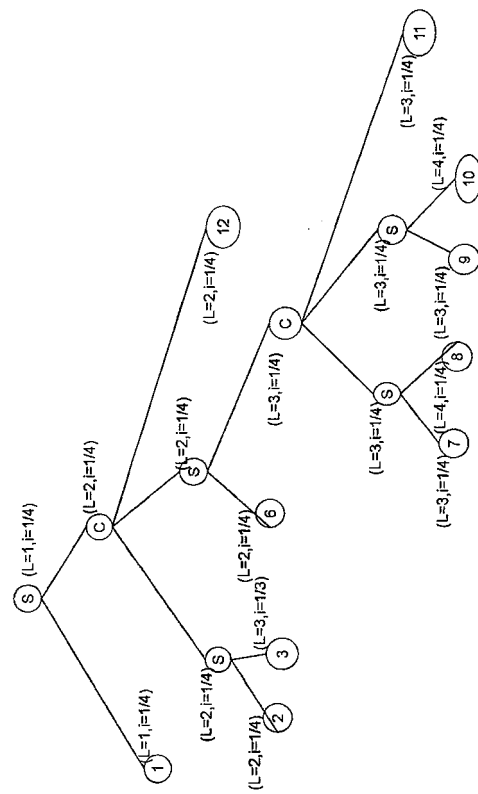
【 図 1 5 】



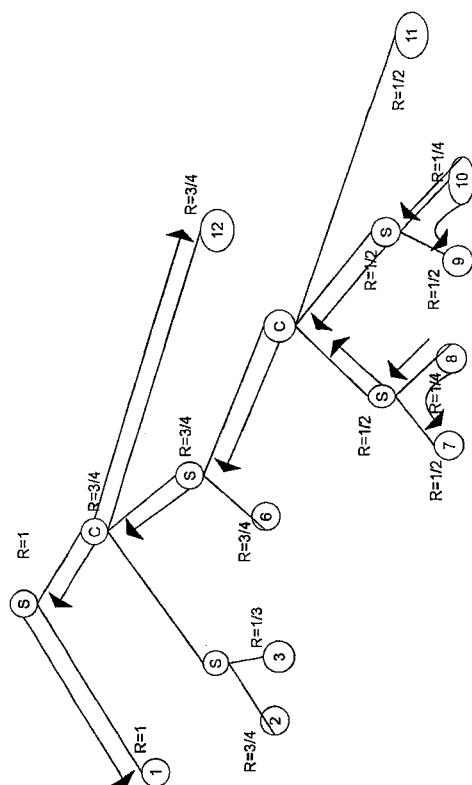
【 図 1 6 】



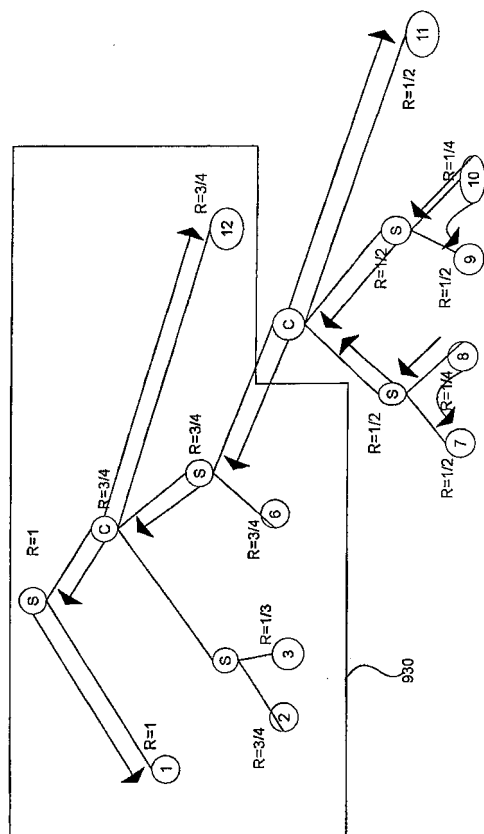
【圖 17】



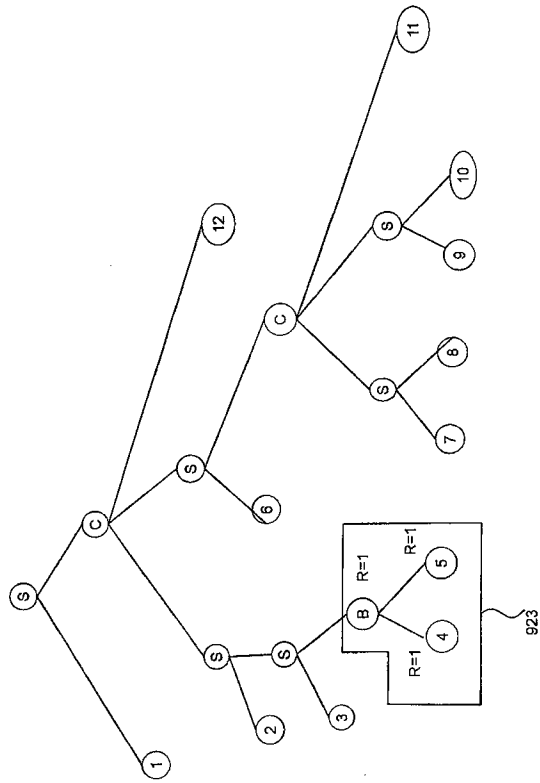
【 図 1 8 】



【 図 1 9 】



【図 20】



 フロントページの続き

- (72)発明者 マーティン ヘンク ヴァン デン バーグ
 アメリカ合衆国 94306 カリフォルニア州 パロ アルト ヴィラ ヴィスタ 4037
- (72)発明者 ジョバンニ エル・ティオーネ
 アメリカ合衆国 94114 カリフォルニア州 サンフランシスコ エイティーンズ ストリート 4533 ナンバー2
- (72)発明者 リチャード エス・クラウチ
 アメリカ合衆国 95014 カリフォルニア州 クパティーノ シーニック ブルバード 10290
- (72)発明者 クリストファー ディー・カリー
 アメリカ合衆国 94043 カリフォルニア州 マウンテン ヴュー ネドソン コート 2513
- (72)発明者 デイビッド ディー・アーン
 アメリカ合衆国 94306 カリフォルニア州 パロ アルト ウィリアムズ ストリート 2052

審査官 岩田 淳

- (56)参考文献 特公平07-011801(JP, B2)
 特開平11-345233(JP, A)
 特開平06-348696(JP, A)
 特開2002-091276(JP, A)
 特開2003-085193(JP, A)
 国際公開第2001/086491(WO, A1)

- (58)調査した分野(Int.Cl., DB名)
 G06F 17/30
 G06F 17/21
 G06F 17/27

- (54)【発明の名称】ハイブリッドテキスト要約を決定する方法、システム、及び、制御プログラム、談話構文解析方法、システム、及び、該システムにおける方法、談話構成要素にテキストを分割する方法及びシステム、談話の構造表現を決定する方法及びシステム、ハイブリッドテキスト要約システム