(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2011-170790
(P2011-170790A)
(43)公開日　平成23年9月1日(2011.9.1)

(51)Int.Cl.		F I		テーマコード（参考）
G06F 17/27	(2006.01)	G06F 17/27	Z	5B091

審査請求 未請求　請求項の数 10　ＯＬ　（全 26 頁）

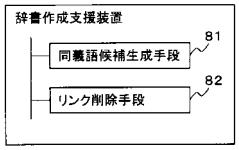
(21)出願番号　特願2010-36415(P2010-36415)
(22)出願日　平成22年2月22日(2010.2.22)

(71)出願人　000004237
日本電気株式会社
東京都港区芝五丁目７番１号
(71)出願人　899000068
学校法人早稲田大学
東京都新宿区戸塚町１丁目１０４番地
(74)代理人　100103090
弁理士　岩壁　冬樹
(74)代理人　100124501
弁理士　塩川　誠人
(72)発明者　立石　健二
東京都港区芝五丁目７番１号　日本電気株式会社内

最終頁に続く

(54)【発明の名称】辞書作成支援装置、辞書作成支援方法及び辞書作成支援プログラム

(57)【要約】
【課題】生成される同義語候補の精度を向上させて辞書作成を支援できる辞書作成支援装置を提供する。
【解決手段】同義語候補生成手段８１は、Ｗｅｂページの識別子である各資源位置指定子に対してリンクする文字列を表すアンカーテキストを用いて、同義語を生成する対象の語として入力される入力語の同義語候補を生成する。リンク削除手段８２は、一の資源位置指定子に対して、アンカーテキストが入力語もしくは同義語候補の中でその入力語の同義語と判定された同義語候補になっているリンクである第一のリンクと、アンカーテキストが同義語候補の中で入力語の非同義語と判定された同義語候補になっているリンクである第二のリンクのうち、少なくとも一方のリンクを削除する。このとき、リンク削除手段８２は、アンカーテキストごとのリンクの数に基づいて、上記一の資源位置指定子とアンカーテキストとのリンクを削除する。
【選択図】図１５



【特許請求の範囲】
【請求項１】
辞書作成を支援する辞書作成支援装置であって、
Ｗｅｂページを識別する識別子である各資源位置指定子に対してリンクする文字列を表すアンカーテキストを用いて、同義語を生成する対象の語として当該辞書作成支援装置に入力される入力語の同義語候補を生成する同義語候補生成手段と、
一の資源位置指定子に対して、前記アンカーテキストが入力語もしくは前記同義語候補の中で当該入力語の同義語と判定された同義語候補になっているリンクである第一のリンクと、前記アンカーテキストが前記同義語候補の中で入力語の非同義語と判定された同義語候補になっているリンクである第二のリンクのうち、少なくとも一方のリンクを削除するリンク削除手段とを備え、
前記リンク削除手段は、前記一の資源位置指定子に対するアンカーテキストごとのリンクの数に基づいて、当該一の資源位置指定子と前記アンカーテキストとのリンクを削除し、
同義語候補決定手段は、前記リンク削除手段が削除する対象から除かれたリンクのアンカーテキストを用いて入力語の同義語候補を生成する
ことを特徴とする辞書作成支援装置。
【請求項２】
資源位置指定子に対してリンクするアンカーテキストが当該資源位置指定子の実体を表す確率である実体確率を、資源位置指定子とアンカーテキストとのリンクごとに計算する実体確率計算手段を備え、
リンク削除手段は、第一のリンクと第二のリンクのうち、前記実体確率が小さいリンクを削除する
請求項１記載の辞書作成支援装置。
【請求項３】
実体確率計算手段は、各アンカーテキストから一の資源位置指定子へのリンクの総数に対する一のアンカーテキストから当該一の資源位置指定子へのリンクの数の割合を実体確率として算出する
請求項２記載の辞書作成支援装置。
【請求項４】
実体確率計算手段は、各アンカーテキストから一の資源位置指定子へのリンクの総数に対する一のアンカーテキストから当該一の資源位置指定子へのリンクの数の割合、及び、前記アンカーテキストを持つ各資源位置指定子へのリンクの総数に対する当該アンカーテキストから前記一の資源位置指定子へのリンクの数の割合を用いて実体確率を算出する
請求項２記載の辞書作成支援装置。
【請求項５】
アンカーテキストが同義語候補になっている各資源位置指定子に対するリンクのうち、入力語の同義語と判定された同義語候補がアンカーテキストであるリンクを統合するリンク統合手段を備え、
同義語候補決定手段は、統合された前記リンクのアンカーテキストを用いて入力語の同義語候補を生成する
請求項１から請求項４のうちのいずれか１項に記載の辞書作成支援装置。
【請求項６】
リンク削除手段は、第一のリンクと第二のリンクの双方が一の資源位置指定子に存在しない場合に、当該一の資源位置指定子に対するアンカーテキストのリンクを削除対象から除く
請求項１から請求項５のうちのいずれか１項に記載の辞書作成支援装置。
【請求項７】
辞書作成を支援する辞書作成支援方法であって、
Ｗｅｂページを識別する識別子である各資源位置指定子に対してリンクする文字列を表

すアンカーテキストを用いて、同義語を生成する対象の語として入力される入力語の同義語候補を生成し、
一の資源位置指定子に対して、前記アンカーテキストが入力語もしくは前記同義語候補の中で当該入力語の同義語と判定された同義語候補になっているリンクである第一のリンクと、前記アンカーテキストが前記同義語候補の中で入力語の非同義語と判定された同義語候補になっているリンクである第二のリンクのうち、少なくとも一方のリンクを削除し、
前記リンクを削除する際に、前記一の資源位置指定子に対するアンカーテキストごとのリンクの数に基づいて、当該一の資源位置指定子と前記アンカーテキストとのリンクを削除し、
同義語候補を生成する際に、削除対象から除かれたリンクのアンカーテキストを用いて入力語の同義語候補を生成する
ことを特徴とする辞書作成支援方法。
【請求項８】
資源位置指定子に対してリンクするアンカーテキストが当該資源位置指定子の実体を表す確率である実体確率を、資源位置指定子とアンカーテキストとのリンクごとに計算し、
リンクを削除する際に、第一のリンクと第二のリンクのうち、前記実体確率が小さいリンクを削除する
請求項７記載の辞書作成支援方法。
【請求項９】
辞書作成を支援するコンピュータに搭載される辞書作成支援プログラムであって、
前記コンピュータに、
Ｗｅｂページを識別する識別子である各資源位置指定子に対してリンクする文字列を表すアンカーテキストを用いて、同義語を生成する対象の語として当該コンピュータに入力される入力語の同義語候補を生成する同義語候補生成処理、および、
一の資源位置指定子に対して、前記アンカーテキストが入力語もしくは前記同義語候補の中で当該入力語の同義語と判定された同義語候補になっているリンクである第一のリンクと、前記アンカーテキストが前記同義語候補の中で入力語の非同義語と判定された同義語候補になっているリンクである第二のリンクのうち、少なくとも一方のリンクを削除するリンク削除処理を実行させ、
前記リンク削除処理で、前記一の資源位置指定子に対するアンカーテキストごとのリンクの数に基づいて、当該一の資源位置指定子と前記アンカーテキストとのリンクを削除させ、
同義語候補決定処理で、前記リンク削除処理で削除する対象から除かれたリンクのアンカーテキストを用いて入力語の同義語候補を生成させる
ことを特徴とする辞書作成支援プログラム。
【請求項１０】
コンピュータに、
資源位置指定子に対してリンクするアンカーテキストが当該資源位置指定子の実体を表す確率である実体確率を、資源位置指定子とアンカーテキストとのリンクごとに計算する実体確率計算処理を実行させ、
リンク削除処理で、第一のリンクと第二のリンクのうち、前記実体確率が小さいリンクを削除させる
請求項９記載の辞書作成支援プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、Ｗｅｂのアンカーテキストを用いて入力語に対する同義語候補を生成する辞書作成支援装置、辞書作成支援方法及び辞書作成支援プログラムに関する。
【背景技術】

【０００２】
同義語辞書は、文書検索、顧客データの名寄せなど、様々なソフトウェアの基本的な資源として使用される。同義語の定義としては様々なものが存在するが、ここでは、表記が異なり、同じ対象物を示す２つの語を同義語とする。
【０００３】
非特許文献１には、同義語辞書の作成支援方法として、利用者が入力した語の同義語候補をＷｅｂ（ウェブ）のアンカーテキストを用いて生成する方法が開示されている。非特許文献１に記載された方法は、あるＷｅｂページを示すＵＲＬに対する複数のアンカーテキストは、それぞれが同様の表現を含んでいると判断されることから、それらの表現を同義であるとみなすものである。
【先行技術文献】
【非特許文献】
【０００４】
【非特許文献１】WEN-HSIANG LU, LEE-FENG CHIEN, HIS-JIAN LEE, Anchor Text Mining for Translation of Web Queries: A Transitive Translation Approach, ACM Transactions on Information Systems, Vol. 22, No. 2, pp. 242-269, 2004.
【発明の概要】
【発明が解決しようとする課題】
【０００５】
例えば、非特許文献１に記載された方法を用いることで、同義語候補を生成することが可能である。以下、非特許文献１に記載された方法を利用した同義語候補の生成方法について説明する。
【０００６】
今、ＵＲＬ（Uniform Resource Locator）ｕに対して２種類のアンカーテキストｓ及びｔが存在するとする。ここでアンカーテキストとは、ＵＲＬｕへリンクするＷｅｂ文書におけるリンク上の文字列を表す。つまり、アンカーテキストｓは、ｕに対してｓという文字列で別のＷｅｂ文書からリンクしていることになる。また、以下の説明では、アンカーテキストｓが、ＵＲＬｕへリンクするリンク上の文字列であることを、ＵＲＬｕがアンカーテキストｓを持つと表現することもある。
【０００７】
次に、Ｐ（ｘ｜ｕ）をＵＲＬｕにリンクするアンカーテキストｘがＵＲＬｕの実体を表す確率と定義する。例えば、ＵＲＬｕが「ｗｗｗ．ｎｅｃ．ｃｏ．ｊｐ」であるとする。ここで、アンカーテキストｘが「日本電気」の場合、アンカーテキストｘは、ＵＲＬｕの実体を表すので、Ｐ（日本電気｜ｗｗｗ．ｎｅｃ．ｃｏ．ｊｐ）＝１になることが理想である。言い換えると、Ｐ（ｘ｜ｕ）は、アンカーテキストｘがＵＲＬｕによって示す対象をどれだけ正確に表しているかを示す指標である。
【０００８】
このＰ（ｘ｜ｕ）を用いて、アンカーテキストｓとアンカーテキストｔの同義性Ｒｅｌ（ｓ，ｔ）を次の式１で定義する。ここで、同義性とは、表記が異なる２つの語が同じ対象物を示す尤もらしさのことを言う。
【０００９】
Ｒｅｌ（ｓ，ｔ）＝Ｅ［Ｐ（ｓ｜ｕ）＊Ｐ（ｔ｜ｕ）］＿ｕ　　（式１）
【００１０】
なお、Ｐ（ｓ｜ｕ）＊Ｐ（ｔ｜ｕ）は、アンカーテキストｓとアンカーテキストｔの両方がＵＲＬｕの実体を表す確率を意味する。今、ＵＲＬの出現確率を全て一様とすれば、同義性Ｒｅｌ（ｓ，ｔ）は、以下の式２により算出される。
【００１１】
Ｒｅｌ（ｓ，ｔ）＝Σｉ［Ｐ（ｓ｜ｕ＿ｉ）＊Ｐ（ｔ｜ｕ＿ｉ）］／Ｎ　（式２）
【００１２】
ここで、Ｎは、アンカーテキストｓもしくはアンカーテキストｔが出現するＵＲＬの数

である。式２による同義性Ｒｅｌ（ｓ，ｔ）の算出方法は、アンカーテキストｓ及びアンカーテキストｔが出現する全てのＵＲＬでそのＵＲＬの実体を表していれば、アンカーテキストｓとアンカーテキストｔとは同義語であると判断するという考えに基づく。

【００１３】

ところで、Ｐ（ｓ｜ｕ）＊Ｐ（ｔ｜ｕ）は、アンカーテキストｓとアンカーテキストｔの両方がＵＲＬｕの実体を表す確率であり、その確率の上限値は、アンカーテキストｓとアンカーテキストｔのいずれか一方がＵＲＬｕの実体を表す確率である。つまり、Ｐ（ｓ｜ｕ）＊Ｐ（ｔ｜ｕ）について、以下の式３の関係が成り立つ

【００１４】

Ｐ（ｓ｜ｕ）＊Ｐ（ｔ｜ｕ）≦１－｛（１－Ｐ（ｓ｜ｕ））＊（１－Ｐ（ｔ｜ｕ））｝
＝Ｐ（ｓ｜ｕ）＋Ｐ（ｔ｜ｕ）－Ｐ（ｓ｜ｕ）＊Ｐ（ｔ｜ｕ）　（式３）

【００１５】

したがって、同義性Ｒｅｌ（ｓ，ｔ）を正規化したＮＲｅｌ（ｓ，ｔ）は、下記の式４により算出される。

【００１６】

【数１】

$$NRel(s,t)=\frac{\Sigma i[P(s|u\_i)*P(t|u\_i)]}{\Sigma i[P(s|u\_i)+P(t|u\_i)-P(s|u\_i)*P(t|u\_i)]} \qquad (式４)$$

【００１７】

一方、Ｐ（ｘ｜ｕ）は、ＵＲＬｕへアンカーテキストｘでリンクする数を用いて推定される。具体的には、ＵＲＬｕに対するリンクの総数をＬ＿ｕ、ＵＲＬｕに対するアンカーテキストｘによるリンクの数（すなわち、アンカーテキストｘでＵＲＬｕにリンクする数）をＬ＿ｕ，ｘとする。このとき、Ｐ（ｘ｜ｕ）は、以下の式５により算出される。

【００１８】

Ｐ（ｘ｜ｕ）＝Ｌ＿ｕ，ｘ／Ｌ＿ｕ　（式５）

【００１９】

なお、式５により算出されるＰ（ｘ｜ｕ）は、「実体を表すアンカーテキストほど多くの人が用いる」ことを前提とした確率である。

【００２０】

図１６及び図１７を用いて、通常の辞書作成支援装置が正規化された同義性ＮＲｅｌを算出する方法について具体的に説明する。図１６は、ＵＲＬとアンカーテキストとのリンク情報を示す説明図である。また、図１７は、図１６に示すリンク情報をもとに同義性ＮＲｅｌを算出する過程を示す説明図である。

【００２１】

図１６に示すｕ１～ｕ４はＵＲＬを示し、ｓ１～ｓ４はアンカーテキストを示す。また、ＵＲＬｕ１～ｕ４とアンカーテキストｓ１～ｓ４とを結ぶ実線は、アンカーテキストｓ１～ｓ４がどのＵＲＬへリンクしているかを示す。また、表中の「Ｐ→」は、式５により算出されるＰ（ｘ｜ｕ）の値を示し、ＵＲＬｕ１～ｕ４の下部に記載された値は、アンカーテキストｓ１～ｓ４によるＵＲＬｕへのリンク数を示す。

【００２２】

例えば、利用者から入力語「ＮＥＣ」を受け付けると、辞書作成支援装置は、アンカーテキスト「ＮＥＣ」と他のアンカーテキスト間のＮＲｅｌを計算し、その値（ＮＲｅｌ）の上位を同義語候補として利用者に提示する。

【００２３】

辞書作成支援装置は、図１６に示すリンク情報をもとに式４を用いて同義性を示す値を算出する。具体的には、図１７に示すように、「ＮＥＣ」との同義性をそれぞれ、ＮＲｅｌ（ＮＥＣ，日本電気）＝０．２６、ＮＲｅｌ（ＮＥＣ，ＢＩＧＬＯＢＥ）＝０．０７、ＮＲｅｌ（ＮＥＣ，ビッグローブ）＝０．０３と算出する。そして、辞書作成支援装置は

、これらを同義語候補としてこの順序で利用者に提示する。利用者は、この結果を順番に閲覧したり、取捨選択したりして、同義語辞書を作成する。なお、ＮＥＣ、ＢＩＧＬＯＢＥ及びビッグローブは登録商標である。

【００２４】

しかし、上述の方法を用いて同義語辞書を作成する辞書作成支援装置では、Ｐ（ｘ｜ｕ）の値を正しく推定できない場合が存在する。例えば、アンカーテキストが誤って記述されたり、Ｗｅｂ文書の作成者がＵＲＬの実体を表す文字列の一部にアンカーテキストを設定したりする場合などである。

【００２５】

具体例を用いて説明する。例えば、Ｗｅｂ文書の記述者が、テキスト「ＮＥＣが提供するプロバイダー」を記述し、テキスト中の「ＮＥＣ」の部分をＵＲＬ「ｂｉｇｌｏｂｅ．ｎｅ．ｊｐ」へのアンカーテキストとして設定する場合がある。このような場合には、ＵＲＬ「ｂｉｇｌｏｂｅ．ｎｅ．ｊｐ」にリンクするアンカーテキスト「ＮＥＣ」が存在することになってしまい、Ｐ（ｘ｜ｕ）の値を正しく推定できなくなる。そのため、同義語候補を生成する精度が悪化してしまうという課題がある。

【００２６】

図１６に示すリンク情報を用いて、さらに説明する。図１６に示すアンカーテキスト「ＮＥＣ」は、実際には、ＵＲＬ「ｂｉｇｌｏｂｅ．ｎｅ．ｊｐ」の実体を表すものではない。そのため、本来は、Ｐ（ＮＥＣ｜ｂｉｇｌｏｂｅ．ｎｅ．ｊｐ）の値が０に近くなることが望ましい。しかし、誤った情報が用いられた場合、図１７に示すように、Ｐ（ＮＥＣ｜ｂｉｇｌｏｂｅ．ｎｅ．ｊｐ）の値は、０．１７と算出されてしまうことになる。同様に、アンカーテキスト「ＮＥＣ」は、実際、ＵＲＬ「ｎｅｃ．ｊｐ」の実体を表すものである。そのため、本来は、Ｐ（ＮＥＣ｜ｎｅｃ．ｊｐ）の値が１に近くなることが望ましい。しかし、誤った情報が用いられた場合、図１７に示すように、Ｐ（ＮＥＣ｜ｎｅｃ．ｊｐ）の値は、０．４と算出されてしまうことになる。

【００２７】

このように算出されたＰ（ｘ｜ｕ）を用いて同義性を算出すると、同義語と判断されるべき語の同義性ＮＲｅｌの値が低く、同義語と判断されるべきでない語の同義性ＮＲｅｌの値が高く算出されてしまう。この結果、精度の低い同義語候補が生成されてしまうという問題がある。

【００２８】

そこで、本発明は、生成される同義語候補の精度を向上させて辞書作成を支援できる辞書作成支援装置、辞書作成支援システム及び辞書支援作成プログラムを提供することを目的とする。

【課題を解決するための手段】

【００２９】

本発明による辞書作成支援装置は、辞書作成を支援する辞書作成支援装置であって、Ｗｅｂページを識別する識別子である各資源位置指定子に対してリンクする文字列を表すアンカーテキストを用いて、同義語を生成する対象の語としてその辞書作成支援装置に入力される入力語の同義語候補を生成する同義語候補生成手段と、一の資源位置指定子に対して、アンカーテキストが入力語もしくは同義語候補の中でその入力語の同義語と判定された同義語候補になっているリンクである第一のリンクと、アンカーテキストが同義語候補の中で入力語の非同義語と判定された同義語候補になっているリンクである第二のリンクのうち、少なくとも一方のリンクを削除するリンク削除手段とを備え、リンク削除手段が、一の資源位置指定子に対するアンカーテキストごとのリンクの数に基づいて、その一の資源位置指定子とアンカーテキストとのリンクを削除し、同義語候補決定手段が、リンク削除手段が削除する対象から除かれたリンクのアンカーテキストを用いて入力語の同義語候補を生成することを特徴とする。

【００３０】

本発明による辞書作成支援方法は、辞書作成を支援する辞書作成支援方法であって、Ｗ

ｅｂページを識別する識別子である各資源位置指定子に対してリンクする文字列を表すアンカーテキストを用いて、同義語を生成する対象の語として入力される入力語の同義語候補を生成し、一の資源位置指定子に対して、アンカーテキストが入力語もしくは同義語候補の中でその入力語の同義語と判定された同義語候補になっているリンクである第一のリンクと、アンカーテキストが同義語候補の中で入力語の非同義語と判定された同義語候補になっているリンクである第二のリンクのうち、少なくとも一方のリンクを削除し、リンクを削除する際に、一の資源位置指定子に対するアンカーテキストごとのリンクの数に基づいて、その一の資源位置指定子とアンカーテキストとのリンクを削除し、同義語候補を生成する際に、削除対象から除かれたリンクのアンカーテキストを用いて入力語の同義語候補を生成することを特徴とする。

【００３１】

本発明による辞書作成支援プログラムは、辞書作成を支援するコンピュータに搭載される辞書作成支援プログラムであって、コンピュータに、Ｗｅｂページを識別する識別子である各資源位置指定子に対してリンクする文字列を表すアンカーテキストを用いて、同義語を生成する対象の語としてそのコンピュータに入力される入力語の同義語候補を生成する同義語候補生成処理、および、一の資源位置指定子に対して、アンカーテキストが入力語もしくは同義語候補の中でその入力語の同義語と判定された同義語候補になっているリンクである第一のリンクと、アンカーテキストが同義語候補の中で入力語の非同義語と判定された同義語候補になっているリンクである第二のリンクのうち、少なくとも一方のリンクを削除するリンク削除処理を実行させ、リンク削除処理で、一の資源位置指定子に対するアンカーテキストごとのリンクの数に基づいて、その一の資源位置指定子とアンカーテキストとのリンクを削除させ、同義語候補決定処理で、リンク削除処理で削除する対象から除かれたリンクのアンカーテキストを用いて入力語の同義語候補を生成させることを特徴とする。

【発明の効果】

【００３２】

本発明によれば、生成される同義語候補の精度を向上させて辞書作成を支援できる。

【図面の簡単な説明】

【００３３】

【図１】本発明の第１の実施形態における辞書作成支援装置の例を示すブロック図である。

【図２】第１の実施形態における動作の例を示すフローチャートである。

【図３】リンク情報の例を示す説明図である。

【図４】リンク情報を削除する処理の例を示す説明図である。

【図５】リンク情報が削除された後の状態の例を示す説明図である。

【図６】削除されたリンク情報をもとに同義性の値を算出する過程を示す説明図である。

【図７】本発明の第２の実施形態における辞書作成支援装置の例を示すブロック図である。

【図８】第２の実施形態における動作の例を示すフローチャートである。

【図９】ＵＲＬとアンカーテキストとのリンク情報を示す説明図である。

【図１０】リンク情報をもとに同義性の値を算出する過程を示す説明図である。

【図１１】本発明の第３の実施形態における辞書作成支援装置の例を示すブロック図である。

【図１２】第３の実施形態における動作の例を示すフローチャートである。

【図１３】ＵＲＬとアンカーテキストとのリンク情報を示す説明図である。

【図１４】同義語候補を統合した際のＵＲＬとアンカーテキストとのリンク情報の例を示す説明図である。

【図１５】本発明による辞書作成支援装置の最小構成の例を示すブロック図である。

【図１６】ＵＲＬとアンカーテキストとのリンク情報を示す説明図である。

【図１７】リンク情報をもとに同義性の値を算出する過程を示す説明図である。

【発明を実施するための形態】
【００３４】
以下、本発明の実施形態を図面を参照して説明する。
【００３５】
実施形態１．
図１は、本発明の第１の実施形態における辞書作成支援装置の例を示すブロック図である。本実施形態における辞書作成支援システムは、データ処理部１と、記憶部２とを備えている。なお、データ処理部１と、記憶部２とは、それぞれが独立の装置であってもよい。
【００３６】
記憶部２は、リンク情報記憶部２０と、関連リンク情報記憶部２１とを備えている。リンク情報記憶部２０は、リンク情報を記憶する。リンク情報には、アンカーテキストと、そのアンカーテキストでリンクするＵＲＬと、ＵＲＬに対するアンカーテキストを持つリンクの数とを対応付けた情報が含まれる。
【００３７】
関連リンク情報記憶部２１は、リンク情報抽出手段１１が抽出したリンク情報を格納する。すなわち、関連リンク情報記憶部２１が記憶するリンク情報は、リンク情報記憶部２０が記憶するリンク情報のサブセットである。
【００３８】
リンク情報記憶部２０及び関連リンク情報記憶部２１は、記憶部２が備える磁気ディスク等によって実現される。
【００３９】
データ処理部１は、入力語保持手段１０と、リンク情報抽出手段１１と、実体確率計算手段１２と、同義性計算手段１３と、表示指示手段１４と、判定結果保持手段１５と、リンク情報削除手段１６とを備えている。
【００４０】
入力語保持手段１０は、利用者からキーボードなどの入力装置（図示せず）を介し、同義語を生成する対象の語として入力された入力語を記憶する。
【００４１】
リンク情報抽出手段１１は、リンク情報記憶部２０に記憶されたリンク情報の中から、入力語に関するリンク情報と、入力語がアンカーテキストとして出現するＵＲＬに一回以上出現するアンカーテキストに関するリンク情報を抽出し、関連リンク情報記憶部２１に記憶させる。すなわち、リンク情報抽出手段１１は、入力語が含まれるアンカーテキストがリンクするＵＲＬが存在する場合、そのＵＲＬにリンクする他のアンカーテキストのリンク情報も抽出し、関連リンク情報記憶部２１に記憶させる。なお、ＵＲＬは、Ｗｅｂ上の資源であるＷｅｂページを識別する識別子であることから、資源位置指定子と呼ぶことができる。
【００４２】
実体確率計算手段１２は、あるＵＲＬｕへリンクするアンカーテキストｘがＵＲＬｕの実体を表す確率である実体確率Ｐ（ｘ｜ｕ）を、関連リンク情報記憶部２１に記憶されたリンク情報ごとに計算する。すなわち、実体確率計算手段１２は、ＵＲＬとアンカーテキストとのリンクごとに実体確率を計算する。
【００４３】
同義性計算手段１３は、入力語と関連リンク情報記憶部２１に記憶されたアンカーテキスト間の同義性ＮＲｅｌ（ｓ，ｔ）を実体確率Ｐ（ｘ｜ｕ）に基づいて計算する。
【００４４】
表示指示手段１４は、同義性計算手段１３が計算した同義性の上位を同義語候補としてディスプレイ装置などの出力装置（図示せず）に送信し、出力させる。
【００４５】
以上のように、リンク情報抽出手段１１、実体確率計算手段１２、同義性計算手段１３

及び表示指示手段１４は、アンカーテキストを用いて入力語の同義語候補を生成する。なお、アンカーテキストを用いて入力語の同義語候補を生成する方法は、非特許文献１に記載された方法であってもよい。ただし、アンカーテキストを用いて入力語の同義語候補を生成する方法は、非特許文献１に記載された方法に限定されない。アンカーテキストを用いて入力語の同義語候補を生成する具体的な方法については後述する。

【００４６】

判定結果保持手段１５は、同義語候補が入力語の同義語か非同義語かについて判定された結果を記憶する。判定結果保持手段１５は、同義語候補を入力語の同義語もしくは非同義語と判定することを規定したルールに基づいて同義語判定手段（図示せず）が判定した結果を記憶してもよい。もしくは、判定結果保持手段１５は、表示指示手段１４が出力装置（図示せず）に出力させた同義語候補に対して利用者が同義語か非同義語かを判定した結果を記憶してもよい。

【００４７】

リンク情報削除手段１６は、関連リンク情報記憶部２１に格納されたリンク情報の中で、あるＷｅｂページのＵＲＬに対して入力語および同義語であると判断された同義語候補がそれぞれアンカーテキストであるリンク情報と、そのＵＲＬに対して非同義語であると判断された同義語候補がアンカーテキストであるリンク情報のうち、いずれかのリンク情報を、上記ＵＲＬを指すそれぞれのリンクの数に基づいて削除する。

【００４８】

すなわち、リンク情報削除手段１６は、あるＵＲＬに対し、アンカーテキストが入力語もしくは同義語候補の中でその入力語の同義語と判定された同義語候補になっているリンクの情報（以下、同義語等に関するリンク情報と記す。）を抽出する。また、リンク情報削除手段１６は、上記ＵＲＬに対し、アンカーテキストが同義語候補の中で入力語の非同義語と判定された同義語候補になっているリンクの情報（以下、非同義語に関するリンク情報と記す。）を抽出する。そして、リンク情報削除手段１６は、同義語等に関するリンク情報と非同義語に関するリンク情報の少なくとも一方のリンク情報を削除する。このとき、リンク情報削除手段１６は、上記ＵＲＬに対する各アンカーテキストを用いたリンクの数に基づいてリンク情報を削除する。

【００４９】

言い換えると、リンク情報削除手段１６は、あるＵＲＬに対してあるアンカーテキストを持つリンクの数（あるＵＲＬに対してあるアンカーテキストでリンクする回数）に基づいてリンクを削除すると言うことができる。そして、リンク情報抽出手段１１、実体確率計算手段１２、同義性計算手段１３及び表示指示手段１４は、各ＵＲＬに対して残された（すなわち、削除対象から除かれた）リンクのアンカーテキストを用いて入力語の同義語候補を生成する。

【００５０】

入力語保持手段１０及び判定結果保持手段１５は、データ処理部１が備える磁気ディスク等によって実現される。

【００５１】

また、リンク情報抽出手段１１と、実体確率計算手段１２と、同義性計算手段１３と、表示指示手段１４と、リンク情報削除手段１６とは、プログラム（辞書作成支援プログラム）に従って動作するコンピュータのＣＰＵによって実現される。例えば、プログラムは、データ処理部１の記憶部（図示せず）に記憶され、ＣＰＵは、そのプログラムを読み込み、プログラムに従って、リンク情報抽出手段１１、実体確率計算手段１２、同義性計算手段１３、表示指示手段１４及びリンク情報削除手段１６として動作してもよい。また、リンク情報抽出手段１１と、実体確率計算手段１２と、同義性計算手段１３と、表示指示手段１４と、リンク情報削除手段１６とは、それぞれが専用のハードウェアで実現されていてもよい。

【００５２】

次に、動作について説明する。図２は、第１の実施形態における動作の例を示すフロー

チャートである。

【００５３】

まず、入力語保持手段１０は、利用者から入力装置（図示せず）を介して入力された入力語を記憶する（ステップＳ１）。本実施形態では、入力語保持手段１０は、入力語として「ＮＥＣ」を保持しているものとする。

【００５４】

次に、リンク情報抽出手段１１は、リンク情報記憶部２０に記憶されたリンク情報の中で、入力語に関するリンク情報と、入力語がアンカーテキストとして出現するＵＲＬに一回以上出現するアンカーテキストに関するリンク情報を抽出し、関連リンク情報記憶部２１に記憶させる（ステップＳ２）。

【００５５】

図３は、リンク情報の例を示す説明図である。リンク情報とは、どのアンカーテキストがどのＵＲＬに何回リンクしたかを示す情報であり、一つのリンク情報には、図３に例示するように、「アンカーテキスト」項目、「ＵＲＬ」項目、「リンク回数」項目が含まれる。なお、「アンカーテキストｘに関するリンク情報」とは、「アンカーテキスト」項目が、アンカーテキストｘと一致するリンク情報を表す。

【００５６】

本実施形態では、リンク情報抽出手段１１がリンク情報記憶部２０に記憶されたリンク情報の中から、入力語「ＮＥＣ」に関するリンク情報及び入力語がアンカーテキストとして出現するＵＲＬに一回以上出現するアンカーテキストに関するリンク情報として、図３に例示する情報を抽出し、関連リンク情報記憶部２１に記憶させたものとする。また、図３に例示する表の内容をグラフ化したものが、図１６に例示するグラフである。以下の説明では、図１６に例示する内容をリンク情報として用いるものとする。

【００５７】

次に、実体確率計算手段１２は、あるＵＲＬｕにリンクするアンカーテキストｘがＵＲＬｕの実体を表す確率である実体確率Ｐ（ｘ｜ｕ）を、関連リンク情報記憶部２１に格納されたリンク情報ごとに計算する（ステップＳ３）。各アンカーテキストからＵＲＬｕに対するリンクの総数をＬ＿ｕ、ＵＲＬｕに対するアンカーテキストｘを持つリンクの数（すなわち、アンカーテキストｘでＵＲＬｕにリンクする数)をＬ＿ｕ，ｘとしたとき、実体確率Ｐ（ｘ｜ｕ）は、以下の式６によって算出される。

【００５８】

Ｐ（ｘ｜ｕ）＝Ｌ＿ｕ，ｘ／Ｌ＿ｕ　（式６）

【００５９】

ここで、Ｐ（ｘ｜ｕ）は、「実体を表すアンカーテキストほど多くの人が用いる」ことを前提とした確率である。図１６に示す例では、「Ｐ→」の下部に示す数値が実体確率の計算結果である。このように、実体確率計算手段１２は、各アンカーテキストからＵＲＬｕへのリンクの総数（Ｌ＿ｕ）に対するアンカーテキストｘからＵＲＬｕへのリンクの数（Ｌ＿ｕ，ｘ）の割合を実体確率として算出する。

【００６０】

次に、同義性計算手段１３は、入力語と関連リンク情報記憶部２１に格納されたアンカーテキスト間の同義性ＮＲｅｌ（ｓ，ｔ）を実体確率に基づいて計算する（ステップＳ４）。アンカーテキストｓとアンカーテキストｔとの間の同義性ＮＲｅｌ（ｓ，ｔ）は下記の式７により算出される。なお、Σにより総和を求める範囲（すなわち、ｉの取りうる範囲）は、アンカーテキストｓあるいはアンカーテキストｔがアンカーテキストとして出現するＵＲＬの数である。

【００６１】

【数２】

$$NRel(s,t)=\frac{\Sigma i[P(s|u\_i)*P(t|u\_i)]}{\Sigma i[P(s|u\_i)+P(t|u\_i)-P(s|u\_i)*P(t|u\_i)]} \qquad (式７)$$

【００６２】
図１６に例示するグラフ（すなわち、図３に例示する関連リンク情報記憶部２１のリンク情報）の内容に対し、同義性計算手段１３が式７に基づいて同義性ＮＲｅｌ（ｓ，ｔ）を計算した結果が図１７に例示する内容である。図１７に示す例では、同義性計算手段１３が式７を用いて、「ＮＥＣ」と「日本電気」、「ＮＥＣ」と「ＢＩＧＬＯＢＥ」及び「ＮＥＣ」と「ビッグローブ」の同義性を算出していることを示す。

【００６３】
次に、表示指示手段１４は、同義性計算手段１３が計算した同義性を示す値の上位を同義語候補として出力装置（図示せず）に送信し、表示させる（ステップＳ５）。例えば、アンカーテキストｓとアンカーテキストｔとが同義であると判断する時における同義性ＮＲｅｌ（ｓ，ｔ）の値の閾値を０．０１に設定したとする。この場合、表示指示手段１４は、図１７に例示する同義性の計算結果に対して、「日本電気」、「ＢＩＧＬＯＢＥ」及び「ビッグローブ」を同義語候補として出力装置（図示せず）に送信する。

【００６４】
上記説明では、表示指示手段１４は、同義性ＮＲｅｌ（ｓ，ｔ）が示す値と比較する閾値を設定して、同義語候補を出力する場合について説明した。他にも、表示指示手段１４は、同義性を示す値の上位ｎ件もしくは同義性を示す値の上位ｎ％などの指標を用いて同義語候補を出力してもよい。

【００６５】
次に、判定結果保持手段１５は、同義語候補が同義語か非同義語かを利用者が判断した判断結果を入力装置（図示せず）を介して受け取り、その判断結果を記憶する（ステップＳ６）。

【００６６】
本実施形態では、利用者に提示した同義語候補のうち、入力語「ＮＥＣ」に対する同義語として「日本電気」が、非同義語として「ＢＩＧＬＯＢＥ」がそれぞれ、判断結果として入力装置（図示せず）に入力されたものとする。

【００６７】
次に、リンク情報削除手段１６は、関連リンク情報記憶部２１に記憶されたリンク情報の中で、以下のリンク情報のうちのいずれかを、ＵＲＬを指すそれぞれのリンクの数に基づいて削除する。一つは、あるＷｅｂページのＵＲＬに対して、アンカーテキストが入力語もしくはその入力語の同義語であると判断された同義語候補になっているリンク情報（すなわち、同義語等に関するリンク情報）であり、もう一つは、そのＵＲＬに対して、入力語の非同義語であると判断された同義語候補がアンカーテキストであるリンク情報（すなわち、非同義語に関するリンク情報）である。

【００６８】
例えば、図１６に例示するリンク情報には、ＵＲＬ「ｂｉｇｌｏｂｅ．ｎｅ．ｊｐ」に対するアンカーテキストとして、入力語である「ＮＥＣ」と、同義語候補である「日本電気」、「ＢＩＧＬＯＢＥ」及び「ビッグローブ」が存在する。このうち、「ＮＥＣ」は入力語であり、「日本電気」は同義語、「ＢＩＧＬＯＢＥ」は非同義語と判断されている。

【００６９】
このとき、ＵＲＬ「ｂｉｇｌｏｂｅ．ｎｅ．ｊｐ」に対する「ＮＥＣ」および「日本電気」のアンカーテキストと、「ＢＩＧＬＯＢＥ」のアンカーテキストの少なくともいずれか一方は、ＵＲＬ「ｂｉｇｌｏｂｅ．ｎｅ．ｊｐ」の実体を表していないと言える。「ＮＥＣ」と「ＢＩＧＬＯＢＥ」とは非同義語の関係にあり、両者は別の対象物を示すはずだからである。

【００７０】
リンク情報削除手段１６は、削除の判断基準に実体確率を用いる。具体的には、リンク情報削除手段１６は、実体確率が小さいアンカーテキストに関するリンク情報を削除する。上記の例では、ＵＲＬ「ｂｉｇｌｏｂｅ．ｎｅ．ｊｐ」に着目した場合、実体確率は、それぞれ、Ｐ（ＢＩＧＬＯＢＥ｜ｂｉｇｌｏｂｅ．ｎｅ．ｊｐ）＝０．３３、Ｐ（日本電気｜ｂｉｇｌｏｂｅ．ｎｅ．ｊｐ）＝０．１７、Ｐ（ＮＥＣ｜ｂｉｇｌｏｂｅ．ｎｅ．ｊｐ）＝０．１７と算出される。よって、リンク情報削除手段１６は、実体確率が小さいＵＲＬ「ｂｉｇｌｏｂｅ．ｎｅ．ｊｐ」に対する「日本電気」「ＮＥＣ」に関するリンク情報を削除する。

【００７１】
同様に、リンク情報削除手段１６は、ＵＲＬ「ｎｅｃ．ｊｐ」に対する「ＮＥＣ」に関するリンク情報と、「ＢＩＧＬＯＢＥ」に関するリンク情報のうち、実体確率の小さい「ＢＩＧＬＯＢＥ」に関するリンク情報を削除する。

【００７２】
なお、リンク情報削除手段１６がリンク情報を削除するのは、同じＵＲＬに対して、上記の２種類のリンク情報（すなわち、同義語等に関するリンク情報と非同義語に関するリンク情報）の双方が存在する場合のみである。例えば、「ｎｅｃ．ｃｏ．ｊｐ」に対しては入力語「ＮＥＣ」と同義語「日本電気」のリンクは存在するが、非同義語「ＢＩＧＬＯＢＥ」のリンクは存在しない。そのため、リンク情報削除手段１６は、「ｎｅｃ．ｃｏ．ｊｐ」に関するリンク情報を削除の対象としない。

【００７３】
また、リンク情報削除手段１６が、アンカーテキストが入力語であるリンク情報を削除する場合、そのリンク情報のＵＲＬは指定された入力語に関する同義語候補抽出では利用されなくなる。これは、入力語に関する同義語候補を抽出するタスクにおいて実質的にＵＲＬを削除したことと同じ意味を持つ。

【００７４】
図４は、リンク情報削除手段１６がリンク情報を削除する処理の例を示す説明図である。また、図５は、リンク情報が削除された後の状態の例を示す説明図である。図４に例示する点線部が、リンク情報削除手段１６により削除される対象になるリンク情報である。その他の内容は、図１６に記載した内容と同様である。以上の処理により、図４に例示する点線部のリンク情報が削除され、その結果、関連リンク情報記憶部２１には、図５に例示するリンク情報が残ることになる。

【００７５】
この処理の後、ステップＳ３に戻り、以降の処理を繰り返す。リンク情報削除手段１６がリンク情報を削除した結果をもとに実体確率計算手段１２が算出した実体確率が、図５に例示する「Ｐ→」の下部に示す数値である。その他の内容は、図１６に記載した内容と同様である。図５に示す例では、Ｐ（ＮＥＣ｜ｂｉｇｌｏｂｅ．ｎｅ．ｊｐ）＝０になり、図４に例示する１回目に算出された実体確率の値である０．１７よりも減少していることが分かる。同様に、図５に示す例では、Ｐ（ＮＥＣ｜ｎｅｃ．ｃｏ．ｊｐ）＝０．５になり、図４に例示する１回目に算出された実体確率の値である０．４よりも増加していることが分かる。この例からも、リンク情報の削除によって、実体確率をより正確に推定できるようになったと言える。

【００７６】
なお、本実施形態において算出した実体確率は、一つのＵＲＬに対して実体を表す語が一つであることを前提としている。したがって、同義語が存在する場合は、実体を表すアンカーテキストであってもその実体確率は１にならない。例えば、Ｐ（ＮＥＣ｜ｎｅｃ．ｃｏ．ｊｐ）の理論上の最大値は０．５になる。したがって、リンク削除手段により実体確率を理論上の最大値にまで上昇できたことになる。

【００７７】
図６は、削除されたリンク情報をもとに同義性の値を算出する過程を示す説明図である

。図６に示す例では、同義性計算手段１３が式７を用いて同義性の値を算出する２回目の過程を示している。図１７に例示する１回目の計算結果と比較すると、「ＮＥＣ」と「日本電気」の同義性が、１回目の０．２６から０．３３に増加していることがわかる。また、「ＮＥＣ」と「ＢＩＧＬＯＢＥ」及び「ＮＥＣ」と「ビッグローブ」の同義性が、１回目の０．０７及び０．０３から、それぞれ０に減少していることがわかる。

【００７８】

以上のことから、実体確率計算手段１２が実体確率を正確に推定できたことにより、同義性計算手段１３が同義性の値をより正確に算出できたと言える。

【００７９】

また、同義語候補を表示する際の閾値を１回目と同様に０．０１に設定したとする。この場合、表示指示手段１４は、２回目の判断結果として、同義語候補「日本電気」のみを表示することになる。このことからも、提示する同義語候補の精度がリンク情報の削除によって向上することがわかる。

【００８０】

以上、第１の実施形態の動作を説明した。なお、上記説明では、判定結果保持手段１５が、入力装置を通じて利用者から入力された同義語候補に対する判断結果を記憶する場合について説明した。それ以外の方法として、判定結果保持手段１５が、あらかじめ部分的な同義語辞書と非同義語辞書を記憶しておき、リンク情報削除手段１６は、同義語候補をそれらの辞書にあてはめて、同義語及び非同義語を判断してもよい。

【００８１】

また、判定結果保持手段１５が、利用者から入力された判定結果を記憶する代わりに、同義性計算手段１３が、計算した同義性の値の上位を同義語、同義性の値の下位を非同義語とする判断結果を判定結果保持手段１５に記憶させてもよい。なお、同義性計算手段１３は、この場合の上位及び下位の判断を、予め定められた閾値や範囲に基づいて行えばよい。この場合、同義性計算手段１３の初期の計算結果がある程度正しければ、利用者に負担をかけずに、実体確率の推定精度を向上させることが可能になる。

【００８２】

また、リンク情報削除手段１６は、２種類のリンク情報の中で実体確率が小さい方のアンカーテキストに関するリンク情報を削除する代わりに、単純にＬ＿ｕ，ｘ（ｕに対するｘを持つリンクの数）が小さい方のアンカーテキストに関するリンク情報を削除してもよい。

【００８３】

式６に例示する実体確率の算出式「実体確率Ｐ（ｘ｜ｕ）＝Ｌ＿ｕ，ｘ／Ｌ＿ｕ」におけるＬ＿ｕの値は、比較する両者で同一となる。すなわち、実体確率は、リンクの数に基づいて算出される値であり、実質的にはＬ＿ｕ，ｘで両者を比較しているのと同一だからである。

【００８４】

また、上記説明では、リンク情報削除手段１６が、２種類のリンク情報（すなわち、同義語等に関するリンク情報と非同義語に関するリンク情報）のうち、実体確率が小さい方のアンカーテキストに関するリンク情報を削除する場合について説明した。その代わりに、リンク情報削除手段１６は、実体確率が閾値以下のアンカーテキストを削除するようにしても良い。この場合、実体確率が閾値以下であれば、両方のアンカーテキストに関するリンク情報が削除される場合も、両方が削除されない場合もありうることになる。

【００８５】

また、同義性計算手段１３が同義性を計算する方法は、式７に例示した計算式を利用する場合に限定されない。同義性計算手段１３は、例えば、正規化部分を用いない式２に例示するＲｅｌ（ｓ，ｔ）の計算式を利用して同義性を計算してもよい。また、同義性計算手段１３は、以下の式８に示す計算式を用いて同義性を計算してもよい。

【００８６】

Ｒｅｌ２（ｓ，ｔ）

＝Ａｖｅ（Σｉ（Ｌ＿ｕ＿ｉ，ｓ）／Ｌ＿ｓ，Σ＿ｉ（Ｌ＿ｕ＿ｉ，ｔ）／Ｌ＿ｔ）
（式８）

【００８７】
ここで、Ｌ＿ｕ＿ｉ，ｘはｕ＿ｉに対するアンカーテキストｘを持つリンクの数（ｕ＿ｉに対してｘでリンクする数）を表し、Ｌ＿ｘはアンカーテキストｘを持つリンクの総数を表す。Σにより総和を求める範囲（すなわち、ｉの取りうる範囲）は、アンカーテキストｓとアンカーテキストｔの両方がアンカーテキストとして出現するＵＲＬの数である。また、Ａｖｅは２つの値の平均を表す。この平均は、相加平均、相乗平均及び調和平均のいずれであってもよい。

【００８８】
また、本実施形態では、リンク情報削除手段１６が、リンク情報を削除すると判断した場合に関連リンク情報記憶部２１のリンク情報を削除する場合について説明した。加えて、リンク情報削除手段１６は、対応するリンク情報記憶部２０のリンク情報を削除してもよい。このようにすることで、リンク情報を削除したことを、他の入力語に対しても反映することが可能になる。

【００８９】
また、上記説明では、リンク情報抽出手段１１が、リンク情報記憶部２０に記憶されたリンク情報の中から、入力語に関するリンク情報と、入力語がアンカーテキストとして出現するＵＲＬに一回以上出現するアンカーテキストに関するリンク情報を抽出し、関連リンク情報記憶部２１に記憶させる場合について説明した。ただし、リンク情報抽出手段１１は、上記リンク情報の中で、入力語または同義語と判断されたアンカーテキストが最も多く出現するＵＲＬ（あるいは最も多いものからＮ個のＵＲＬ）に関するリンク情報のみを抽出し、関連リンク情報記憶部２１に記憶させてもよい。この方法では、各アンカーテキストに関する公式ページに近いＵＲＬのみを採用するため、同義語候補の精度が向上する可能性がある。

【００９０】
以上のように、本実施形態によれば、リンク情報抽出手段１１、実体確率計算手段１２、同義性計算手段１３及び表示指示手段１４が、アンカーテキストを用いて入力語の同義語候補を生成する。リンク情報削除手段１６は、あるＵＲＬに対して、アンカーテキストが入力語もしくは同義語候補の中でその入力語の同義語と判定された同義語候補になっているリンクの情報（すなわち、同義語等に関するリンク情報）を抽出する。また、リンク情報削除手段１６は、上記ＵＲＬに対して、アンカーテキストが同義語候補の中で入力語の非同義語と判定された同義語候補になっているリンクの情報（すなわち、非同義語に関するリンク情報）を抽出する。そして、リンク情報削除手段１６は、同義語等に関するリンク情報と非同義語に関するリンク情報のうちの少なくとも一方のリンクを削除する。このとき、リンク情報削除手段１６は、上記ＵＲＬに対するアンカーテキストごとのリンクの数に基づいて、そのＵＲＬとアンカーテキストとのリンクの情報を削除する。そして、リンク情報抽出手段１１、実体確率計算手段１２、同義性計算手段１３及び表示指示手段１４が、削除対象から除かれたリンクのアンカーテキストを用いて入力語の同義語候補を生成する。このような構成により、生成される同義語候補の精度を向上させて辞書作成を支援できる。

【００９１】
具体的には、本実施形態では、推定された実体確率をもとに同義性を判断することで、同義性候補の精度を向上させている。したがって、同義性を判断するためには、実体確率を正しく推定することが必要になる。本実施形態では、リンク情報削除手段１６が、例えば、ユーザの判断結果に基づいてＵＲＬの実体を表さないアンカーテキストに関するリンク情報を削除することで、実体確率を０に減少させている。また、ＵＲＬの実体を表さないアンカーテキストに関するリンク情報を削除することで、ＵＲＬの実体を表す正しいアンカーテキストに関する実体確率を１に近づけることが可能になる。このように、不要なリンク情報を削除して実体確率の推定精度を向上させることで、同義語候補の精度を向上

させることができる。
【００９２】
実施形態２．
図７は、本発明の第２の実施形態における辞書作成支援装置の例を示すブロック図である。なお、第１の実施形態と同様の構成については、図１と同一の符号を付し、説明を省略する。本実施形態における辞書作成支援システムも、データ処理部１と、記憶部２とを備えている。データ処理部１と、記憶部２とは、それぞれが独立の装置であってもよい。
【００９３】
記憶部２は、リンク情報記憶部２０と、関連リンク情報記憶部２１とを備えている。リンク情報記憶部２０及び関連リンク情報記憶部２１の構成は、第１の実施形態と同様である。
【００９４】
データ処理部１は、入力語保持手段１０と、リンク情報抽出手段１１と、改良実体確率計算手段１７と、同義性計算手段１３と、表示指示手段１４と、判定結果保持手段１５と、リンク情報削除手段１６とを備えている。すなわち、実体確率計算手段１２の代わりに改良実体確率計算手段１７を備えている点で、第１の実施形態と異なる。それ以外の構成については、第１の実施形態と同様である。
【００９５】
改良実体確率計算手段１７は、あるＵＲＬｕにリンクするあるアンカーテキストｘがＵＲＬｕの実体を表す確率である実体確率を関連リンク情報記憶部２１に格納されたリンク情報ごとに計算する。このとき、改良実体確率計算手段１７は、あるＵＲＬｕへのリンクの総数に対するそのＵＲＬｕへのアンカーテキストｘを持つリンクの数の割合に加え、そのアンカーテキストｘを持つリンクの総数に対するそのＵＲＬｕへのそのアンカーテキストｘを持つリンクの数の割合を用いて、実体確率を計算する。
【００９６】
すなわち、改良実体確率計算手段１７は、まず、あるＵＲＬｕへのリンクの総数に対する、あるアンカーテキストｘから上記ＵＲＬｕへのリンクの数の割合を計算する。さらに、改良実体確率計算手段１７は、上記アンカーテキストｘを持つ各ＵＲＬへのリンクの総数に対する、上記アンカーテキストから上記ＵＲＬｕへのリンクの数の割合を計算する。改良実体確率計算手段１７は、このように算出した２つの割合を用いて実体確率を計算する。例えば、改良実体確率計算手段１７は、この２つの割合を乗じた値を実体確率として計算してもよい。
【００９７】
なお、リンク情報抽出手段１１と、改良実体確率計算手段１７と、同義性計算手段１３と、表示指示手段１４と、リンク情報削除手段１６とは、プログラム（辞書作成支援プログラム）に従って動作するコンピュータのＣＰＵによって実現される。また、リンク情報抽出手段１１と、改良実体確率計算手段１７と、同義性計算手段１３と、表示指示手段１４と、リンク情報削除手段１６とは、それぞれが専用のハードウェアで実現されていてもよい。
【００９８】
次に、動作について説明する。図８は、第２の実施形態における動作の例を示すフローチャートである。リンク情報抽出手段１１がリンク情報を抽出して、関連リンク情報記憶部２１に記憶させるステップＳ１～ステップＳ２までの処理、及び、同義性計算手段１３が実体確率に基づいて同義性を示す値を算出してから、リンク情報削除手段１６がリンク情報を削除するステップＳ４～ステップＳ７までの処理は、図２に例示する第１の実施形態おける処理と同様である。
【００９９】
改良実体確率計算手段１７は、あるＵＲＬｕにリンクするあるアンカーテキストｘがＵＲＬｕの実体を表す確率である実体確率Ｐ（ｘ｜ｕ）を、関連リンク情報記憶部２１に記憶されたリンク情報ごとに求める（ステップＳ３ａ）。具体的には、改良実体確率計算手

段１７は、上記ＵＲＬｕに対するリンクの総数におけるそのＵＲＬｕに対してアンカーテキストｘを持つリンクの数の割合に加え、そのアンカーテキストｘを持つリンクの総数における上記ＵＲＬに対して上記アンカーテキストｘを持つリンクの数の割合を用いて実体確率を計算する。

【０１００】

ここで、各アンカーテキストからＵＲＬｕに対するリンクの総数をＬ＿ｕ、アンカーテキストｘを持つ各ＵＲＬへのリンクの総数をＬ＿ｘ、ＵＲＬｕに対するアンカーテキストｘを持つリンクの数（アンカーテキストｘでＵＲＬｕにリンクする数）をＬ＿ｕ，ｘとしたとき、改良実体確率計算手段１７は、以下の式９を用いて実体確率Ｐ（ｘ｜ｕ）を算出する。

【０１０１】

Ｐ（ｘ｜ｕ）＝（Ｌ＿ｕ，ｘ／Ｌ＿ｕ）＊（Ｌ＿ｕ，ｘ／Ｌ＿ｘ）　　（式９）

【０１０２】

図９は、ＵＲＬとアンカーテキストとのリンク情報を示す説明図である。図９に例示する「Ｐ→」の下部に示された値は、図３に例示する関連リンク情報記憶部２１のリンク情報をもとに式９を用いて計算されたリンクごとの実体確率である。それ以外については、図１７に示す内容と同様である。

【０１０３】

例えば、図９に示す例では、Ｐ（ＮＥＣ｜ｂｉｇｌｏｂｅ．ｎｅ．ｊｐ）＝０．０３、Ｐ（ＮＥＣ｜ｎｅｃ．ｊｐ）＝０．２であり、両者の比は６．７である。一方、第１の実施形態における方法で実体確率を算出した場合、上記値は、図１６に例示するように、Ｐ（ＮＥＣ｜ｂｉｇｌｏｂｅ．ｎｅ．ｊｐ）＝０．１７、Ｐ（ＮＥＣ｜ｎｅｃ．ｊｐ）＝０．４であり両者の比は２．３５である。

【０１０４】

第１の実施形態における方法で算出した実体確率に比べ、第２の実施形態における方法で算出した実体確率は、Ｐ（ＮＥＣ｜ｂｉｇｌｏｂｅ．ｎｅ．ｊｐ）の値とＰ（ＮＥＣ｜ｎｅｃ．ｊｐ）の値の両方とも減少している。しかし、Ｐ（ＮＥＣ｜ｂｉｇｌｏｂｅ．ｎｅ．ｊｐ）とＰ（ＮＥＣ｜ｎｅｃ．ｊｐ）との比の値を比較した場合、第２の実施形態における方法で算出した実体確率の比の値がより大きくなっている。このことから、上記方法によれば、ＵＲＬの実体を表わさないアンカーテキストの実体確率をより減少できることが分かる。

【０１０５】

図１０は、図９に例示する実体確率を用いて同義性を示す値を計算する過程を示す説明図である。図１０に示す例では、改良実体確率計算手段１７が算出した実体確率を用いて、同義性計算手段１３が、「ＮＥＣ」と「日本電気」、「ＮＥＣ」と「ＢＩＧＬＯＢＥ」及び「ＮＥＣ」と「ビッグローブ」の同義性を式７を用いて算出していることを示す。

【０１０６】

ここで、「ＮＥＣ」と「日本電気」は同義語であり、「ＮＥＣ」と「ＢＩＧＬＯＢＥ」及び「ＮＥＣ」と「ビッグローブ」は非同義語であるとする。このとき、同義性計算手段１３は、同義語の同義性ＮＲｅｌ（ＮＥＣ，日本電気）＝０．１０、非同義語の同義性ＮＲｅｌ（ＮＥＣ，ＢＩＧＬＯＢＥ）＝０．０１、ＮＲｅｌ（ＮＥＣ，ビッグローブ）＝０．０１と算出する。この場合、同義語と非同義語との間の同義性の値の比の平均は１０である。

【０１０７】

一方、図１７に例示する算出結果によれば、第１の実施形態において算出される同義語と非同義語との間の同義性の値の比の平均は６．２である。以上の具体例からも、第２の実施形態における方法で算出される実体確率を用いることで、非同義語の同義性を相対的に減少させられることがわかる。

【０１０８】

以上、第２の実施形態の動作を説明した。なお、上記説明では、改良実体確率計算手段

１７が、式９に例示する算出式を用いて実体確率を算出する方法について説明した。他にも、改良実体確率計算手段１７は、式９に例示する算出式の一部を用いた下記の式１０を用いて実体確率を計算してもよい。

【０１０９】

Ｐ（ｘ｜ｕ）＝Ｌ＿ｕ，ｘ／Ｌ＿ｘ　（式１０）

【０１１０】

以上のように、本実施形態によれば、改良実体確率計算手段１７が、各アンカーテキストからＵＲＬｕへのリンクの総数に対するアンカーテキストｘからそのＵＲＬｕへのリンクの数の割合を算出する。さらに、改良実体確率計算手段１７が、アンカーテキストｘを持つ各ＵＲＬへのリンクの総数に対するそのアンカーテキストｘから上記ＵＲＬｕへのリンクの数の割合を用いて実体確率を算出する。このような構成により、生成される同義語候補の精度を向上させて辞書作成を支援できる。

【０１１１】

具体的には、第１の実施形態と同様、推定された実体確率Ｐ（ｘ｜ｕ）をもとに同義性を判断することで、同義性候補の精度を向上させている。したがって、同義性を判断するためには、実体確率を正しく推定することが必要になる。本実施形態では、改良実体確率計算手段１７が、アンカーテキストｘを持つリンクがＵＲＬｕに対するリンクの中でどの程度多数派をしめるかを表す指標（例えば、第１の実施形態における式６で算出される実体確率）を算出する。さらに、改良実体確率計算手段１７が、アンカーテキストｘがＵＲＬｕのみにどの程度多くリンクするかを表す指標を算出して実体確率を計算する。したがって、アンカーテキストのリンク状況をより反映させた実体確率の値を推定できるため、同義語候補の精度を向上させることができる。

【０１１２】

実施形態３．

図１１は、本発明の第３の実施形態における辞書作成支援装置の例を示すブロック図である。なお、第１の実施形態と同様の構成については、図１と同一の符号を付し、説明を省略する。本実施形態における辞書作成支援システムも、データ処理部１と、記憶部２とを備えている。データ処理部１と、記憶部２とは、それぞれが独立の装置であってもよい。

【０１１３】

記憶部２は、リンク情報記憶部２０と、関連リンク情報記憶部２１とを備えている。リンク情報記憶部２０及び関連リンク情報記憶部２１の構成は、第１の実施形態と同様である。

【０１１４】

データ処理部１は、入力語保持手段１０と、リンク情報抽出手段１１と、実体確率計算手段１２と、同義性計算手段１３と、表示指示手段１４と、判定結果保持手段１５と、リンク情報削除手段１６と、リンク情報統合手段１８を備えている。すなわち、リンク情報統合手段１８をさらに備えている点で、第１の実施形態と異なる。それ以外の構成については、第１の実施形態と同様である。

【０１１５】

リンク情報統合手段１８は、あるＷｅｂページのＵＲＬに対して入力語および同義語であると判断された同義語候補がそれぞれアンカーテキストであるリンク情報を統合する。ここで、リンク情報の統合とは、各同義語候補のアンカーテキストを同一のアンカーテキスト（以下、統合アンカーテキストと記す。）とみなし、各同義語候補のアンカーテキストでリンクしていたＵＲＬを、統合アンカーテキストでリンクするＵＲＬとみなしたリンク情報を生成することである。このとき、リンク情報におけるリンクの数も、統合アンカーテキストを持つリンクの数として集約する。

【０１１６】

すなわち、リンク情報統合手段１８は、アンカーテキストが同義語候補になっている各ＵＲＬに対するリンクのうち、入力語の同義語と判定された同義語候補がアンカーテキス

トであるリンク情報を統合する。したがって、リンク情報が統合された後、リンク情報抽出手段１１、実体確率計算手段１２、同義性計算手段１３及び表示指示手段１４は、統合されたリンクのアンカーテキストを用いて入力語の同義語候補を生成する。

【０１１７】

リンク情報抽出手段１１と、実体確率計算手段１２と、同義性計算手段１３と、表示指示手段１４と、リンク情報削除手段１６と、リンク情報統合手段１８とは、プログラム（辞書作成支援プログラム）に従って動作するコンピュータのＣＰＵによって実現される。また、リンク情報抽出手段１１と、実体確率計算手段１２と、同義性計算手段１３と、表示指示手段１４と、リンク情報削除手段１６と、リンク情報統合手段１８とは、それぞれが専用のハードウェアで実現されていてもよい。

【０１１８】

次に、動作について説明する。図１２は、第３の実施形態における動作の例を示すフローチャートである。リンク情報抽出手段１１がリンク情報を抽出してから、リンク情報削除手段１６がリンク情報を削除するステップＳ１～ステップＳ７までの処理は、図２に例示する第１の実施形態おける処理と同様である。

【０１１９】

リンク情報統合手段１８は、あるＷｅｂページのＵＲＬに対して、アンカーテキストが入力語もしくは入力語の同義語であると判断された同義語候補になっているリンク情報を統合する（ステップＳ８）。

【０１２０】

図１３は、ＵＲＬとアンカーテキストとのリンク情報を示す説明図である。以下、関連リンク情報記憶部２１が、図１３に例示するリンク情報を記憶している場合について説明する。図１３に例示する「Ｐ→」の下部に示された値は、図１３に例示するリンク情報を用いて実体確率計算手段１２が計算した実体確率である。それ以外については、図１７に示す内容と同様である。

【０１２１】

図１３に例示するように、アンカーテキスト「ＮＥＣ」とアンカーテキスト「日電」とは、同一のＵＲＬにリンクしていない。よって、アンカーテキスト「日電」がリンクするＵＲＬに対する「ＮＥＣ」の実体確率は０になるため（Ｐ（ＮＥＣ｜ｎｅｃ．ｃｏｍ．ｃｎ）＝０）、「ＮＥＣ」と「日電」の同義性は０になる。したがって、このままでは「日電」を「ＮＥＣ」の同義語候補にすることできない。

【０１２２】

ここで、表示指示手段１４が「ＮＥＣ」と「日本電気」とを同義語候補として出力し、利用者が、両者を同義語と判断したとする。このとき、リンク情報統合手段１８は、「ＮＥＣ」と「日本電気」に関するリンク情報を統合する。同義語候補を統合した際のＵＲＬとアンカーテキストとのリンク情報を図１４に示す。

【０１２３】

この処理の後、ステップＳ３に戻り、以降の処理を繰り返す。リンク情報統合手段１８がリンク情報を統合した結果をもとに実体確率計算手段１２が算出した実体確率が、図１４に例示する「Ｐ→」の下部に示す数値である。図１４に示す例では、「日電」がリンクするＵＲＬに対する「ＮＥＣ」の実体確率は、Ｐ（ＮＥＣ｜ｎｅｃ．ｃｏｍ．ｃｎ）＝０．２３と算出される。このことから、図１３に例示する１回目に算出された実体確率の値０よりも増加していることが分かる。この実体確率を用いて同義性を示す値を計算することで、「ＮＥＣ」と「日電」との同義性も向上する。よって、「日電」を「ＮＥＣ」の同義語候補とすることができる。

【０１２４】

以上、第３の実施形態の動作を説明した。なお、上記説明では、リンク情報削除手段１６がリンク情報を削除した後、リンク情報統合手段１８がリンク情報を統合する場合について説明した。ただし、リンク情報を削除する処理と、リンク情報を統合する処理とが行われる順番は、上記順番に限定されない。リンク情報統合手段１８がリンク情報を統合し

た後で、リンク情報削除手段１６がリンク情報を削除してもよい。

【０１２５】

以上のように、本実施形態によれば、リンク情報統合手段１８が、アンカーテキストが同義語候補になっている各ＵＲＬに対するリンクのうち、入力語の同義語と判定された同義語候補がアンカーテキストであるリンクを統合することで、生成される同義語候補の精度を向上させて辞書作成を支援する。

【０１２６】

すなわち、本実施形態でも、推定された実体確率Ｐ（ｘ｜ｕ）をもとに同義性を判断して、同義性候補の精度を向上させる。したがって、同義性を判断するためには、実体確率を正しく推定することが必要になる。本実施形態では、リンク情報統合手段１８が、同義語と判断されたリンク情報を統合する。リンク情報を統合することにより、実体確率の値を推定するためのリンク情報を増加させることができる。このように、他のアンカーテキストのリンク情報も反映させた実体確率の値を推定できるため、同義語候補の精度をより向上させることができる。

【０１２７】

次に、本発明による辞書作成支援装置の最小構成の例を説明する。図１５は、本発明による辞書作成支援装置の最小構成の例を示すブロック図である。本発明による辞書作成支援装置は、辞書作成を支援する辞書作成支援装置であって、Ｗｅｂページを識別する識別子である各資源位置指定子（例えば、ＵＲＬ）に対してリンクする文字列を表すアンカーテキストを用いて、同義語を生成する対象の語として辞書作成支援装置に入力される入力語の同義語候補を生成する同義語候補生成手段８１（例えば、リンク情報抽出手段１１、実体確率計算手段１２、同義性計算手段１３及び表示指示手段１４）と、一の資源位置指定子に対して、アンカーテキストが入力語もしくは同義語候補の中でその入力語の同義語と判定された同義語候補になっているリンクである第一のリンク（例えば、同義語等に関するリンク情報）と、アンカーテキストが同義語候補の中で入力語の非同義語と判定された同義語候補になっているリンクである第二のリンク（例えば、非同義語に関するリンク情報）のうち、少なくとも一方のリンクを削除するリンク削除手段８２（例えば、リンク情報削除手段１６）とを備えている。

【０１２８】

リンク削除手段８２は、一の資源位置指定子に対するアンカーテキストごとのリンクの数に基づいて、上記一の資源位置指定子とアンカーテキストとのリンクを削除し、同義語候補決定手段８１は、リンク削除手段８２が削除する対象から除かれたリンクのアンカーテキストを用いて入力語の同義語候補を生成する。

【０１２９】

そのような構成により、生成される同義語候補の精度を向上させて辞書作成を支援できる。

【０１３０】

なお、少なくとも以下に示すような辞書作成支援装置も、上記に示すいずれかの実施形態に開示されている。

【０１３１】

（１）辞書作成を支援する辞書作成支援装置であって、Ｗｅｂページを識別する識別子である各資源位置指定子（例えば、ＵＲＬ）に対してリンクする文字列を表すアンカーテキストを用いて、同義語を生成する対象の語として辞書作成支援装置に入力される入力語の同義語候補を生成する同義語候補生成手段（例えば、リンク情報抽出手段１１、実体確率計算手段１２、同義性計算手段１３及び表示指示手段１４）と、、一の資源位置指定子に対して、アンカーテキストが入力語もしくは同義語候補の中でその入力語の同義語と判定された同義語候補になっているリンクである第一のリンク（例えば、同義語等に関するリンク情報）と、アンカーテキストが同義語候補の中で入力語の非同義語と判定された同義語候補になっているリンクである第二のリンク（例えば、非同義語に関するリンク情報）のうち、少なくとも一方のリンクを削除するリンク削除手段（例えば、リンク情報削除手

段１６）とを備え、リンク削除手段が、一の資源位置指定子に対するアンカーテキストごとのリンクの数に基づいて、上記一の資源位置指定子とアンカーテキストとのリンクを削除し、同義語候補決定手段が、リンク削除手段が削除する対象から除かれたリンクのアンカーテキストを用いて入力語の同義語候補を生成する辞書作成支援装置。

【０１３２】

（２）資源位置指定子に対してリンクするアンカーテキストがその資源位置指定子の実体を表す確率である実体確率（例えば、Ｐ（ｘ｜ｕ））を、資源位置指定子とアンカーテキストとのリンクごとに計算する実体確率計算手段（例えば、実体確率計算手段１２）を備え、リンク削除手段が、第一のリンクと第二のリンクのうち、実体確率が小さいリンクを削除する辞書作成支援装置。

【０１３３】

（３）実体確率計算手段（例えば、実体確率計算手段１２）が、各アンカーテキストから一の資源位置指定子へのリンクの総数に対する一のアンカーテキストからその一の資源位置指定子へのリンクの数の割合を実体確率として算出する（例えば、式６を用いて算出する）辞書作成支援装置。

【０１３４】

（４）実体確率計算手段（例えば、改良実体確率計算手段１７）は、各アンカーテキストから一の資源位置指定子へのリンクの総数に対する一のアンカーテキストからその一の資源位置指定子へのリンクの数の割合、及び、アンカーテキストを持つ各資源位置指定子へのリンクの総数に対するそのアンカーテキストから一の資源位置指定子へのリンクの数の割合を用いて実体確率を算出する（例えば、式９を用いて算出する）辞書作成支援装置。

【０１３５】

（５）アンカーテキストが同義語候補になっている各資源位置指定子に対するリンクのうち、入力語の同義語と判定された同義語候補がアンカーテキストであるリンクを統合するリンク統合手段（例えば、リンク情報統合手段１８）を備え、同義語候補決定手段が、統合されたリンクのアンカーテキストを用いて入力語の同義語候補を生成する辞書作成支援装置。

【０１３６】

（６）リンク削除手段が、第一のリンクと第二のリンクの双方が一の資源位置指定子に存在しない場合に、その一の資源位置指定子に対するアンカーテキストのリンクを削除対象から除く辞書作成支援装置。

【産業上の利用可能性】

【０１３７】

本発明は、Ｗｅｂのアンカーテキストを用いて入力語に対する同義語候補を生成する辞書作成支援装置に好適に適用される。また、本発明による辞書作成支援装置で作成された同義語辞書は、文書検索、顧客データの名寄せなど、様々なソフトウェアの基本的な資源として使用可能である。

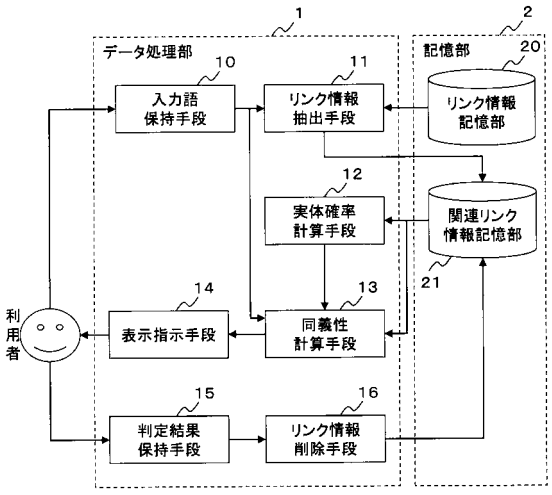
【符号の説明】

【０１３８】

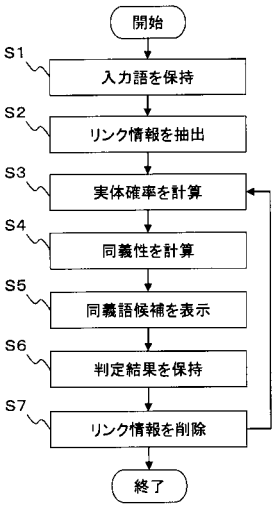
１　データ処理部
２　記憶部
１０　入力語保持手段
１１　リンク情報抽出手段
１２　実体確率計算手段
１３　同義性計算手段
１４　表示指示手段
１５　判定結果保持手段
１６　リンク情報削除手段
１７　改良実体確率計算手段
１８　リンク情報統合手段

２０　リンク情報記憶部
２１　関連リンク情報記憶部

【図１】



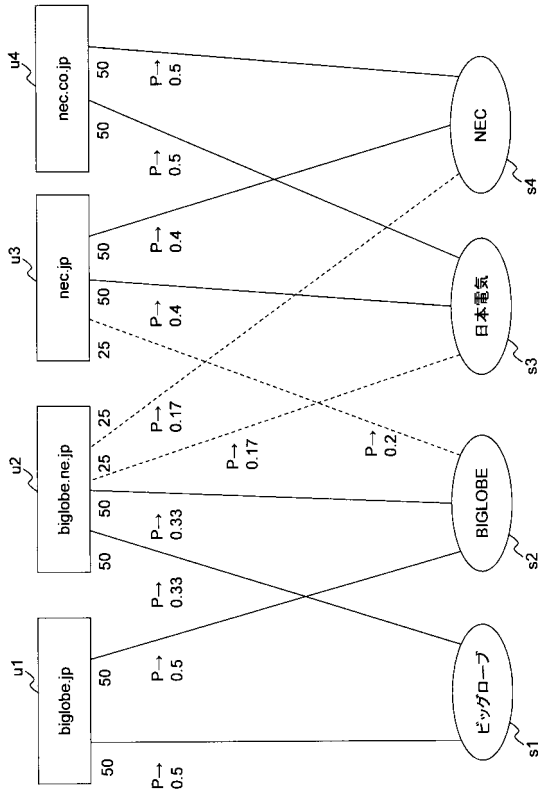
【図２】



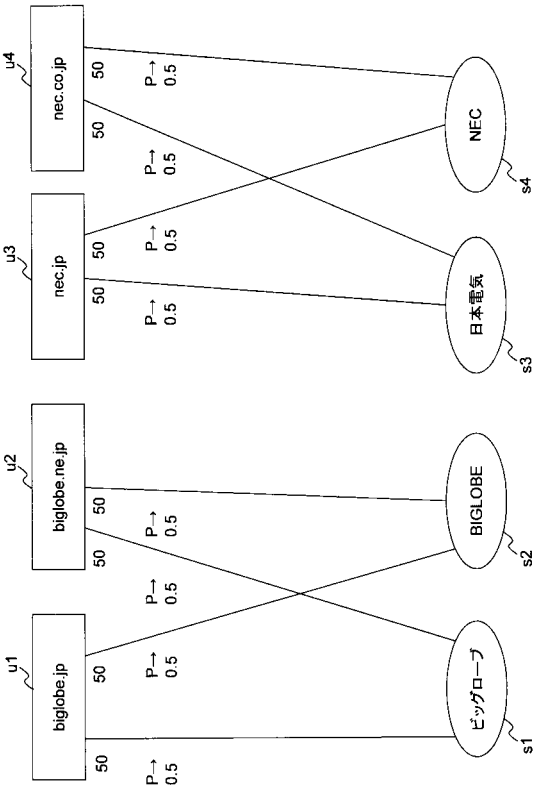
【図３】

アンカーテキスト	URL	リンク回数
ビッグローブ	biglobe.jp	50
ビッグローブ	biglobe.ne.jp	50
BIGLOBE	biglobe.jp	50
BIGLOBE	biglobe.ne.jp	50
BIGLOBE	nec.jp	25
日本電気	biglobe.ne.jp	25
日本電気	nec.jp	50
日本電気	nec.co.jp	50
NEC	biglobe.ne.jp	25
NEC	nec.jp	50
NEC	nec.co.jp	50

【図４】



【図５】



【図６】

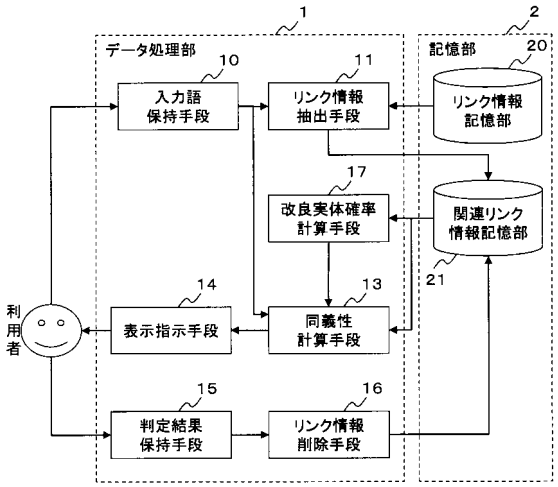
入力語: NEC 2回目

$$\mathrm{NRel(NEC,日本電気)} = \frac{0.5*0.5+0.5*0.5}{0.5+0.5-0.5*0.5+0.5+0.5-0.5*0.5} = 0.5/1.5 = 0.33$$

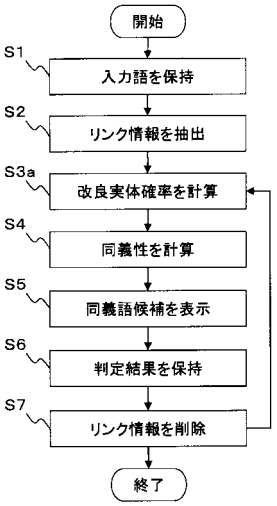
$$\mathrm{NRel(NEC,BIGLOBE)} = 0$$

$$\mathrm{NRel(NEC,ビッグローブ)} = 0$$

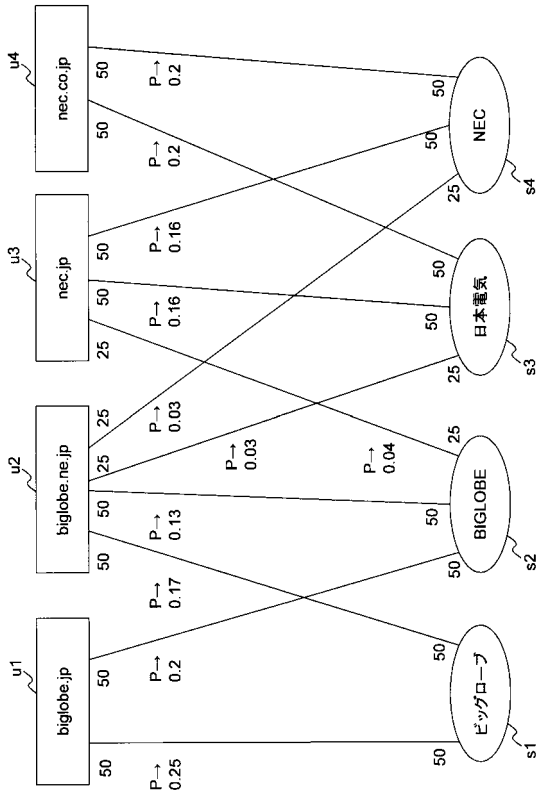
【図７】



【図８】



【図９】



【図１０】

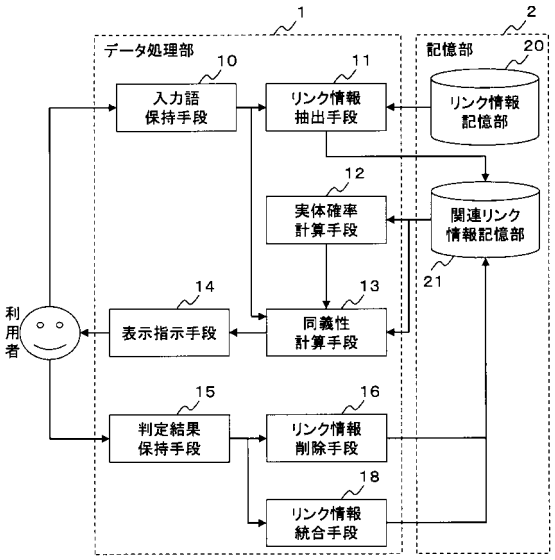
入力語: NEC 1回目

$$\mathrm{NRel}(\mathrm{NEC},\text{日本電気}) = \frac{0.03*0.03+0.16*0.16+0.2*0.2}{0.03+0.03-0.03*0.03+0.16+0.16-0.16*0.16+0.2+0.2-0.2*0.2}$$
$$= 0.00+0.03+0.04/0.03+0.03-0.00+0.16+0.16-0.03+0.2+0.2-0.04$$
$$= 0.07/0.71$$
$$= 0.10$$

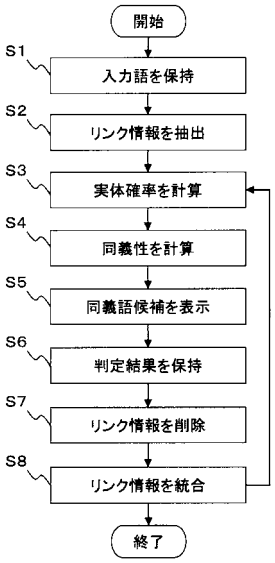
$$\mathrm{NRel}(\mathrm{NEC},\mathrm{BIGLOBE}) = \frac{0.2*0+0.13*0.03+0.04*0.16+0*0.2}{0.2+0-0.2*0+0.13+0.03-0.13*0.03+0.04+0.16-0.04*0.16+0.2+0-0.2*0}$$
$$= 0.00+0.01/0.2+0.13+0.03-0.00+0.04+0.16-0.01+0.2$$
$$= 0.01/0.75$$
$$= 0.01$$

$$\mathrm{NRel}(\mathrm{NEC},\text{ビッグローブ}) = \frac{0.25*0+0.17*0.03+0*0.16+0*0.2}{0.25+0-0.25*0+0.17+0.03-0.17*0.03+0+0.16-0*0.16+0+0.2-0*0.2}$$
$$= 0.01/0.25+0.17+0.03-0.01+0.16+0.2$$
$$= 0.01/0.8$$
$$= 0.01$$

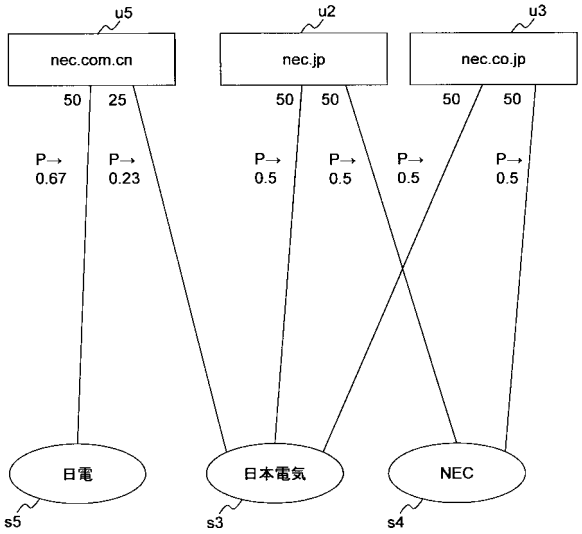
【図１１】



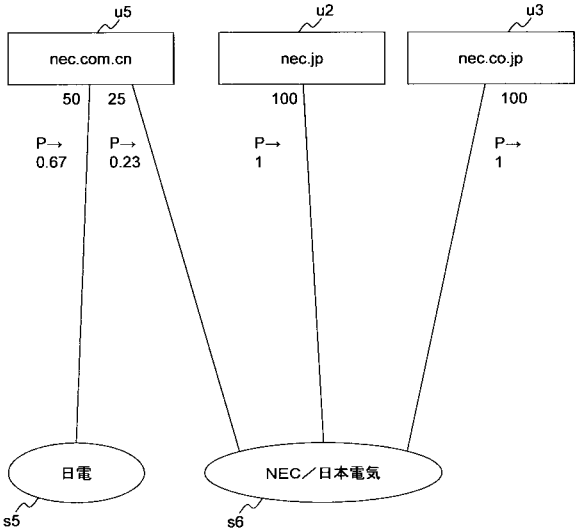
【図１２】



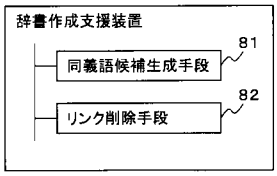
【図１３】



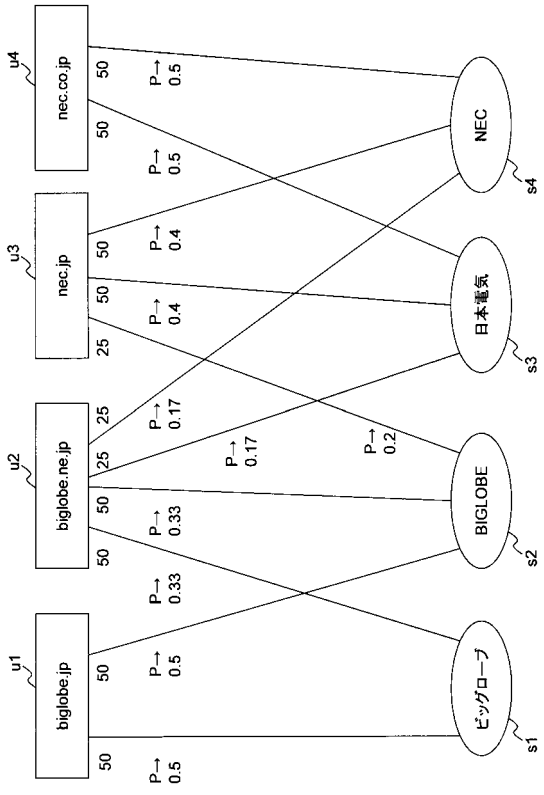
【図１４】



【図１５】



【図１６】



【図１７】

入力語：NEC　1回目

$$\mathrm{NRel}(\mathrm{NEC},\text{日本電気}) = \frac{0.17*0.17+0.4*0.4+0.5*0.5}{0.17+0.17-0.17*0.17+0.4+0.4-0.4*0.4+0.5+0.5-0.5*0.5}$$
$$= 0.03+0.16+0.25/0.34-0.03+0.8-0.16+1-0.25$$
$$= 0.44/1.7$$
$$= 0.26$$

$$\mathrm{NRel}(\mathrm{NEC},\mathrm{BIGLOBE}) = \frac{0.5*0+0.33*0.17+0.2*0.4+0.5*0}{0.5+0-0.5*0+0.33+0.17-0.33*0.17+0.2+0.4-0.2*0.4+0.5+0-0.5*0}$$
$$= 0+0.06+0.08+0/0.5+0.33+0.17-0.06+0.2+0.4-0.08+0.5$$
$$= 0.14/1.96$$
$$= 0.07$$

$$\mathrm{NRel}(\mathrm{NEC},\text{ビッグローブ}) = \frac{0.5*0+0.33*0.17+0*0.4+0*0.5}{0.5+0-0.5*0+0.33+0.17-0.33*0.17+0+0.4-0*0.4+0+0.5-0*0.5}$$
$$= 0+0.06+0+0/0.5+0.33+0.17-0.06+0.4+0.5$$
$$= 0.06/1.84$$
$$= 0.03$$

フロントページの続き

(72)発明者　細見　格
東京都港区芝五丁目７番１号　日本電気株式会社内
(72)発明者　山名　早人
東京都新宿区戸塚町１丁目１０４番地　学校法人早稲田大学内
Ｆターム(参考) 5B091 AA15　AB17　CC16