

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B1)

(11) 特許番号

特許第6919961号
(P6919961)

(45) 発行日 令和3年8月18日(2021.8.18)

(24) 登録日 令和3年7月28日(2021.7.28)

(51) Int.Cl. F 1
G 0 6 F 1 6 / 9 5 3 8 (2 0 1 9 . 0 1) G O 6 F 1 6 / 9 5 3 8

請求項の数 16 (全 38 頁)

<p>(21) 出願番号 特願2021-510244 (P2021-510244)</p> <p>(86) (22) 出願日 令和2年12月4日(2020.12.4)</p> <p>(86) 国際出願番号 PCT/JP2020/045268 審査請求日 令和3年2月24日(2021.2.24)</p> <p>(31) 優先権主張番号 特願2019-230822 (P2019-230822)</p> <p>(32) 優先日 令和1年12月20日(2019.12.20)</p> <p>(33) 優先権主張国・地域又は機関 日本国(JP)</p> <p>特許法第30条第2項適用 集会 第11回計算科学情報技術国際学会 発表日平成30年12月22日</p> <p>特許法第30条第2項適用 集会 第39回国際社会ネットワーク学会サンベルト会議 発表日 令和1年6月22日</p>	<p>(73) 特許権者 317018251 杉原 桂太 愛知県名古屋市天白区植田1-701 パ ンペール植田ヒルズ2-504</p> <p>(74) 代理人 110000578 名古屋国際特許業務法人</p> <p>(72) 発明者 杉原 桂太 愛知県名古屋市天白区植田1-701 パ ンペール植田ヒルズ2-504</p> <p>審査官 早川 学</p>
---	--

最終頁に続く

(54) 【発明の名称】 情報処理システム及び情報処理方法

(57) 【特許請求の範囲】

【請求項1】

文書間の接続関係を表すデータに基づき、少なくとも弱連結で連結された複数の文書で構成される文書ネットワークを判別するように構成される文書ネットワーク判別部と、

判別された前記文書ネットワークに含まれる、二つ以上の文書からのインリンクを有する特定文書を判別するように構成される文書判別部と、

判別された前記特定文書を基準に、前記文書ネットワークに含まれる複数のサブネットワークを判別するように構成されるサブネットワーク判別部と、

判別された前記複数のサブネットワークのそれぞれに対する個別処理を実行することにより、前記文書ネットワークを構成する前記複数の文書のそれぞれのスコアを算出するように構成されるスコア算出部であって、前記個別処理では、対応するサブネットワークに含まれる各文書のスコアを算出するスコア算出部と、

を備え、

前記文書ネットワークには、前記複数のサブネットワークのうち二つ以上のサブネットワークに属する文書である重複文書が一つ以上含まれ、

前記スコア算出部は、一つ以上の前記重複文書のそれぞれに関して、対応する重複文書の前記二つ以上のサブネットワークでのスコアを統合することにより、前記対応する重複文書に対する一つのスコアを算出する情報処理システム。

【請求項2】

請求項1記載の情報処理システムであって、

10

20

前記サブネットワーク判別部は、前記特定文書を境界に有する複数のサブネットワークを判別し、

前記複数のサブネットワークは、前記特定文書が有する二つ以上のインリンクに対応する二つ以上の上流ネットワークであって、上流ネットワークのそれぞれでは、前記特定文書に対応する一つのインリンクを有する、二つ以上の上流サブネットワークと、前記特定文書が有するアウトリンクを通じて前記特定文書と接続される下流サブネットワークとを少なくとも有し、

前記特定文書は、前記二つ以上の上流サブネットワークに属する前記重複文書であり、前記スコア算出部は、前記特定文書の前記上流サブネットワークでのスコアを統合することにより、前記特定文書に対して一つの統合スコアを算出し、前記下流サブネットワークに属する各文書のスコアを、前記特定文書の前記統合スコアを基準に算出する情報処理システム。

10

【請求項 3】

請求項 1 記載の情報処理システムであって、

前記サブネットワーク判別部は、前記複数のサブネットワークとして、前記特定文書が有するインリンク毎のサブネットワークであって、対応するインリンクより上流に位置する文書群と、前記特定文書と、前記特定文書が有するアウトリンクより下流に位置する文書群と、を備えるインリンク毎のサブネットワークを判別する情報処理システム。

【請求項 4】

請求項 1 記載の情報処理システムであって、

前記サブネットワーク判別部は、前記複数のサブネットワークとして、前記特定文書が有するインリンク及びアウトリンクの組合せ毎のサブネットワークであって、前記組合せに対応するインリンクより上流に位置する文書群と、前記特定文書と、前記組合せに対応するアウトリンクより下流に位置する文書群と、を備える組合せ毎のサブネットワークを判別する情報処理システム。

20

【請求項 5】

請求項 3 又は請求項 4 記載の情報処理システムであって、

前記統合は、前記対応する重複文書の、前記二つ以上のサブネットワークでのスコアの合計を算出することにより、又は、前記二つ以上のサブネットワークでのスコアの代表値を算出することにより実現される情報処理システム。

30

【請求項 6】

請求項 1 ~ 請求項 5 のいずれか一項記載の情報処理システムであって、

前記個別処理は、前記対応するサブネットワークにおける文書間の接続関係に基づくエルミート隣接行列を用いて、前記対応するサブネットワークに含まれる各文書のスコアを算出する処理を含む情報処理システム。

【請求項 7】

前記個別処理は、前記対応するサブネットワークに含まれる、アウトリンクを有さない後端文書に対して、ダミー文書を付加することにより、前記後端文書に仮想的にアウトリンクを設けるように、前記対応するサブネットワークを変更し、変更後のサブネットワークにおける文書間の接続関係に基づくエルミート隣接行列を定義する処理を含む請求項 6 記載の情報処理システム。

40

【請求項 8】

前記エルミート隣接行列は、前記対応するサブネットワークを構成する文書 $D[m]$ ($1 \leq m \leq N$ (整数)) 間の接続関係に基づく N 行 N 列のエルミート行列であって、第 p 行第 q 列の成分 $h(p, q)$ が、文書 $D[p]$ から文書 $D[q]$ へのリンクが存在し且つ文書 $D[q]$ から文書 $D[p]$ へのリンクが存在するとき、値 1 を示し、文書 $D[p]$ から文書 $D[q]$ へのリンク及び文書 $D[q]$ から文書 $D[p]$ へのリンクのいずれも存在しないとき、値 0 を示し、文書 $D[p]$ から文書 $D[q]$ へのリンクが存在するが文書 $D[q]$ から文書 $D[p]$ へのリンクが存在しないとき、値 $+i$ (i は虚数単位) を示し、文書 $D[p]$ から文書 $D[q]$ へのリンクが存在しないが文書 $D[q]$ から文書 $D[p]$ へ

50

のリンクが存在するとき、値 $-i$ を示す、対角成分がゼロのエルミート行列に対応する請求項 6 又は請求項 7 記載の情報処理システム。

【請求項 9】

前記個別処理は、前記エルミート隣接行列を変形して特殊エルミート隣接行列を定義し、前記特殊エルミート隣接行列の固有ベクトルを用いて前記対応するサブネットワークに含まれる各文書のスコアを算出する処理を含み、前記エルミート隣接行列は、前記固有ベクトルの各成分を仮に複素平面上に配置したときに、全ての成分が前記複素平面において $/2$ ラジアン の角度範囲内に収まるように変形される請求項 8 記載の情報処理システム。

【請求項 10】

プロセッサと、
前記プロセッサに特定の処理を実行させるための命令を含むメモリと、
を備え、前記特定の処理は、
文書間の接続関係を表すデータに基づき、少なくとも弱連結で連結された複数の文書で構成される文書ネットワークを判別することと、
判別された前記文書ネットワークに含まれる、二つ以上の文書からのインリンクを有する特定文書を判別することと、
判別された前記特定文書を基準に、前記文書ネットワークに含まれる複数のサブネットワークを判別することと、
判別された前記複数のサブネットワークのそれぞれに対する個別処理として、対応するサブネットワークに含まれる各文書のスコアを算出する処理を実行することにより、前記文書ネットワークを構成する前記複数の文書のそれぞれのスコアを算出することと、
を含み、

前記文書ネットワークには、前記複数のサブネットワークのうち二つ以上のサブネットワークに属する文書である重複文書が一つ以上含まれ、
前記文書ネットワークを構成する前記複数の文書のそれぞれのスコアを算出することは、一つ以上の前記重複文書のそれぞれに関して、対応する重複文書の前記二つ以上のサブネットワークでのスコアを統合することにより、前記対応する重複文書に対する一つのスコアを算出することを含む情報処理システム。

【請求項 11】

コンピュータにより実行される情報処理方法であって、
文書間の接続関係を表すデータに基づき、少なくとも弱連結で連結された複数の文書で構成される文書ネットワークを判別することと、
判別された前記文書ネットワークに含まれる、二つ以上の文書からのインリンクを有する特定文書を判別することと、
判別された前記特定文書を基準に、前記文書ネットワークに含まれる複数のサブネットワークを判別することと、
判別された前記複数のサブネットワークのそれぞれに対する個別処理として、対応するサブネットワークに含まれる各文書のスコアを算出する処理を実行することにより、前記文書ネットワークを構成する前記複数の文書のそれぞれのスコアを算出することと、
を含み、

前記文書ネットワークには、前記複数のサブネットワークのうち二つ以上のサブネットワークに属する文書である重複文書が一つ以上含まれ、
前記文書ネットワークを構成する前記複数の文書のそれぞれのスコアを算出することは、一つ以上の前記重複文書のそれぞれに関して、対応する重複文書の前記二つ以上のサブネットワークでのスコアを統合することにより、前記対応する重複文書に対する一つのスコアを算出することを含む情報処理方法。

【請求項 12】

請求項 11 記載の情報処理方法であって、
前記複数のサブネットワークを判別することは、前記特定文書を境界に有する複数のサ

10

20

30

40

50

ブネットワークを判別することを含み、

前記複数のサブネットワークは、前記特定文書が有する二つ以上のインリンクに対応する二つ以上の上流ネットワークであって、上流ネットワークのそれぞれでは、前記特定文書が対応する一つのインリンクを有する、二つ以上の上流サブネットワークと、前記特定文書が有するアウトリンクを通じて前記特定文書と接続される下流サブネットワークとを少なくとも有し、

前記特定文書は、前記二つ以上の上流サブネットワークに属する前記重複文書であり、前記文書ネットワークを構成する前記複数の文書のそれぞれのスコアを算出することは、前記特定文書の前記上流サブネットワークでのスコアを統合することにより、前記特定文書に対して一つの統合スコアを算出し、前記下流サブネットワークに属する各文書のスコアを、前記特定文書の前記統合スコアを基準に算出することを含む情報処理方法。

10

【請求項 1 3】

請求項 1 1 記載の情報処理方法であって、

前記複数のサブネットワークを判別することは、前記特定文書が有するインリンク毎のサブネットワークであって、対応するインリンクより上流に位置する文書群と、前記特定文書と、前記特定文書が有するアウトリンクより下流に位置する文書群と、を備えるインリンク毎のサブネットワークを判別することを含む情報処理方法。

【請求項 1 4】

請求項 1 1 記載の情報処理方法であって、

前記複数のサブネットワークを判別することは、前記特定文書が有するインリンク及びアウトリンクの組合せ毎のサブネットワークであって、前記組合せに対応するインリンクより上流に位置する文書群と、前記特定文書と、前記組合せに対応するアウトリンクより下流に位置する文書群と、を備える組合せ毎のサブネットワークを判別することを含む情報処理方法。

20

【請求項 1 5】

請求項 1 3 又は請求項 1 4 記載の情報処理方法であって、

前記統合は、前記対応する重複文書の、前記二つ以上のサブネットワークでのスコアの合計を算出することにより、又は、前記二つ以上のサブネットワークでのスコアの代表値を算出することにより実現される情報処理方法。

【請求項 1 6】

請求項 1 ~ 請求項 9 のいずれか一項記載の情報処理システムが備える前記文書ネットワーク判別部と、前記文書判別部と、前記サブネットワーク判別部と、前記スコア算出部として、コンピュータを機能させるためのコンピュータプログラム。

30

【発明の詳細な説明】

【関連出願への相互参照】

【0001】

本国際出願は、令和 1 年 1 2 月 2 0 日に日本国特許庁に出願された日本国特許出願第 2 0 1 9 - 2 3 0 8 2 2 号に基づく優先権を主張するものであり、日本国特許出願第 2 0 1 9 - 2 3 0 8 2 2 号の全内容を本国際出願に参照により援用する。

40

【技術分野】

【0002】

本開示は、情報処理システム及び情報処理方法に関する。

【背景技術】

【0003】

ウェブページのランク付けを行う技術が既に知られている（特許文献 1 参照）。この技術の単純な例では、ページランクを、多くのウェブページからリンクされるウェブページほど高く判定する。ページランクの計算には、ウェブページ間の接続関係（換言すれば接続状態）を値 0, 1 で二値表現した隣接行列、及び、隣接行列を変形した値 0, 1 と他の実数とを成分に含む行列が用いられる。

50

【先行技術文献】

【特許文献】

【0004】

【特許文献1】特開2017-102712号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

上述の隣接行列に基づきページランクを判定する従来方法では、ウェブページ間の実際の接続関係に加え、全てのウェブページから全てのウェブページへの仮想的な接続関係を指定する必要がある。このため、ウェブページの良好なランク付けを行うことが難しい。

10

【0006】

そこで、本開示の一側面によれば、従来よりも適切に複数文書のスコアリングを行うことが可能な技術を提供できることが望ましい。

【課題を解決するための手段】

【0007】

本開示の一側面によれば、情報処理システムが提供される。情報処理システムは、文書ネットワーク判別部と、文書判別部と、サブネットワーク判別部と、スコア算出部と、を備える。

【0008】

文書ネットワーク判別部は、文書間の接続関係を表すデータに基づき、少なくとも弱連結で連結された複数の文書で構成される文書ネットワークを判別するように構成される。文書判別部は、判別された文書ネットワークに含まれる、二つ以上の文書からのインリンクを有する特定文書を判別するように構成される。

20

【0009】

サブネットワーク判別部は、判別された特定文書を基準に、文書ネットワークに含まれる複数のサブネットワークを判別するように構成される。スコア算出部は、判別された複数のサブネットワークのそれぞれに対する個別処理を実行することにより、文書ネットワークを構成する複数の文書のそれぞれのスコアを算出するように構成される。個別処理では、対応するサブネットワークに含まれる各文書のスコアが算出される。

【0010】

文書ネットワークには、重複文書が一つ以上含まれる。重複文書のそれぞれは、複数のサブネットワークのうちの一つ以上のサブネットワークに属する文書である。スコア算出部は、一つ以上の重複文書のそれぞれに関して、対応する重複文書の一つ以上のサブネットワークでのスコアを統合することにより、対応する重複文書に対する一つのスコアを算出する。

30

【0011】

本開示の一側面に係る情報処理システムによれば、二つ以上の文書からのインリンクを有する文書が含まれる文書ネットワークにおける複数文書のスコアリングを適切に実行できる。従って、本開示の一側面に係る情報処理システムは、複雑な接続関係を有する文書ネットワークにおける複数文書のスコアリングに大変役立つ。

40

【0012】

本開示の一側面によれば、サブネットワーク判別部は、特定文書を境界に有する複数のサブネットワークを判別してもよい。複数のサブネットワークは、二つ以上の上流サブネットワークと、下流サブネットワークとを少なくとも有し得る。二つ以上の上流サブネットワークは、特定文書が有する二つ以上のインリンクに対応し、上流ネットワークのそれぞれでは、特定文書が、対応する一つのインリンクを有する。下流サブネットワークは、特定文書が有するアウトリンクを通じて特定文書と接続される。

【0013】

この場合、特定文書は、二つ以上の上流サブネットワークに属する重複文書である。スコア算出部は、特定文書の上流サブネットワークでのスコアを統合することにより、特定

50

文書に対して一つの統合スコアを算出し、下流サブネットワークに属する各文書のスコアを、特定文書の統合スコアを基準に算出してもよい。

【0014】

本開示の一側面によれば、サブネットワーク判別部は、複数のサブネットワークとして、特定文書が有するインリンク毎のサブネットワークを判別してもよい。サブネットワークのそれぞれは、対応するインリンクより上流に位置する文書群と、特定文書と、特定文書が有するアウトリンクより下流に位置する文書群と、を備え得る。

【0015】

本開示の一側面によれば、サブネットワーク判別部は、複数のサブネットワークとして、特定文書が有するインリンク及びアウトリンクの組合せ毎のサブネットワークを判別してもよい。サブネットワークのそれぞれは、組合せに対応するインリンクより上流に位置する文書群と、特定文書と、組合せに対応するアウトリンクより下流に位置する文書群と、を備え得る。

10

【0016】

本開示の一側面によれば、統合は、対応する重複文書の、二つ以上のサブネットワークでのスコアの合計を算出することにより実現されてもよい。本開示の一側面によれば、統合は、対応する重複文書の、二つ以上のサブネットワークでのスコアの代表値を算出することにより実現されてもよい。代表値は、平均値であってもよい。

【0017】

本開示の一側面によれば、個別処理は、対応するサブネットワークにおける文書間の接続関係に基づくエルミート隣接行列を用いて、対応するサブネットワークに含まれる各文書のスコアを算出する処理を含んでいてもよい。

20

【0018】

本開示の一側面によれば、個別処理は、対応するサブネットワークに含まれる、アウトリンクを有さない後端文書に対して、ダミー文書を付加することにより、後端文書に仮想的にアウトリンクを設けるように、対応するサブネットワークを変更する処理を含んでいてもよい。個別処理は、変更後のサブネットワークにおける文書間の接続関係に基づくエルミート隣接行列を定義する処理を含んでもよい。

【0019】

本開示の一側面では、エルミート隣接行列は、対応するサブネットワークを構成する文書 $D[m]$ ($1 \leq m \leq N$ (整数)) 間の接続関係に基づく N 行 N 列のエルミート行列であってもよい。

30

【0020】

エルミート隣接行列は、第 p 行第 q 列の成分 $h(p, q)$ が、文書 $D[p]$ から文書 $D[q]$ へのリンクが存在し且つ文書 $D[q]$ から文書 $D[p]$ へのリンクが存在するとき、値 1 を示し、文書 $D[p]$ から文書 $D[q]$ へのリンク及び文書 $D[q]$ から文書 $D[p]$ へのリンクのいずれも存在しないとき、値 0 を示し、文書 $D[p]$ から文書 $D[q]$ へのリンクが存在するが文書 $D[q]$ から文書 $D[p]$ へのリンクが存在しないとき、値 $+i$ (i は虚数単位) を示し、文書 $D[p]$ から文書 $D[q]$ へのリンクが存在しないが文書 $D[q]$ から文書 $D[p]$ へのリンクが存在するとき、値 $-i$ を示す、対角成分がゼロのエルミート行列に対応してもよい。

40

【0021】

個別処理は、エルミート隣接行列を変形して特殊エルミート隣接行列を定義し、特殊エルミート隣接行列の固有ベクトルを用いて、対応するサブネットワークに含まれる各文書のスコアを算出する処理を含んでいてもよい。

【0022】

エルミート隣接行列は、固有ベクトルの各成分を仮に複素平面上に配置したときに、全ての成分が複素平面において $\pm \pi/2$ ラジアン の角度範囲内に収まるように変形されてもよい。

【0023】

50

本開示の一側面によれば、文書間の接続関係を $1, 0, +i, -i$ の4値で表現可能なエルミート隣接行列に対応する特殊エルミート隣接行列を用いて複数の文書をスコアリングする。このため、全文書から全文書への仮想的な接続関係を指定する必要がなく、文書間の接続関係に基づく各文書のスコアリングを従来よりも適切に実現することができる。

【0024】

本開示の一側面によれば、コンピュータプログラムが提供されてもよい。コンピュータプログラムは、上述した情報処理システムが備える文書ネットワーク判別部と、文書判別部と、サブネットワーク判別部と、スコア算出部の少なくとも一つとして、コンピュータを機能させるためのコンピュータプログラムであってもよい。

【0025】

10

本開示の一側面によれば、コンピュータにより実行される情報処理方法が提供されてもよい。情報処理方法は、文書間の接続関係を表すデータに基づき、少なくとも弱連結で連結された複数の文書で構成される文書ネットワークを判別することを含んでいてもよい。

【0026】

情報処理方法は、判別された文書ネットワークに含まれる、二つ以上の文書からのインリンクを有する特定文書を判別することを含んでいてもよい。情報処理方法は、判別された特定文書を基準に、文書ネットワークに含まれる複数のサブネットワークを判別することを含んでいてもよい。

【0027】

情報処理方法は、判別された複数のサブネットワークのそれぞれに対する個別処理として、対応するサブネットワークに含まれる各文書のスコアを算出する処理を実行することにより、文書ネットワークを構成する複数の文書のそれぞれのスコアを算出することを含んでいてもよい。

20

【0028】

文書ネットワークには、複数のサブネットワークのうち二つ以上のサブネットワークに属する文書である重複文書が一つ以上含まれていてもよい。文書ネットワークを構成する複数の文書のそれぞれのスコアを算出することは、一つ以上の重複文書のそれぞれに関して、対応する重複文書の二つ以上のサブネットワークでのスコアを統合することにより、対応する重複文書に対する一つのスコアを算出することを含んでいてもよい。

【0029】

30

本開示の一側面によれば、上述した情報処理システムが実行する手順の少なくとも一部を備える情報処理方法が提供され得る。

【0030】

本開示の一側面によれば、プロセッサと、プロセッサに特定の処理を実行させるための命令を含むメモリと、を備える情報処理システムが提供されてもよい。特定の処理は、上述の情報処理方法に対応する処理であり得る。

【図面の簡単な説明】

【0031】

【図1】第1実施形態の情報処理システムの構成を表すブロック図である。

【図2】情報処理システムに関する機能ブロック図である。

40

【図3】クエリ応答部の詳細を表す機能ブロック図である。

【図4】第2スコアリング部が実行する処理を表すフローチャートである。

【図5】結合ノードを含まない第1の文書ネットワークを示す図である。

【図6】結合ノードを含む第2の文書ネットワークを示す図である。

【図7】第2スコアリング部が実行するスコア算出処理の第1部分を表すフローチャートである。

【図8】第2スコアリング部が実行する副処理を表すフローチャートである。

【図9】ダミーノードが付加された第1の文書ネットワークを示す図である。

【図10】特殊エルミート隣接行列の第1例を示す図である。

【図11】図11Aは、各ノードの複素平面上の配置に関する説明図であり、図11Bは

50

、スコアの算出方法を説明した図である。

【図 1 2】特殊エルミート隣接行列の第 2 例を示す図である。

【図 1 3】先端ノードから第 0 層結合ノードに導かれるサブグラフを示す図である。

【図 1 4】ダミーノードが付加されたサブグラフを示す図である。

【図 1 5】スコア算出処理の第 2 部分を表すフローチャートである。

【図 1 6】第 1 層結合ノードに導かれるサブグラフを示す図である。

【図 1 7】第 2 層結合ノードに導かれるサブグラフを示す図である。

【図 1 8】第 3 層結合ノードに導かれるサブグラフを示す図である。

【図 1 9】スコア算出処理の第 3 部分を表すフローチャートである。

【図 2 0】結合ノードから後端ノードに導かれる非循環型のサブグラフを示す図である。 10

【図 2 1】循環型のサブグラフを示す図である。

【図 2 2】第 2 実施形態のスコア算出処理を表すフローチャートである。

【図 2 3】第 3 の文書ネットワークの例を示す図である。

【図 2 4】図 2 4 A 及び図 2 4 B は、第 2 実施形態における第 3 の文書ネットワークのサブグラフを示す図である。

【図 2 5】第 3 実施形態のスコア算出処理を表すフローチャートである。

【図 2 6】図 2 6 A 及び図 2 6 B は、第 3 実施形態における第 3 の文書ネットワークのサブグラフを示す図である。

【図 2 7】第 4 の文書ネットワークの例を示す図である。

【図 2 8】第 5 の文書ネットワークの例を示す図である。 20

【図 2 9】図 2 9 A 及び図 2 9 B は、第 4 の文書ネットワークのサブグラフを示す図である。

【図 3 0】第 4 実施形態において第 2 スコアリング部が実行する副処理を表すフローチャートである。

【図 3 1】第 5 実施形態において第 2 スコアリング部が実行する副処理を表すフローチャートである。

【図 3 2】第 5 実施形態における特殊エルミート隣接行列に関する説明図である。

【図 3 3】第 5 実施形態における特殊エルミート隣接行列に関する説明図である。

【符号の説明】

【 0 0 3 2】 30

1 ... 情報処理システム、5 ... ユーザ端末、10 ... 演算部、11 ... プロセッサ、15 ... メモリ、20 ... 記憶部、30 ... 通信部、110 ... クローラ、120 ... インデクサ、130 ... クエリ処理部、140 ... クエリ応答部、141 ... 第 1 スコアリング部、143 ... 第 2 スコアリング部、145 ... ランク付け部、147 ... 出力部、210 ... ページリポジトリ、220 ... インデックス記憶部。

【発明を実施するための形態】

【 0 0 3 3】

本開示の例示的实施形態を、以下に図面を参照しながら説明する。

【 0 0 3 4】 40

[第 1 実施形態]

図 1 に示す本実施形態の情報処理システム 1 は、ユーザ端末 5 から入力される検索クエリに回答して、ユーザ端末 5 に検索クエリに対応する文書のリストを提供するように構成される。文書は、ウェブ文書、具体的にはウェブページである。情報処理システム 1 は、通信ネットワークを通じてユーザ端末 5 から利用可能な検索エンジンとして機能する。通信ネットワークは、例えばインターネットである。

【 0 0 3 5】

情報処理システム 1 は、演算部 10 と、記憶部 20 と、通信部 30 とを備える。演算部 10 は、プロセッサ 11 及びメモリ 15 を備える。記憶部 20 は、プロセッサ 11 により実行されるコンピュータプログラム及びデータを記憶する。記憶部 20 は、ハードディスクドライブ及びソリッドステートドライブの一方を備えることができる。 50

【 0 0 3 6 】

通信部 3 0 は、ユーザ端末 5 と通信可能な通信インタフェースを備える。演算部 1 0 は、記憶部 2 0 に記憶されたコンピュータプログラムに従う処理を実行することにより、検索機能を実現する。検索機能を実現するための処理は、具体的には、プロセッサ 1 1 により実行される。図 1 に簡略的に示される情報処理システム 1 は、一つ以上の協働するサーバ装置群で構成され得る。

【 0 0 3 7 】

検索機能は、演算部 1 0 が、図 2 に示すクローラ 1 1 0、インデクサ 1 2 0、クエリ処理部 1 3 0、及び、クエリ応答部 1 4 0 として機能し、記憶部 2 0 が、ページリポジトリ 2 1 0、及びインデックス記憶部 2 2 0 として機能することにより実現される。

10

【 0 0 3 8 】

クローラ 1 1 0 は、周知のクローラと同様に、通信ネットワークに存在する複数のウェブページを収集するように構成される。クローラ 1 1 0 により収集されたウェブページは、ページリポジトリ 2 1 0 に蓄積される。

【 0 0 3 9 】

インデクサ 1 2 0 は、ページリポジトリ 2 1 0 に蓄積された各ウェブページを解析してインデックス化するように構成される。インデックス化により、ウェブページ毎のインデックスデータが生成される。ウェブページ毎のインデックスデータは、インデックス記憶部 2 2 0 に記憶される。

【 0 0 4 0 】

各インデックスデータは、内容インデックス、及び構造インデックスを含む。内容インデックスは、対応するウェブページのキーワード、タイトル、及びキーとなる文章の情報を含む。構造インデックスは、対応するウェブページのハイパーリンク構造を表す情報を含む。インデックスデータの一部は、ウェブページ間の接続関係を表す。

20

【 0 0 4 1 】

クエリ処理部 1 3 0 は、ユーザからの検索クエリを受け付け、検索クエリに対応するウェブページの集合である関連ページ群を、全ウェブページの中から抽出する。ここでいう全ウェブページは、クローラ 1 1 0 により通信ネットワーク内で見つけれられ、インデックス記憶部 2 2 0 にインデックスデータが登録されたウェブページ群に対応する。

【 0 0 4 2 】

具体的に、クエリ処理部 1 3 0 は、インデックス記憶部 2 2 0 が記憶するウェブページの内容インデックスに基づき、検索クエリに対応する語彙を含むウェブページの集合を関連ページ群として、全ウェブページの中から抽出する。抽出された関連ページ群の情報は、クエリ応答部 1 4 0 に提供される。

30

【 0 0 4 3 】

クエリ応答部 1 4 0 は、提供される関連ページ群の情報に基づき、関連ページ群をページランク順に配列した検索結果リストを、検索クエリに対する応答データとして、ユーザ端末 5 に送信する。

【 0 0 4 4 】

関連ページの夫々は、検索クエリとの関連度及び重要度が高いウェブページほど上位にランク付けされ、検索結果リストの上位に配置される。検索結果リストは、従来の検索エンジンからの応答データと同様に、リストアップされた関連ページへのリンクを有したウェブページとして構成される。ここで言うリンクは、所謂ハイパーリンクである。

40

【 0 0 4 5 】

クエリ応答部 1 4 0 は、図 3 に示すように、第 1 スコアリング部 1 4 1 と、第 2 スコアリング部 1 4 3 と、ランク付け部 1 4 5 と、出力部 1 4 7 とを備える。第 1 スコアリング部 1 4 1 は、検索クエリに対応する関連ページ群について、関連ページの夫々を、ページコンテンツの検索クエリとの関連度に基づいてスコアリングする。具体的に、第 1 スコアリング部 1 4 1 は、関連ページの夫々に、第 1 スコアとして、内容得点を与えるように構成される。

50

【 0 0 4 6 】

第2スコアリング部143は、検索クエリとは独立して動作し、クローラ110により収集されたウェブページの夫々に、第2スコアとして、ウェブページ間の接続関係に基づく重要得点を与えるように構成される。

【 0 0 4 7 】

第2スコアは、ウェブページ間の接続関係から重要度が高いと推定されるウェブページほど大きな値を示すように算出される。第2スコアは、多くのリンクが集まるウェブページほど、高い重要得点を持つウェブページからリンクされるウェブページほど、他のウェブページへのリンクの少ないウェブページからのリンクを持つウェブページほど大きな値を示す。

10

【 0 0 4 8 】

ランク付け部145は、第1スコアリング部141が関連ページの夫々に対して算出した第1スコアと、第2スコアリング部143が関連ページの夫々に対して算出した第2スコアとに基づき、関連ページの夫々のページランクを算出するように構成される。

【 0 0 4 9 】

一例によれば、関連ページの夫々のページランクは、第1スコアと第2スコアとの重み付け和に対応する。例えば、第1スコア $\times 1$ 、第2スコア $\times 2$ 、及び、0から1の間の値を採る重み付け係数を用いて、各関連ページのページランク Y は、式 $Y = \quad \cdot X 1 + (1 - \quad) \cdot X 2$ に従って算出され得る。各関連ページのページランクは、検索クエリに基づく内容得点とウェブページ間の接続関係に基づく重要得点とを成分に含む全体得点として理解されてもよい。

20

【 0 0 5 0 】

出力部147は、検索クエリに対応する関連ページ群を、ランク付け部145により算出された各関連ページのページランクに基づき、ページランクの高い順に並べたページリストを、検索結果リストとして検索クエリ送信元のユーザ端末5に送信する。

【 0 0 5 1 】

具体的には、第2スコアリング部143は、図4に示す処理を定期的に行うことにより、インデックス記憶部220が記憶する最新のインデックスデータに基づき、少なくとも弱連結で連結されたウェブページ群毎に、対応するウェブページ群に属する各ウェブページの第2スコアを算出する。

30

【 0 0 5 2 】

図4に示す処理を開始すると、第2スコアリング部143は、全ウェブページの中から、一つ以上の文書ネットワークを抽出する(S110)。第2スコアリング部143は、インデックス記憶部220が記憶するインデックスデータを参照することにより、全ウェブページの中で、少なくとも弱連結で連結されたウェブページ群のそれぞれを、一つの文書ネットワークとして抽出することができる。一つの文書ネットワークは、少なくとも弱連結で連結されたウェブページ群から構成される。

【 0 0 5 3 】

少なくとも弱連結で連結されたノード群から構成されるネットワークは、ノード間のリンクの接続方向を無視したときに、そのネットワークに属するノード群の任意の一つのノードから、残りのノードにリンクをたどって到達可能なネットワークに対応する。

40

【 0 0 5 4 】

すなわち、文書ネットワークは、その文書ネットワークに属するウェブページの任意の一つが、リンクの接続方向を無視したときに、残りのウェブページと少なくとも間接的に接続されるウェブページ群から構成される。

【 0 0 5 5 】

図5及び図6は、異なる二つの文書ネットワークの例を示す。文書ネットワークは、有向グラフとして表現される。図5及び図6における一つの円は、一つのノード、換言すれば一つのウェブページに対応する。同図における矢印は、矢印の始点に対応するウェブページに、矢印の終点に対応するウェブページへのリンク(ハイパーリンク)が形成されて

50

いることを示す。即ち、矢印の始点に対応するウェブページから矢印の終点に対応するウェブページにリンクを介して移動可能であることを意味する。

【 0 0 5 6 】

図5及び図6に示される文書ネットワーク内の各ウェブページは、明らかに、矢印の方向を無視したとき、文書ネットワーク内の他のウェブページと少なくとも間接的に接続されている。以下では、一つの文書ネットワーク内の複数のウェブページのそれぞれを、図において円内に示される数字kを用いて、第kウェブページとも表現する。文書ネットワーク内の各ウェブページのことを、ノードとも表現する。第kノードは、第kウェブページを意味する。

【 0 0 5 7 】

S 1 1 0に続くS 1 2 0において、第2スコアリング部1 4 3は、上記抽出した一つ以上の文書ネットワークのうちの一つを、処理対象の文書ネットワークに選択する。その後、第2スコアリング部1 4 3は、処理対象の文書ネットワーク内の各ノードの第2スコアを算出するために、図7に示すスコア算出処理を実行する(S 1 3 0)。

【 0 0 5 8 】

第2スコアリング部1 4 3は、全ての文書ネットワークに対して、スコア算出処理を実行するまで、スコア算出処理を繰り返し実行する(S 1 2 0 - S 1 4 0)。すなわち、第2スコアリング部1 4 3は、各文書ネットワークを順に処理対象に選択し(S 1 2 0)、選択した処理対象の文書ネットワークに対するスコア算出処理を実行する(S 1 3 0)。

【 0 0 5 9 】

第2スコアリング部1 4 3は、全ての文書ネットワークに対するスコア算出処理を終了すると(S 1 4 0でYes)、図4に示す処理を終了する。第2スコアリング部1 4 3は、このようにして、文書ネットワーク毎に、対応する文書ネットワークを構成する各ウェブページの第2スコアを算出する。算出された第2スコアは、ランク付け部1 4 5に提供される。

【 0 0 6 0 】

図7に示すスコア算出処理(S 1 3 0)を開始すると、第2スコアリング部1 4 3は、処理対象の文書ネットワークにおける、インリンクを持たない先端ノードを判別する(S 2 1 0)。

【 0 0 6 1 】

インリンクを持つノードは、このノードへのリンクが他のノードにおいて形成されたノードを意味する。換言すれば、インリンクを持つウェブページは、このウェブページに移動可能なリンク(ハイパーリンク)が他のウェブページにおいて形成されたウェブページを意味する。以下では、インリンクを持たないノードのことを「先端ノード」とも表現する。

【 0 0 6 2 】

図5においてインリンクを持つノードは、第2、第3、第4、第5、第6、第7、及び第8ノードであり、インリンクを持たないノードは、第1ノードである。図6においてインリンクを持たないノードは、第1、第2、第4、第10、第13、及び第14ノードである。

【 0 0 6 3 】

処理対象の文書ネットワークが、先端ノードを有さない文書ネットワークである場合、文書ネットワークには、インリンクを持たないダミーノードDPが付加される。具体的には、文書ネットワーク内の全てのノードへのアウトリンクを持つダミーノードDPが文書ネットワークに付加される。

【 0 0 6 4 】

アウトリンクを持つノードは、他ノードへのリンクを持つノードを意味する。換言すれば、アウトリンクを持つウェブページは、他のウェブページに移動可能なリンク(ハイパーリンク)が形成されたウェブページを意味する。以下では、アウトリンクを持たないノードのことを「後端ノード」とも表現する。

10

20

30

40

50

【 0 0 6 5 】

文書ネットワークに、インリンクを持たないダミーノードDPが付加された場合、第2スコアリング部143は、ダミーノードDPが付加された文書ネットワークを、処理対象の文書ネットワークとみなし、付加したダミーノードDPを、先端ノードと判別する。

【 0 0 6 6 】

S210に続くS220において、第2スコアリング部143は、複数のインリンクを持つ結合ノードを判別する。一つの結合ノードは、複数のインリンクを持つ一つのノードのことを意味する。

【 0 0 6 7 】

図5に示す文書ネットワークには、結合ノードがない。図6に示す文書ネットワークにおける結合ノードは、二重丸で示される第3、第6、第7、第12、及び第15ノードである。例えば、第3ノードは、第1ノードからのインリンクと、第2ノードからのインリンクと、を有する。

10

【 0 0 6 8 】

S220での処理により、処理対象の文書ネットワークが結合ノードを有することが判明した場合(S230でYes)、第2スコアリング部143は、S250の処理を実行する。処理対象の文書ネットワークが結合ノードを有さないことが判明した場合、第2スコアリング部143は、S240の処理を実行する。

【 0 0 6 9 】

S240において、第2スコアリング部143は、処理対象の文書ネットワーク内の各ノードのスコアを、文書ネットワークに対応するエルミート隣接行列Hを用いて算出する。算出される各ノードのスコアは、ノード間の接続関係に基づくスコアである。

20

【 0 0 7 0 】

第2スコアリング部143は、S240で算出した各ノードのスコアを、各ウェブページの第2スコアとしてランク付け部145に出力する(S245)。その後、図7に示すスコア算出処理を終了する。

【 0 0 7 1 】

S240において、第2スコアリング部143は、同一出願人によって2018年7月13日に出願された国際出願PCT/JP2018/026560と同様の手法で、各ノードのスコアを算出することができる。具体的には、第2スコアリング部143は、図8

30

【 0 0 7 2 】

以下では、スコアの算出方法を説明するために、処理対象の文書ネットワークを構成するノードのそれぞれを、ノードD[m]と表現する。変数mは、値1からNまでの整数値を採る(1 ≤ m ≤ N)。Nは、処理対象の文書ネットワークのノード数Nである。ノードD[m]は、対応する文書ネットワークにおける第mノード、すなわち第mウェブページに対応する。

【 0 0 7 3 】

図8に示す副処理を開始すると、第2スコアリング部143は、処理対象の文書ネットワークに対応するエルミート隣接行列Hを生成する(S1010)。具体的には、第2スコアリング部143は、処理対象の文書ネットワーク内のノード間の接続関係を、値1, 0, +i, -iで表すエルミート隣接行列Hを生成する。ここでiは、虚数単位を表す。

40

【 0 0 7 4 】

エルミート隣接行列Hは、処理対象の文書ネットワークのノード数Nに対応したN行N列(N×N)の行列であり、各成分が、値1, 0, +i, -iのいずれかの値を採る行列である。以下における表現「成分h(p, q)」は、エルミート隣接行列Hにおける第p行第q列の成分を示す。

【 0 0 7 5 】

処理対象の文書ネットワークにおいて、ノードD[p]からノードD[q]へのリンクが存在し且つノードD[q]からノードD[p]へのリンクが存在するとき、対応する成

50

分 $h(p, q)$ は、値 1 に設定される。

【0076】

ノード $D[p]$ からノード $D[q]$ へのリンク及びノード $D[q]$ からノード $D[p]$ へのリンクのいずれもが存在しないとき、対応する成分 $h(p, q)$ は、値 0 に設定される。従って、エルミート隣接行列 H の対角成分 $h(p, p)$ は、値ゼロである。

【0077】

ノード $D[p]$ からノード $D[q]$ へのリンクが存在するがノード $D[q]$ からノード $D[p]$ へのリンクが存在しないとき、対応する成分 $h(p, q)$ は、値 $+i$ に設定される。ノード $D[p]$ からノード $D[q]$ へのリンクが存在しないがノード $D[q]$ からノード $D[p]$ へのリンクが存在するとき、対応する成分 $h(p, q)$ は、値 $-i$ に設定される。

10

【0078】

第2スコアリング部143は、処理対象の文書ネットワークのノード間の接続関係に従って、上述したように各成分 $h(p, q)$ の値を設定し、エルミート隣接行列 H を生成する(S1010)。

【0079】

【数1】

$$H = \begin{bmatrix} 0 & & \dots & & h(1,N) \\ & 0 & h(p,q) & & \\ & & 0 & & \\ \vdots & & & \ddots & \vdots \\ & & & & 0 \\ h(N,1) & & & \dots & 0 \end{bmatrix}$$

20

$$h(p,q) = \begin{cases} 1 \\ 0 \\ +i \\ -i \end{cases}$$

30

上述した規則に従って各成分 $h(p, q)$ の値が設定される場合、第 p 行第 q 列の成分 $h(p, q)$ と対角成分を挟んで対称的な位置にある第 q 行第 p 列の成分 $h(q, p)$ は、成分 $h(p, q)$ の複素共役である。従って、エルミート隣接行列 H は、エルミート行列である。

40

【0080】

S1010では、エルミート隣接行列 H を生成する前に、処理対象の文書ネットワークにおけるアウトリンクを持たない各後端ノードに対し、ダミーノード DP が付加される。

【0081】

後端ノードに付加されるダミーノード DP は、図9に示されるように、後端ノードからのインリンクを一つ持つが、アウトリンクを持たないノードである。図9に示す文書ネットワークは、図5に示す文書ネットワークにおいてアウトリンクを持たない第5及び第8ノードのそれぞれに、ダミーノード DP が付加された文書ネットワークである。S1010では、このようにダミーノード DP が付加された処理対象の文書ネットワークに対して、エルミート隣接行列 H が生成される。

50

【 0 0 8 2 】

続く S 1 0 2 0 において、第 2 スコアリング部 1 4 3 は、上記生成したエルミート隣接行列 H を変形した特殊エルミート隣接行列 H 1 を生成する。変形は、特殊エルミート隣接行列 H 1 の固有ベクトル V の各成分を複素平面に配置したときに、成分の全てが / 2 ラジアン の角度範囲に収まるように行われる。変形に際して、第 2 スコアリング部 1 4 3 は、第 1 補正量 C 1 及び第 2 補正量 C 2 を算出する。

【 0 0 8 3 】

【 数 2 】

$$C1 = \sin\left(\frac{\pi}{2n}\right)$$

$$C2 = \frac{\cos\left(\frac{\pi}{2n}\right)}{\sin\left(\frac{\pi}{2n}\right)}$$

10

第 1 補正量 C 1 及び第 2 補正量 C 2 に含まれるパラメータ n の値は、S 1 1 0 で抽出される一つ以上の文書ネットワークにおけるノード数 N の最大値以上の自然数に定められる。文書ネットワークには、上述したようにダミーノード D P が付加され得る。この場合、ノード数 N は、ダミーノード D P を含む文書ネットワーク内のノード数である。パラメータ n の値が、このように定められることで、固有ベクトル V の成分の全ては、 / 2 ラジアン の角度範囲に収まる。

20

【 0 0 8 4 】

パラメータ n の値が大きいほど、固有ベクトル V の成分は、 / 2 ラジアン の角度範囲より小さい角度範囲内に収まる。成分の全てを / 2 ラジアン の角度範囲に収めること の目的は、成分の全てが複素平面上の一つの象限内に収まるようにするためである。第 2 スコアの良好な算出のために、パラメータ n は、この目的が達成可能な範囲で、小さい値に定められる。上述の第 2 補正量 C 2 は、角度範囲の調整に寄与し、第 1 補正量 C 1 は、第 2 補正量 C 2 によって行列成分の絶対値が変化するのを回避するのに役立つ。

30

【 0 0 8 5 】

第 2 スコアリング部 1 4 3 は、第 1 補正量 C 1 及び第 2 補正量 C 2 の算出後、エルミート隣接行列 H における値 + i の成分を、値 C 1 (C 2 + i) に置換し、値 - i を示す成分を値 C 1 (C 2 - i) に置換する。第 2 スコアリング部 1 4 3 は更に、当該置換後のエルミート隣接行列 H における各行の成分の値 C 1 (C 2 + i) を、同じ行において値 C 1 (C 2 + i) を示す成分の数及び値 1 を示す成分の数の和 W で除算した値 { C 1 (C 2 + i) / W } に変更する。

【 0 0 8 6 】

第 2 スコアリング部 1 4 3 は更に、値 { C 1 (C 2 + i) / W } に変更された成分と対角成分を挟んで対称的な位置にある成分の値 C 1 (C 2 - i) を、値 { C 1 (C 2 + i) / W } の複素共役 { C 1 (C 2 - i) / W } に変更する。第 2 スコアリング部 1 4 3 は、このような置換及び変更によって定義されるエルミート行列を、特殊エルミート隣接行列 H 1 として生成する。

40

【 0 0 8 7 】

エルミート隣接行列 H から特殊エルミート隣接行列 H 1 への変形手順の具体例が図 1 0 に示される。例えば、第 p 1 行における合計 N 個の成分 h (p 1 , 1) , h (p 1 , 2) , ... , h (p 1 , N) の内、値 + i を採る成分及び値 1 を採る成分が合計 W 1 個である場合には、エルミート隣接行列 H における第 p 1 行の値 + i を示す各成分は、値 { C 1 (C 2 + i) / W 1 } に変更される。

50

【 0 0 8 8 】

第 $p 2$ 行における合計 N 個の成分 $h(p 2, 1), h(p 2, 2), \dots, h(p 2, N)$ の内、値 $+i$ を採る成分及び値 1 を採る成分が合計 $W 2$ 個である場合には、エルミート隣接行列 H における第 $p 2$ 行の値 $+i$ を示す各成分は、値 $\{C 1(C 2 + i) / W 2\}$ に変更される。

【 0 0 8 9 】

更に、値 $-i$ を示す成分の値は、対角成分を挟んで対称的な位置にある成分の複素共役に変更される。例えば、値 $\{C 1(C 2 + i) / W 1\}$ を示す成分 $h(p 1, q 1)$ と対角成分を挟んで対称的な位置にある成分 $h(q 1, p 1)$ の値は、 $\{C 1(C 2 - i) / W 1\}$ に変更される。同様に、値 $\{C 1(C 2 + i) / W 2\}$ を示す成分 $h(p 2, q 2)$ と対角成分を挟んで対称的な位置にある成分 $h(q 2, p 2)$ の値は、 $\{C 1(C 2 - i) / W 2\}$ に変更される。

10

【 0 0 9 0 】

続く $S 1 0 3 0$ において、第 2 スコアリング部 1 4 3 は、 $S 1 0 2 0$ で生成した特殊エルミート隣接行列 $H 1$ の固有値及び固有ベクトル V を算出する。特殊エルミート隣接行列 $H 1$ が $N \times N$ の行列であることから、固有ベクトル V は、 N 個の成分を含む N 次元ベクトルである。

【 0 0 9 1 】

以下では、絶対値最大の固有値に対応する固有ベクトル V の各成分を $V[m]$ を用いて表す。変数 m は値 1 から値 N までの整数値を採る。即ち、固有ベクトル V は、 $V = \{V[1], V[2], \dots, V[N]\}$ である。固有ベクトル V の各成分 $V[m]$ ($1 \leq m \leq N$) は、文書ネットワークを構成するノード $D[m]$ ($1 \leq m \leq N$) に対応する。

20

【 0 0 9 2 】

続く $S 1 0 4 0$ において、第 2 スコアリング部 1 4 3 は、特殊エルミート隣接行列 $H 1$ の絶対値最大の固有値に対応する固有ベクトル V の各成分 $V[m]$ ($1 \leq m \leq N$) を、文書ネットワークの始点ノードに対応する成分 E で除算する。始点ノードが第 s ノード $D[s]$ であるとき、成分 E は、固有ベクトル V の第 s 成分 $V[s]$ である ($E = V[s]$)。

【 0 0 9 3 】

始点ノードは、文書ネットワークにおけるノードのうち、最も小さいスコアを付与すべきノードに対応する。始点ノードは、処理対象の文書ネットワーク内でリンクの向きに従って移動可能な先端ノードと後端ノードとの組合せのうち、先端ノードから後端ノードまでのノード数が最も多い組合せに対応する先端ノードに設定され得る。

30

【 0 0 9 4 】

固有ベクトル V の各成分 $V[m]$ ($1 \leq m \leq N$) が成分 E で除算されると、始点ノードに対応する固有ベクトル V の成分は、値 1 に変換される。以下では、除算後の固有ベクトル V を、固有ベクトル $V 1$ と表現する。固有ベクトル $V 1$ は、 $V 1 = \{V[1] / E, V[2] / E, \dots, V[s] / E = 1, \dots, V[N] / E\}$ である。除算により、始点ノードに対応する固有ベクトル $V 1$ の成分は、複素平面において、実軸上に配置される。

【 0 0 9 5 】

$S 1 0 4 0$ での処理を終えると、第 2 スコアリング部 1 4 3 は、除算後の固有ベクトル $V 1$ の各成分 $V 1[m] = V[m] / E$ ($1 \leq m \leq N$) に基づいて、文書ネットワーク内の各ノード $D[m]$ のスコアを算出する ($S 1 0 5 0$)。

40

【 0 0 9 6 】

$S 1 0 5 0$ において、第 2 スコアリング部 1 4 3 は、各成分 $V 1[m]$ ($1 \leq m \leq N$) を、複素平面上で回転変換する。具体的に、第 2 スコアリング部 1 4 3 は、複素平面上において、最も第 1 象限側に位置する成分が、実軸から角度 θ だけ第 4 象限側に位置するように、固有ベクトル $V 1$ の各成分 $V 1[m]$ ($1 \leq m \leq N$) を複素平面上において回転させる。

【 0 0 9 7 】

50

図 1 1 A によれば、始点ノードに対応する成分 $V_1 [s]$ が複素平面の実軸上にある。図 1 1 A 及び図 1 1 B における黒丸及び白丸の夫々は、固有ベクトル V_1 の成分の一つ、換言すれば、文書ネットワーク内のノードの一つに対応し、黒丸は、始点ページに対応する。

【 0 0 9 8 】

上記回転変換によって、始点ノードに対応する成分は、図 1 1 B に示すように、複素平面上で実軸から角度 θ_1 だけ第 4 象限側に位置するように回転移動する。この回転変換は、適切なスコアリングを目的として、始点ノードを実軸から第 4 象限側にずらすために実行される。角度 θ_1 は、回転変換によっても、固有ベクトル V_1 の全成分が依然として第 4 象限に位置する小さい角度に定められる。スコアリングに悪影響がなければ、角度 θ_1 はゼロであってもよい。

10

【 0 0 9 9 】

回転変換後の固有ベクトル V_1 のことを、以下では、固有ベクトル $V_c = \{ V_c [1], V_c [2], \dots, V_c [s], \dots, V_c [N] \}$ と表現する。固有ベクトル V_c の各成分 $V_c [m]$ ($1 \leq m \leq N$) は、複素数である。

【 0 1 0 0 】

S 1 0 5 0 では、各成分 $V_c [m]$ ($1 \leq m \leq N$) の複素平面上の位置に基づいて、各ノードのスコアを算出する。以下では、回転変換後の固有ベクトル V_c の各成分 $V_c [m]$ ($1 \leq m \leq N$) のことを、各ノードのスコア基準値 $V_c [m]$ ($1 \leq m \leq N$) とも表現する。スコア基準値 $V_c [m]$ は、第 m ノードのスコア基準値であり、第 m ノードのスコアリングに用いられる。角度 θ_1 がゼロであるとき、各ノードのスコア基準値 $V_c [m]$ ($1 \leq m \leq N$) は、 $V_1 [m]$ ($1 \leq m \leq N$) に一致する。

20

【 0 1 0 1 】

本明細書において以下に記載される関数 $\arg(x)$ は、複素数 x の複素平面上の偏角であると理解されてよい。 x は、例えば、 $V_c [m]$ である。図 1 1 B に示される成分 $V_c [m]$ の複素平面上の実軸から第 4 象限への角度 $\theta [m]$ は、 $\{ 2\pi - \arg(V_c [m]) \}$ に等しい。以下で表現する $|x|$ は、複素数 x の絶対値を意味する。 $x = V_c [m]$ である場合、 $|x|$ は、図 1 1 B に示す $V_c [m]$ の複素平面上の長さ $L [m]$ に対応する。

30

【 0 1 0 2 】

S 1 0 5 0 において、第 2 スコアリング部 1 4 3 は、各ノードのスコア相当値 $Z [m]$ ($1 \leq m \leq N$) として、各ノードのスコア基準値 $V_c [m]$ ($1 \leq m \leq N$) の複素平面における実軸からの距離に対応する値 $Z [m] = L [m] \cdot \theta [m]$ ($1 \leq m \leq N$) $= |V_c [m]| \cdot \{ 2\pi - \arg(V_c [m]) \}$ を算出する。

【 0 1 0 3 】

別例としてスコア相当値 $Z [m]$ は、式 $Z [m] = |V_c [m]|^{d_1} \cdot \{ 2\pi - \arg(V_c [m]) \}^{d_2}$ に従って算出されてもよい。値 d_1 , d_2 は、ゼロより大きい任意の実数である。 d_1 が大きいほど、始点ノードからの各点のアウトリンク数の少なさに応じて、 $Z [m]$ の値は大きくなる。 d_2 が大きいほど始点ノードからの距離に応じて、 $Z [m]$ の値は大きくなる。

40

【 0 1 0 4 】

その後、第 2 スコアリング部 1 4 3 は、文書ネットワーク内の各ノード $D [m]$ ($1 \leq m \leq N$) のスコアを、スコア相当値 $Z [m]$ に基づいて算出する。S 1 0 5 0 において、ノード $D [m]$ に対応するスコア X は、 $X = Z [m] - Z_0$ に従って算出される。 Z_0 は、例えば、文書ネットワーク全体における $Z [m]$ の最小値である。この場合、最も小さい $Z [m]$ を示すノード $D [m]$ のスコアは、値ゼロである。 Z_0 は、値ゼロであってもよい。すなわち、 Z_0 の項はなくてもよい。

【 0 1 0 5 】

S 2 4 5 では、このようにして算出された各ノードのスコア X が、各ウェブページの第 2 スコアとしてランク付け部 1 4 5 に出力される。

50

【 0 1 0 6 】

別例として、第2スコアリング部143は、S1020で上述の特殊エルミート隣接行列H1に代えて、図12に示す特殊エルミート隣接行列H2を生成してもよい。図12に示される特殊エルミート隣接行列H2は、図10上段に示すエルミート隣接行列Hに対応する特殊エルミート隣接行列H2の例である。

【 0 1 0 7 】

第2スコアリング部143は、特殊エルミート隣接行列H2の生成に際して、エルミート隣接行列Hにおける値 $+i$ の成分を、値 $C1(C2+i)$ に置換し、値 $-i$ を示す成分を値 $C1(C2-i)$ に置換することができる。第2スコアリング部143は更に、次の処理A及び処理Bを行うことができる。

10

【 0 1 0 8 】

(処理A)

第2スコアリング部143は、置換後のエルミート隣接行列Hにおける各行の成分内の値 $C1(C2+i)$ を、同じ行において値 $C1(C2+i)$ 及び値1を示す成分の数 W で除算した値 $\{C1(C2+i)/W\}$ に変更し、更に、値 $\{C1(C2+i)/W\}$ に変更された成分と対角成分を挟んで対称的な位置にある成分内の値 $C1(C2-i)$ を、値 $\{C1(C2+i)/W\}$ の複素共役 $\{C1(C2-i)/W\}$ に変更することができる。

【 0 1 0 9 】

(処理B)

第2スコアリング部143は、置換後のエルミート隣接行列Hにおける各行の成分内の値 $C1(C2-i)$ を、同じ行において $C1(C2-i)$ 及び値1を示す成分の数 Z で乗算した値 $\{C1(C2-i)Z\}$ に変更し、更に、値 $\{C1(C2-i)Z\}$ に変更された成分と対角成分を挟んで対称的な位置にある成分内の値 $C1(C2+i)$ を、値 $\{C1(C2-i)Z\}$ の複素共役 $\{C1(C2+i)Z\}$ に変更することができる。

20

【 0 1 1 0 】

このような置換及び変更によって、特殊エルミート隣接行列H2は生成される。第2スコアリング部143は、処理Aの実行後、処理Bを実行してもよいし、処理Bの実行後、処理Aを実行してもよいし、処理A及び処理Bを同時並行的に実行してもよい。いずれの態様で処理A及び処理Bを実行しても、同じ特殊エルミート隣接行列H2が生成される。

30

【 0 1 1 1 】

図12に示される特殊エルミート隣接行列H2における値 $Z1$ は、第 $p3$ 行における成分 $h(p3, 1), h(p3, 2), \dots, h(p3, N)$ の内、値 $-i$ を採る成分及び値1を採る成分の数に対応する。値 $Z2$ は、第 $p4$ 行における合計 N 個の成分 $h(p4, 1), h(p4, 2), \dots, h(p4, N)$ の内、値 $-i$ を採る成分及び値1を採る成分の数に対応する。第 $p3$ 行は、図12において値 $\{C1(C2-i)Z1/W1\}$ が示される行と理解されてよい。第 $p4$ 行は、図12において値 $\{C1(C2-i)Z2/W2\}$ が示される行と理解されてよい。

【 0 1 1 2 】

第2スコアリング部143は、このように算出した特殊エルミート隣接行列H2を、特殊エルミート隣接行列H1に代えて用いて、S1030-S1050の処理を実行することができる。

40

【 0 1 1 3 】

S250(図7参照)において、第2スコアリング部143は、処理対象の文書ネットワークに含まれる結合ノードの層数 J を判別する。本実施形態では、先端ノードからリンクの向きに従ってノード間を移動したときに、最初に現れる結合ノードが第0層結合ノードと定義される。

【 0 1 1 4 】

第0層結合ノードの次に現れる結合ノードが第1層結合ノードと定義され、第 j 層結合ノードの次に現れる結合ノードが第 $(j+1)$ 層結合ノードと定義される(j は0以上の整

50

数である)。この定義に従えば、文書ネットワーク内に、第(J - 1)層の結合ノードまでが存在するとき、文書ネットワーク内における結合ノードの層数は J である。

【 0 1 1 5 】

図 6 に示す文書ネットワークによれば、第 0 層結合ノードは、第 3 ノード及び第 1 2 ノードであり、第 1 層結合ノードは、第 6 ノードであり、第 2 層結合ノードは、第 7 ノードであり、第 3 層結合ノードは、第 1 5 ノードである。図 6 に示す文書ネットワークにおける結合ノードの層数 J は、4 である。

【 0 1 1 6 】

この説明から理解できるように、先端ノードに依存して複数の層番号を採り得る結合ノードに関しては、採り得る層番号のうちの最大の層番号が、対応する結合ノードに割り当てられる。第 7 ノードは、第 1 層結合ノードではなく、第 2 層結合ノードである。

10

【 0 1 1 7 】

S 2 5 0 の処理後、第 2 スコアリング部 1 4 3 は、j = 0 に設定し (S 2 6 0)、先端ノードから第 j 層 (すなわち第 0 層) 結合ノードに導かれるサブグラフを判別する (S 2 7 0)。

【 0 1 1 8 】

図 6 に示す文書ネットワークの例によれば、S 2 7 0 で判別されるサブグラフは、図 1 3 に示すように、第 1 ノードと第 3 ノードとからなるサブグラフ S G 1 と、第 2 ノードと第 3 ノードとからなるサブグラフ S G 2 と、第 1 0 ノード、第 1 1 ノード、第 1 2 ノード、及び第 2 0 ノードからなるサブグラフ S G 3 と、第 1 3 ノード及び第 1 2 ノードからなるサブグラフ S G 4 である。S 2 7 0 において、サブグラフは、先端ノードと第 0 層結合ノードとの組み合わせ毎に判別される。

20

【 0 1 1 9 】

第 2 スコアリング部 1 4 3 は、S 2 7 0 で判別したサブグラフのそれぞれに関して、図 8 に示す処理と同様の処理を実行する (S 2 8 0)。これにより、サブグラフ毎に、サブグラフ内の各ノードのスコア基準値及びスコアを算出する。

【 0 1 2 0 】

S 2 8 0 において、第 2 スコアリング部 1 4 3 は、判別されたサブグラフを順に処理対象に選択して、図 8 に示す処理を実行することができる。ここでは、処理対象のサブグラフが、図 8 の説明における「処理対象の文書ネットワーク」と同様に扱われて、サブグラフ内の各ノードのスコア基準値及びスコアが算出される。

30

【 0 1 2 1 】

例えば、処理対象のサブグラフが、第 1 0 ノード、第 1 1 ノード、第 1 2 ノード、及び第 2 0 ノードからなるサブグラフ S G 3 である場合には、このサブグラフにおいて、アウトリンクを有さない第 1 2 ノード及び第 2 0 ノードに対しダミーノード D P が付加される (図 1 4 参照)。付加対象のノードには、サブグラフ化前において付加対象のノードが有するアウトリンクの数と同数、ダミーノード D P が付加され得る。

【 0 1 2 2 】

S 2 8 0 では、このようにダミーノード D P が付加されたサブグラフに対応するエルミート隣接行列 H が生成される。このエルミート隣接行列 H に対応する特殊エルミート隣接行列 H 1 又は特殊エルミート隣接行列 H 2 に基づいて、第 1 0 ノード、第 1 1 ノード、第 1 2 ノード、及び第 2 0 ノードのスコア基準値及びスコアが算出される。上述の値 Z 0 は、例えば先端ノードである第 1 0 ノードのスコアがゼロとなるように設定され得る。

40

【 0 1 2 3 】

S 2 8 0 において、サブグラフ毎のスコア基準値及びスコアを算出すると、第 2 スコアリング部 1 4 3 は、サブグラフ間で重複する第 j 層結合ノードのスコア基準値及びスコアを統合する (S 2 8 5)。

【 0 1 2 4 】

S 2 8 5 において、第 2 スコアリング部 1 4 3 は、サブグラフ間で重複する第 j 層結合ノードのスコア基準値を、次のように合成して、処理対象の文書ネットワークにおける第

50

j 層結合ノードのそれぞれに対し唯一のスコア基準値及びスコアを算出する。

【0125】

具体的には、第2スコアリング部143は、一つの第j層結合ノードに関して、当該第j層結合ノードのサブグラフ毎のスコア基準値のうち、複素平面上において実軸からの角度が最も大きいスコア基準値を判別する。その角度が最大のスコア基準値と複素平面上で重なるように、第j層結合ノードの各サブグラフにおけるスコア基準値を複素平面上で回転させる。

【0126】

第2スコアリング部143は、複素平面上で重なった各スコア基準値をベクトル合成し、一つの第j層結合ノードに対して唯一のスコア基準値を、その合成ベクトルに決定する。この唯一のスコア基準値 V_x に基づいて、一つの第j層結合ノードに対応するスコア相当値 $Z_x = |V_x|^{d-1} \cdot \{2 - \arg(V_x)\}^{d-2}$ を算出する。第j層結合ノードのスコア X は、 $X = Z_x - Z_0$ に従って算出され得る。

10

【0127】

サブグラフ間で重複する第j層結合ノードに対して共通する一つのスコア X を与えるために、値 Z_0 は、第j層結合ノードについて上記角度が最大のスコア基準値を有するサブグラフにおける $Z[m]$ の最小値に設定され得る。あるいは、値 Z_0 は、上述したように、値ゼロであってもよい。

【0128】

別例として、第2スコアリング部143は、一つの第j層結合ノードに関して、第j層結合ノードのサブグラフ毎のスコア基準値を重ねないまま複素平面上においてベクトル合成することで、一つの第j層結合ノードに対して唯一のスコア基準値を決定してもよい。

20

【0129】

第2スコアリング部143は、文書ネットワークに複数の第j層結合ノードが存在する場合、S285において、第j層結合ノードのそれぞれに対して上述の処理を実行し、各第j層結合ノードのスコア基準値 V_x 及びスコア X を算出する。これにより第2スコアリング部143は、第j層結合ノード毎に、第j層結合ノードに関するサブグラフ間のスコア基準値 V_c を統合したスコア基準値 V_x 及び対応するスコア X を算出する。

【0130】

S285での処理を終えると、第2スコアリング部143は、変数 j の値を1インクリメントする(S290)、続くS300において、第2スコアリング部143は、変数 j の値が、層数 J 未満であるか否かを判断する。

30

【0131】

変数 j の値が層数 J 以上であると判断すると(S300でNo)、第2スコアリング部143は、S410(図19参照)の処理を実行する。一方、変数 j の値が層数 J 未満であると判断すると(S300でYes)、第2スコアリング部143は、S310(図15参照)の処理を実行する。

【0132】

S310において、第2スコアリング部143は、先端ノードから第j層結合ノードまでのサブグラフを判別する。ここで判別されるサブグラフは、先端ノードから第j層結合ノードまでの間に、他の結合ノードが含まれないサブグラフである。S310では、先端ノードと第j層結合ノードとの組み合わせ毎に、組み合わせに対応する一つの先端ノードと一つの結合ノードとを含むサブグラフが判別される。

40

【0133】

図6に示す文書ネットワークの例によれば、S310で判別されるサブグラフは、図16に示すように、第4ノードと、第5ノードと、第6ノードとからなるサブグラフSG5である。

【0134】

S310での処理によって、該当するサブグラフが存在することが判明した場合(S320でYes)、第2スコアリング部143は、S310で判別されたサブグラフのそれ

50

それぞれに関して、S 2 8 0と同様の処理を実行する(S 3 3 0)。これにより、サブグラフ毎に、サブグラフ内の各ノードのスコア基準値及びスコアを算出する(S 3 3 0)。その後、第2スコアリング部1 4 3は、S 3 4 0の処理を実行する。

【0 1 3 5】

S 3 1 0の処理によって、該当するサブグラフが存在しないことが判明した場合(S 3 2 0でNo)、第2スコアリング部1 4 3は、S 3 3 0の処理を実行せず、S 3 4 0の処理を実行する。

【0 1 3 6】

S 3 4 0において、第2スコアリング部1 4 3は、変数fを値ゼロに設定する。続くS 3 5 0において、第2スコアリング部1 4 3は、第f層結合ノードから第j層結合ノードへのサブグラフを判別する。ここで判別されるサブグラフは、第f層結合ノードから第j層結合ノードまでの間に、他の結合ノードが含まれないサブグラフである。

10

【0 1 3 7】

S 3 5 0では、第f層結合ノードと第j層結合ノードとの組み合わせ毎に、組み合わせに対応する一つの第f層結合ノードと一つの第j層結合ノードとを含むサブグラフが判別される。サブグラフにおいて第f層結合ノードは、先端ノードに対応し、第j層結合ノードは、後端ノードに対応する。

【0 1 3 8】

図6に示す文書ネットワークの例によれば、 $f = 0$ 及び $g = 1$ であるとき、S 3 5 0で判別されるサブグラフは、図16に示す第3ノードと第6ノードとからなるサブグラフS G 6である。

20

【0 1 3 9】

S 3 5 0での処理によって、該当するサブグラフが存在しないことが判明した場合(S 3 6 0でNo)、第2スコアリング部1 4 3は、S 3 7 0の処理を実行することなく、S 3 8 0の処理を実行する。

【0 1 4 0】

一方、該当するサブグラフが存在することが判明した場合(S 3 6 0でYes)、第2スコアリング部1 4 3は、S 3 5 0で判別されたサブグラフのそれぞれに対しS 2 8 0と同様の処理を実行する。これにより、サブグラフ毎に、サブグラフ内の各ノードのスコア基準値及びスコアを算出する(S 3 7 0)。

30

【0 1 4 1】

S 3 7 0において、第2スコアリング部1 4 3は更に、算出したサブグラフ内の各ノードのスコア基準値及びスコアを、既に計算されている第f層結合ノードのスコア基準値及びスコアに応じて修正する。

【0 1 4 2】

サブグラフ内の第f層結合ノードのスコア基準値及びスコアは、S 3 7 0の処理前に計算されている。例えば、 $f = 0$ であるときの第0層結合ノードのスコア基準値及びスコアは、S 2 8 5で計算される。S 3 7 0において、第2スコアリング部1 4 3は、既にスコア基準値及びスコアが計算された第f層結合ノードのスコア基準値及びスコアを基準に、サブグラフ内の残りのノードのスコア基準値及びスコアを修正する。

40

【0 1 4 3】

S 3 7 0の第1例によれば、第2スコアリング部1 4 3は、S 3 7 0の処理前に算出されている第f層結合ノードのスコア基準値がVaである場合、サブグラフ内の各ノードのスコア基準値を次のように修正する。

【0 1 4 4】

すなわち、第2スコアリング部1 4 3は、S 3 7 0で算出した修正前の各ノードのスコア基準値を、第f層結合ノードのスコア基準値が上記Vaと一致するように、複素平面上で回転させる。このようにして回転させたときの各ノードのスコア基準値を、修正後のスコア基準値として決定する。

【0 1 4 5】

50

第2スコアリング部143は、修正後の各ノードのスコア基準値 V_c に基づいて、サブグラフ内の各ノードの修正後のスコア X を算出することができる。スコア X は、第 f 層結合ノードのスコア X が修正前のスコアと同じになるように算出され得る。

【0146】

S370の第2例によれば、第2スコアリング部143は、S370の処理前に算出されている第 f 層結合ノードのスコアが X_a である場合、サブグラフ内の各ノードのスコアを次のように修正する。

【0147】

すなわち、第2スコアリング部143は、S370で算出した修正前の第 f 層結合ノードのスコアと上記 X_a との差分だけ、S370で算出した修正前の各ノードのスコアを加算する。これにより、第2スコアリング部143は、S370で算出した修正前の第 f 層結合ノードのスコアが上記 X_a と一致するように、サブグラフ内の各ノードのスコアを修正する。

10

【0148】

S370での処理を終えると、第2スコアリング部143は、変数 f の値を1インクリメントする(S380)。その後、第2スコアリング部143は、変数 f の値が、変数 j の値未満であるか否かを判断する(S390)。ここで肯定判断すると(S390でYes)、第2スコアリング部143は、S350の処理を実行する。肯定判断すると(S390でNo)、第2スコアリング部143は、S400の処理を実行する。

【0149】

20

S400において、第2スコアリング部143は、S310、S350で判別されたサブグラフ間で重複する第 j 層結合ノードのそれぞれに関して、S330、S370の処理で算出された、対応する第 j 層結合ノードのサブグラフ毎のスコア基準値及びスコアを、S285の処理と同様に統合する。

【0150】

すなわち、第2スコアリング部143は、第 j 層結合ノードのそれぞれに関し、対応する第 j 層結合ノードのサブグラフ毎のスコア基準値 V_c を統合したスコア基準値 V_x を算出し、スコア基準値 V_x に対応するスコア相当値 Z_x に基づくスコア $X = Z_x - Z_0$ を算出する。

【0151】

30

その後、第2スコアリング部143は、変数 j の値を1インクリメントして(S290)、S300～S400の処理を実行する。第2スコアリング部143は、変数 j の値をインクリメントしながら、S300～S400を繰返し実行することにより、第 $(J-1)$ 層結合ノードまでの各ノードのスコア基準値及びスコアを算出する。

【0152】

第 $(J-1)$ 層結合ノードまでの各ノードのスコア基準値及びスコアを算出し終えると、第2スコアリング部143は、S300において否定判断して、S410の処理を実行する(図19参照)。

【0153】

S410より前の処理の流れを、図6に示す文書ネットワークの例に基づいて具体的に説明する。第2スコアリング部143は、 $g = 0$ であるとき、サブグラフSG1(図13参照)に関する処理、及びサブグラフSG2に関する処理の実行により、第1、第2、及び第3ノードのスコア基準値及びスコアを算出する。

40

【0154】

第2スコアリング部143は更に、サブグラフSG3に関する処理、及びサブグラフSG4に関する処理の実行により、第10、第11、第12、第13、及び第20ノードのスコア基準値及びスコアを算出する。

【0155】

その後、 $g = 1$ のプロセスにおいて、第2スコアリング部143は、サブグラフSG5(図16参照)に関する処理を実行し、更には、サブグラフSG6に関する処理を実行し

50

、第4、第5、及び第6ノードのスコア基準値及びスコアを算出する。

【0156】

g = 2のプロセスにおいて、第2スコアリング部143は、サブグラフSG7（図17参照）に関する処理を実行し、更には、サブグラフSG8に関する処理を実行し、第7ノードのスコア基準値及びスコアを算出する。

【0157】

g = 3のプロセスにおいて、第2スコアリング部143は、サブグラフSG9（図18参照）に関する処理を実行し、更には、サブグラフSG10に関する処理を実行し、第8、第14、及び第15ノードのスコア基準値及びスコアを算出する。

【0158】

S410（図19参照）において、第2スコアリング部143は、結合ノードから始まり、非結合ノードの後端ノードで終わる非循環型のサブグラフを判別する。図6に示す文書ネットワークの例によれば、S410で判別されるサブグラフは、図20に示す、第15、第16、第17、第18、及び第19ノードからなるサブグラフSG11である。

【0159】

S410での処理によって、該当するサブグラフが存在しないことが判明した場合（S420でNo）、第2スコアリング部143は、S430の処理を実行せずに、S440の処理を実行する。一方、該当するサブグラフが存在することが判明した場合（S420でYes）、第2スコアリング部143は、S410で判別されたサブグラフ毎に、サブグラフ内の各ノードのスコア基準値及びスコアを算出する（S430）。

【0160】

S430において、第2スコアリング部143は、S370の処理と同様、サブグラフ内の各ノードのスコア基準値及びスコアを、既に計算されている結合ノードのスコア及びスコア基準値に基づいて修正する。このようにして、サブグラフ内の各ノードのスコア及びスコア基準値を決定する。

【0161】

続くS440において、第2スコアリング部143は、循環系のサブグラフを判別する。図6に示す文書ネットワークの例によれば、S440で判別されるサブグラフは、図21に示す、第6、第7、第8、及び第9ノードからなるサブグラフSG12である。

【0162】

循環系のサブグラフが存在しない場合（S450でNo）、第2スコアリング部143は、S460 - S490の処理を実行せずに、S500の処理を実行する。一方、循環系のサブグラフが存在する場合（S450でYes）、第2スコアリング部143は、S460 - S490において、S440で判別されたサブグラフ毎に、サブグラフ内の各ノードのスコア基準値及びスコアを算出する。第2スコアリング部143は、S440で判別されたすべてのサブグラフに関してS470 - S480の処理を実行すると（S490でYes）、S500の処理を実行する。

【0163】

S460において、第2スコアリング部143は、S440で判別されたサブグラフの一つを選択する。S470において、第2スコアリング部143は、選択したサブグラフ内において、既にスコアが算出されているノードのスコア群に基づき、サブグラフ内の各ノードに対する共通の加算スコアを決定する。

【0164】

S470の第1例によれば、第2スコアリング部143は、上記サブグラフ内のスコア群の最大値を、加算スコアに決定する。S470の第2例によれば、第2スコアリング部143は、上記スコア群の平均値を、加算スコアに決定する。

【0165】

S480において、第2スコアリング部143は、決定した加算スコアを、選択したサブグラフ内の各ノードのスコアに加算して、各ノードのスコアを修正する。加算前にスコアが算出されていないノードに対しては、スコアがゼロであるとみなして、上記決定した

10

20

30

40

50

加算スコアを加算することができる。

【0166】

第2スコアリング部143は、このようにして各サブグラフ内のノードのスコアを修正すると、S500の処理を実行する。S500の処理が実行される前に、文書ネットワーク内のすべてのノードのスコアが決定される。

【0167】

S500において、第2スコアリング部143は、決定された文書ネットワーク内の各ノードのスコアを、各ウェブページの第2スコアとしてランク付け部145に出力する。その後、スコア算出処理を終了する。

【0168】

以上に説明した第1実施形態の情報処理システム1は、次のように変形され得る。第1変形例として、第2スコアリング部143は、アウトリンクのないノードに対してダミーノードDPを置かずに、エルミート隣接行列Hを生成し、スコア基準値及びスコアを算出してもよい。

【0169】

第2変形例として、第2スコアリング部143は、各ノードのスコア基準値 $V_c[m]$ を、対応するノードのアウトリンク先の影響を除いた値として修正し、修正したスコア基準値 $V_c^*[m]$ を、修正前のスコア基準値 $V_c[m]$ に代えてを用いて、各ノードに対応するスコア相当値 $Z[m] = |V_c^*[m]|^{d-1} \cdot \{2 - \arg(V_c^*[m])\}^{d-2}$ を算出してもよい。

【0170】

第3変形例として、第2スコアリング部143は、各ノードのスコア基準値 $V_c[m]$ を、対応するノードのインリンク元からの影響を除いた値として修正し、修正したスコア基準値 $V_c^*[m]$ を、修正前のスコア基準値に代えてを用いて、各ノードに対応するスコア相当値 $Z[m] = |V_c^*[m]|^{d-1} \cdot \{2 - \arg(V_c^*[m])\}^{d-2}$ を算出してもよい。第2変形例及び第3変形例は、第1変形例と同様、アウトリンクのないノードに対してダミーノードDPを置かずに、エルミート隣接行列Hを生成して実施され得る。

【0171】

本実施形態の情報処理システム1によれば、ウェブページ間の接続関係を $1, 0, +i, -i$ の4値で表現したエルミート隣接行列Hに対応する特殊エルミート隣接行列H1, H2を用いて複数のウェブページをスコアリングする。このため、全ウェブページから全ウェブページへの仮想的な接続関係を措定する必要がなく、ウェブページ間の接続関係に基づく各ウェブページのスコアリング/ランク付けを従来よりも適切に実現することができる。

【0172】

本実施形態によれば、結合ノードを有する文書ネットワークにおいても、複数のウェブページのスコアリング/ランク付けを、エルミート隣接行列Hを用いて適切に実行できる。従って、出力部147は、第2スコアリング部143からの第2スコアに基づき、ウェブページ間の接続関係に基づいた適切な検索結果リストを、ユーザ端末5に提供することができる。

【0173】

[第2実施形態]

続いて、第2実施形態の情報処理システム1を説明する。第2実施形態の情報処理システム1は、第1実施形態とは異なる内容のスコア算出処理がS130において実行されることを除けば、第1実施形態の情報処理システム1と同様に構成される。従って、以下では、第2スコアリング部143が、S130で実行するスコア算出処理の説明のみをする。以下において言及しない第2実施形態の情報処理システム1の構成は、第1実施形態と同一であると理解されてよい。

【0174】

10

20

30

40

50

第2実施形態において、第2スコアリング部143は、S120で選択した処理対象の文書ネットワークに含まれる各ノードの第2スコアをS130において算出する際、図22に示すスコア算出処理を実行する。

【0175】

図22に示すスコア算出処理を開始すると、第2スコアリング部143は、処理対象の文書ネットワークに含まれるインリンクを持たない先端ノードを判別する(S610)。例えば、処理対象の文書ネットワークが、図23に例示される文書ネットワークである場合、第2スコアリング部143は、第1ノード及び第8ノードを先端ノードとして判別する。

【0176】

その後、第2スコアリング部143は、文書ネットワークに含まれる先端ノードの一つを選択し(S620)、選択した先端ノードを含むサブグラフを判別する(S630)。判別されるサブグラフは、選択した先端ノードと、この先端ノードからリンクの向きに従って移動可能な文書ネットワーク内のすべてのノードとからなるサブグラフである。

【0177】

図23に示される文書ネットワークの例によれば、選択された先端ノードが第1ノードである場合、S630では、図24Aに示すように、第8ノードを除く第1ノードから第9ノードまでのノードからなるサブグラフSG21が判別される。選択された先端ノードが第8ノードである場合、S630では、図24Bに示すように、第3ノードから第9ノードまでのノードからなるサブグラフSG22が判別される。

【0178】

その後、第2スコアリング部143は、S630で判別されたサブグラフを、処理対象の文書ネットワークとみなしたときの図8に示す処理と同様の処理を実行し、サブグラフ内の各ノードのスコア基準値及びスコアを算出する(S640)。

【0179】

S640において、第2スコアリング部143は、後端ノードにダミーノードDPを配置せずに、エルミート隣接行列Hを生成し、スコア基準値及びスコアを算出することができる。第2スコアリング部143は、サブグラフ内における先端ノードのスコア基準値を、複素平面において、実軸上の値1の点、又は実軸に近い第4象限上の特定点に配置するように、サブグラフにおける各ノードのスコア基準値を算出することができる。特定点は、実軸上の値1の点を、角度 θ だけ第4象限側に回転させた点であり得る。

【0180】

S640に続くS650において、第2スコアリング部143は、すべての先端ノードを選択して、S640の処理を実行したか否かを判断する。S650において否定判断すると、第2スコアリング部143は、選択する先端ノードを変更して(S620)、変更後の先端ノードのサブグラフを判別する(S630)。そして、判別したサブグラフのエルミート隣接行列Hに基づいて、サブグラフ内の各ノードのスコア基準値及びスコアを算出する(S640)。

【0181】

第2スコアリング部143は、このようにして、文書ネットワーク内に含まれる先端ノードのそれぞれに対応するサブグラフ毎に、対応するエルミート隣接行列Hに基づく各ノードのスコア基準値及びスコアを算出する。換言すれば、第2スコアリング部143は、結合ノードのインリンク毎のサブグラフを判別し、サブグラフ毎に、対応するエルミート隣接行列Hに基づく各ノードのスコア基準値及びスコアを算出する。その後、第2スコアリング部143は、S650で肯定判断して、S660の処理を実行する。

【0182】

サブグラフ内には、他のサブグラフと重複するノードが含まれるが、S640では、重複するノードのそれぞれに対し、サブグラフ毎に、スコア基準値及びスコアが算出される。

【0183】

10

20

30

40

50

S 6 6 0において、第2スコアリング部143は、文書ネットワーク内の各ノードの第2スコアとして、対応するノードの各サブグラフでのスコアを統合した値を算出する。具体的に、第2スコアリング部143は、一つのノードの第2スコアを、そのノードの各サブグラフでのスコアを合計した値として算出する。

【0184】

あるいは、第2スコアリング部143は、一つのノードの第2スコアを、そのノードの各サブグラフでのスコア基準値の合成ベクトルに基づいて算出してもよい。第2スコアリング部143は、ノード毎に、対応するノードの各サブグラフでのスコア基準値の合成ベクトルを、対応するノードの唯一のスコア基準値 V_x として用いて、式 $Z_x = |V_x|^d \cdot \{2 - \arg(V_x)\}^{d-2}$ に従って、スコア基準値 V_x に対応するスコア相当値 Z_x を算出することができる。第2スコアリング部143は、算出したスコア相当値 Z_x を、対応するノードの第2スコアとして出力することができる。

10

【0185】

上述の第2実施形態によっても、情報処理システム1は、結合ノードを有する文書ネットワークに関する複数のウェブページのスコアリング/ランク付けを、エルミート隣接行列 H を用いて適切に実行することができる。

【0186】

第2実施形態は、第1実施形態と同様に変形されてもよい。すなわち、第1実施形態において、第1、第2、及び第3変形例として説明したスコアの算出に係る変形例は、第2実施形態に適用されてもよい。

20

【0187】

[第3実施形態]

続いて、第3実施形態の情報処理システム1を説明する。第3実施形態の情報処理システム1は、第1実施形態とは異なる内容のスコア算出処理がS130において実行されることを除けば、第1実施形態の情報処理システム1と同様に構成される。従って、以下では、第2スコアリング部143が、S130で実行するスコア算出処理の説明のみをする。以下において言及しない第3実施形態の情報処理システム1の構成は、第1実施形態と同一であると理解されてよい。

【0188】

第3実施形態において、第2スコアリング部143は、S120で選択した処理対象の文書ネットワークに含まれる各ノードの第2スコアをS130において算出する際に、図25に示すスコア算出処理を実行する。

30

【0189】

図25に示すスコア算出処理を開始すると、第2スコアリング部143は、処理対象の文書ネットワーク内に含まれるインリンクを持たない先端ノード、及びアウトリンクを持たない後端ノードを判別する(S710)。

【0190】

アウトリンクを持たない後端ノードがない場合、第2スコアリング部143は、文書ネットワークを有向グラフで表現したときのアウトリンク及びインリンクの合計と、文書ネットワークを無向グラフで表現したときのリンク数とが異なるノードを、形式に後端ノードと判別する。該当するノードがない場合、第2スコアリング部143は、文書ネットワーク内のすべてのノードからのインリンクを有するダミーノード DP を文書ネットワークに追加して、そのダミーノード DP を後端ノードと判別する。

40

【0191】

処理対象の文書ネットワークが、図23に例示される文書ネットワークである場合、第2スコアリング部143は、第1及び第8ノードを、先端ノードとして判別し、第5、第7、及び第9ノードを、後端ノードとして判別する。処理対象の文書ネットワークが、図27及び図28に例示される文書ネットワークである場合、第2スコアリング部143は、第1ノードを、先端ノードとして判別し、第9ノードを、後端ノードとして判別する。

【0192】

50

その後、第2スコアリング部143は、文書ネットワークに含まれる、先端ノードと後端ノードとの組合せ毎のサブグラフを判別する(S720)。サブグラフは、先端ノードから後端ノードまでリンクの向きに従って移動可能なノードの一群からなるサブグラフである。先端ノードと後端ノードとの組合せ毎のサブグラフは、結合ノードが有するインリンク及びアウトリンクの組合せ毎のサブグラフと理解されてもよい。

【0193】

図23に示される文書ネットワークの例によれば、S720において、第2スコアリング部143は、図26Aに示す先端ノードが第1ノードである三つのサブグラフSG31、SG32、SG33、及び、図26Bに示す先端ノードが第8ノードである三つのサブグラフSG34、SG35、SG36を判別する。

10

【0194】

図27に示される文書ネットワークの例によれば、S720において、第2スコアリング部143は、図29Aに示す第3ノードの第一のアウトリンクを通る先端ノードが第1ノード及び後端ノードが第9ノードであるサブグラフと、図29Bに示す第3ノードの第二のアウトリンクを通る先端ノードが第1ノード及び後端ノードが第9ノードであるサブグラフとを判別する。

【0195】

その後、第2スコアリング部143は、判別したサブグラフの一つを選択し(S730)、選択したサブグラフを、処理対象の文書ネットワークとみなしたときの図8に示す処理と同様の処理を実行し、サブグラフ内の各ノードの第1仮スコアを算出する(S740)。

20

【0196】

S740において、第2スコアリング部143は、第mノードの第1仮スコア $X_{p1}[m]$ を、式 $X_{p1}[m] = \{ (2 - \arg(Vc[m])) / (\sqrt{2n}) \}^{d3}$ に従って算出することができる。Vc[m]は、第mノードのスコア基準値である。d3は、0より大きい任意の実数である。d3が大きいほど、第1仮スコア $X_{p1}[m]$ は、インリンクを持たない先端ノードからの距離(リンクの数)に応じて大きくなる。

【0197】

第2スコアリング部143は、すべてのサブグラフに関してS740の処理を実行するまで(S750でNo)、サブグラフのそれぞれを順に選択し(S730)、S740の処理を実行する。これにより、第2スコアリング部143は、サブグラフ毎に、当該サブグラフ内の各ノードの第1仮スコアを算出する(S740)。

30

【0198】

すべてのサブグラフに関して、S740の処理を実行すると(S750でYes)、第2スコアリング部143は、文書ネットワーク内のノード毎に、当該ノードの第2仮スコアとして、当該ノードの第1仮スコアの平均値を算出する(S760)。一つのノードの第2仮スコアは、対応するノードの第1仮スコアの合計を、対応するノードが属するサブグラフの数で除算した値である。

【0199】

その後、第2スコアリング部143は、文書ネットワーク内の各ノードの第2仮スコアに基づいて、各ノードの第3仮スコアを算出する(S770)。S770において、第2スコアリング部143は、第mノードの第3仮スコア $X_{p3}[m]$ を、第mノードの第2仮スコア $X_{p2}[m]$ 及びスコア基準値Vc[m]を用いて、式 $X_{p3}[m] = X_{p2}[m] \cdot |Vc[m]|$ に従い算出する。

40

【0200】

その後、第2スコアリング部143は、文書ネットワーク内の各ノードの第2スコアを、各ノードの第3仮スコアを用いて算出する(S780)。具体的には、第2スコアリング部143は、文書ネットワーク内の第mノードの第2スコアを、第3仮スコア $X_{p3}[m]$ を値 M^{d4} で除算した値 $X_{p3}[m] / M^{d4}$ に算出する。ここで、Mは、後端ノードから第mノードに到達可能な各ノードのアウトリンクの数の積である。

50

【0201】

上述の第3実施形態によっても、結合ノードを有する文書ネットワークに関し、複数のウェブページのスコアリング/ランク付けを、エルミート隣接行列Hを用いて適切に実行することができる。

【0202】

[第4実施形態]

続いて、第4実施形態の情報処理システム1を説明する。第4実施形態の情報処理システム1は、第2スコアリング部143が、図8に示す副処理に代えて図30に示す副処理を実行する点で、第1実施形態とは異なる。一方、第4実施形態の情報処理システム1は、その他の点で基本的に第1実施形態と同じである。従って、以下では、第4実施形態の情報処理システム1の第1実施形態とは異なる構成を選択的に説明し、第1実施形態とは同一構成の説明を省略する。

10

【0203】

図30に示す副処理は、S240で実行される。更に、S280, S330, S370, S430では、スコア基準値及びスコアの算出に際し、図30に示す副処理と同様の処理が実行される。

【0204】

図30に示す副処理を開始すると、第2スコアリング部143は、S1010と同様に、処理対象の文書ネットワークに対応するエルミート隣接行列Hを生成する(S1110)。

20

【0205】

その後、第2スコアリング部143は、上記生成したエルミート隣接行列Hを変形した特殊エルミート隣接行列H3を生成する(S1120)。特殊エルミート隣接行列H3は、S1020において生成される特殊エルミート隣接行列H1における対角成分の全てを、値0から値-1に置換することによって生成される。

【0206】

すなわち、第2スコアリング部143は、S1110で生成したエルミート隣接行列HをS1020での処理と同様に變形することにより、特殊エルミート隣接行列H1を生成し、特殊エルミート隣接行列H1における対角成分の全てを、値0から値-1に置換することによって、特殊エルミート隣接行列H3を生成することができる。

30

【0207】

続くS1130において、第2スコアリング部143は、列ベクトルBを生成する。S1120で生成される特殊エルミート隣接行列H3は、処理対象の文書ネットワークのノード数Nに対応したN行N列(N×N)の行列である。

【0208】

S1130で生成される列ベクトルBは、N行1列の行列に対応し、列ベクトルBは、各成分が、対応するノードがインリンクを有するノードであるか否かに応じた値を示すように生成される。

【0209】

具体的に、第2スコアリング部143は、インリンクを有さないノードに対応する成分を値-1に設定し、インリンクを有するノードに対応する成分を値0に設定するように、列ベクトルBを生成する。

40

【0210】

続くS1140において、第2スコアリング部143は、特殊エルミート隣接行列H3及び列ベクトルBを含む次の連立方程式を解くことにより、連立方程式の解に対応するスコア基準ベクトルUの各成分u[m](1 ≤ m ≤ N)を求める。スコア基準ベクトルUは、列ベクトルBと同様にN行1列の行列に対応する。下式では、スコア基準ベクトルUの第m行成分をu[m]で表現し、列ベクトルBの第m行成分をb[m]で表し、特殊エルミート隣接行列H3を、H'で表す。

【0211】

50

【数3】

$$H'U = B$$

$$U = \begin{bmatrix} u[1] \\ u[2] \\ u[3] \\ \vdots \\ u[N] \end{bmatrix}, B = \begin{bmatrix} b[1] \\ b[2] \\ b[3] \\ \vdots \\ b[N] \end{bmatrix}$$

10

20

例えば、第1ノードのみがインリンクを有さないノード数が5の例示的かつ単純な文書ネットワークを想定する。この場合、上記連立方程式は、例示的な特殊エルミート隣接行列H3を用いて、次のように表され得る。

【0212】

【数4】

$$\begin{bmatrix} -1 & C1(C2+i) & 0 & 0 & 0 \\ C1(C2-i) & -1 & C1(C2+i)/2 & C1(C2+i)/2 & 0 \\ 0 & C1(C2-i)/2 & -1 & 0 & 0 \\ 0 & C1(C2-i)/2 & 0 & -1 & C1(C2+i) \\ 0 & 0 & 0 & C1(C2-i) & -1 \end{bmatrix} \begin{bmatrix} u[1] \\ u[2] \\ u[3] \\ u[4] \\ u[5] \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

30

続くS1150において、第2スコアリング部143は、S1140で算出したスコア基準ベクトルUの各成分u[m]に基づいて、各ノードのスコア基準値Vc[m]を決定する。

40

【0213】

第2スコアリング部143は、第1実施形態における固有ベクトルV1と同様に、スコア基準ベクトルUの各成分u[m] (1 ≤ m ≤ N)を、文書ネットワークの始点ノードに対応する成分E = u[s]で除算することによって、補正することができる。その補正值u[m] / E (1 ≤ m ≤ N)を、複素平面上で角度θだけ実軸から第4象限側に回転移動させるように補正することができる。この補正值を、スコア基準値Vc[m]に決定することができる。

【0214】

続くS1160において、第2スコアリング部143は、各ノードのスコア相当値Z[m] (1 ≤ m ≤ N)として、各ノードのスコア基準値Vc[m] (1 ≤ m ≤ N)に基づい

50

た値 $Z[m] = |Vc[m]| \cdot \{2^{-\arg(Vc[m])}\}$ を算出する。第1実施形態と同様に、スコア相当値 $Z[m]$ は、式 $Z[m] = |Vc[m]|^{d-1} \cdot \{2^{-\arg(Vc[m])}\}^{d-2}$ に従って算出されてもよい。

【0215】

第2スコアリング部143は更に、文書ネットワーク内の各ノード $D[m]$ ($1 \leq m \leq N$) のスコア X を、スコア相当値 $Z[m]$ に基づいて算出する (S1160)。第2スコアリング部143は、第1実施形態と同様に、ノード $D[m]$ に対応するスコア X を、 $X = Z[m] - Z_0$ に従って算出することができる。 Z_0 は、例えば文書ネットワーク全体における $Z[m]$ の最小値である。 Z_0 は、値ゼロであってもよい。

【0216】

本実施形態では、このように各ノードのスコア X が算出される。算出されたスコア X の扱いは、第1実施形態と同様である。本実施形態によれば、第1実施形態のように固有値及び固有ベクトルを算出することなく、エルミート隣接行列 H を用いて各ノードのスコア X を算出することができる。

【0217】

[第5実施形態]

続いて、第5実施形態の情報処理システム1を説明する。第5実施形態の情報処理システム1は、第2スコアリング部143が、図8に示す副処理に代えて図31に示す副処理を実行する点で、第1実施形態とは異なる。一方、第5実施形態の情報処理システム1は、その他の点で基本的に第1実施形態と同じである。従って、以下では、第5実施形態の情報処理システム1の第1実施形態とは異なる構成を選択的に説明し、第1実施形態とは同一構成の説明を省略する。

【0218】

図31に示す副処理は、S240で実行される。更に、S280, S330, S370, S430では、スコア基準値及びスコアの算出に際し、図31に示す副処理と同様の処理が実行される。

【0219】

図31に示す副処理を開始すると、第2スコアリング部143は、S1010と同様に、処理対象の文書ネットワークに対応するエルミート隣接行列 H を生成する (S1210)。

【0220】

その後、第2スコアリング部143は、上記生成したエルミート隣接行列 H を変形した特殊エルミート隣接行列 H_4 を生成する (S1220)。特殊エルミート隣接行列 H_4 を生成するために、第2スコアリング部143は、エルミート隣接行列 H において、値 $+i$ の成分を全て値0に置き換えることができる。更に、値 $-i$ の成分を、全て値 C_1 ($C_2 - i$) に置き換えることができる。この置換により、図32上段に示す例示的なエルミート隣接行列 H は、図32下段に示す行列 $H^{(1)}$ に置換される。

【0221】

第2スコアリング部143は更に、上記行列 $H^{(1)}$ において、アウトリンクの数が2以上のノードに対応する列、換言すれば、エルミート隣接行列 H において値 $-i$ を有する成分の数が2以上である列における値 C_1 ($C_2 - i$) の成分を、値 C_1 ($C_2 - i$) / R に置換する。値 R は、アウトリンクの数であり、エルミート隣接行列 H の対応する列において値 $-i$ を有する成分の個数に対応する。この置換により、図32下段に示される行列 $H^{(1)}$ は、図33上段に示される行列 $H^{(2)}$ に置換される。

【0222】

第2スコアリング部143は更に、図33下段に示すように、行列 $H^{(2)}$ における対角成分を全て値 -1 に置換して、特殊エルミート隣接行列 H_4 を生成する。続くS1230において、第2スコアリング部143は、S1130での処理と同様に、列ベクトル B を生成する。

【0223】

10

20

30

40

50

その後、第2スコアリング部143は、S1140での処理と同様に、特殊エルミート隣接行列H4及び列ベクトルBを含む連立方程式を解くことにより、連立方程式の解に対応するスコア基準ベクトルUの各成分 $u[m]$ ($1 \leq m \leq N$)を求める(S1240)。

【0224】

S1240での処理は、特殊エルミート隣接行列H3に代えて、特殊エルミート隣接行列H4が用いられる点を除けば、S1140での処理と同じである。連立方程式の解において、インリンクを持たない先端ノードに対応する成分は、実数1を示す。

【0225】

続くS1250において、第2スコアリング部143は、S1240で算出したスコア基準ベクトルUの各成分 $u[m]$ に基づいて、各ノードのスコア基準値 $Vc[m]$ を決定する。本実施形態によれば、第2スコアリング部143は、各ノードのスコア基準値 $Vc[m]$ を、スコア基準ベクトルUの、対応する成分 $u[m]$ と同じ値に決定することができる。

10

【0226】

続くS1260において、第2スコアリング部143は、各ノードのスコア相当値 $Z[m]$ ($1 \leq m \leq N$)として、各ノードのスコア基準値 $Vc[m]$ ($1 \leq m \leq N$)に基づいた次式に従う値 $Z[m]$ を算出する。値 d は、ゼロより大きい任意の実数である。値 d は、値1であってもよい。

【0227】

【数5】

20

$$Z[m] = |Vc[m]| \times \left\{ \frac{|2\pi - \arg(Vc[m])|}{\frac{\pi}{2n}} \right\}^d$$

第2スコアリング部143は更に、文書ネットワーク内の各ノード $D[m]$ ($1 \leq m \leq N$)のスコア X を、スコア相当値 $Z[m]$ に基づいて算出する(S1260)。第2スコアリング部143は、第1実施形態と同様に、ノード $D[m]$ に対応するスコア X を、 $X = Z[m] - Z_0$ に従って算出することができる。 Z_0 は、例えば、文書ネットワーク全体における $Z[m]$ の最小値である。 Z_0 は、値ゼロであってもよい。

30

【0228】

本実施形態では、このように各ノードのスコア X が算出される。算出されたスコア X の扱いは、第1実施形態と同様である。本実施形態によれば、第1実施形態のように固有値及び固有ベクトルを求めることなく、エルミート隣接行列Hを用いて、各ノードのスコア X を算出することができる。

【0229】

以上に、本開示の例示的实施形態を説明したが、本開示は、上述の実施形態に限定されない。本開示は、ウェブ文書に限定されないリンク/引用関係を持つ文書のスコアリングに適用されてもよい。第4実施形態及び第5実施形態に係る技術的思想は、第2実施形態又は第3実施形態に適用されてもよい。

40

【0230】

先端ノードを有さない文書ネットワークに、インリンクを持たないダミーノードDPであって、文書ネットワーク内の全てのノードへのアウトリンクを持つダミーノードDPを付加する技術は、第1実施形態から第5実施形態に適用され得る。更に言えば、このインリンクを持たないダミーノードDPは、先端ノードを有する文書ネットワークに付加されてもよい。

【0231】

同様に、後端ノードを有さない文書ネットワークに、文書ネットワーク内のすべてのノードからのインリンクを有するダミーノードDPであって、アウトリンクを持たないダミ

50

ーノードDPを付加する技術は、第1実施形態から第5実施形態に適用され得る。更に言えば、このアウトリンクを持たないダミーノードDPは、後端ノードを有する文書ネットワークに付加されてもよい。

【0232】

上記実施形態における1つの構成要素が有する機能は、複数の構成要素に分散して設けられてもよい。複数の構成要素が有する機能は、1つの構成要素に統合されてもよい。上記実施形態の構成の一部は、省略されてもよい。特許請求の範囲に記載の文言から特定される技術思想に含まれるあらゆる態様が本開示の実施形態である。

【要約】

10

本開示の一側面によれば、少なくとも弱連結で連結された複数の文書で構成される文書ネットワークが判別される。文書ネットワークに含まれる、二つ以上の文書からのインリンクを有する特定文書が判別される。特定文書を基準に、複数のサブネットワークが判別される。サブネットワークのそれぞれに対する個別処理の実行により、文書ネットワークを構成する複数の文書のそれぞれのスコアが算出される。個別処理では、対応するサブネットワークに含まれる各文書のスコアが算出される。二つ以上のサブネットワークに属する重複文書のそれぞれに関しては、対応する重複文書の二つ以上のサブネットワークでのスコアが統合される。

【図1】

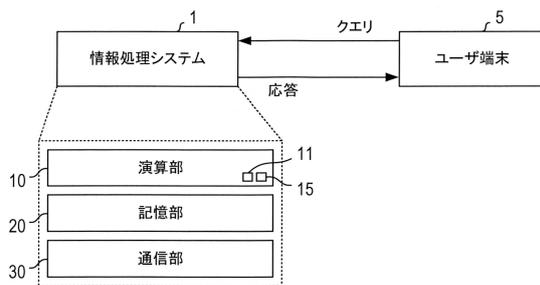


FIG. 1

【図2】

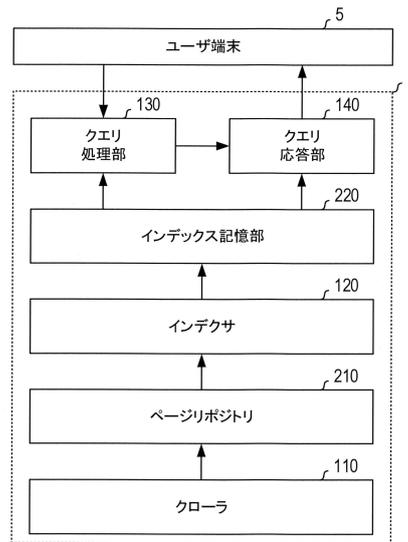


FIG. 2

【図3】

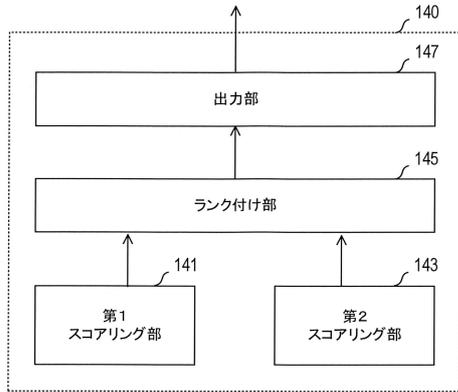


FIG. 3

【図4】

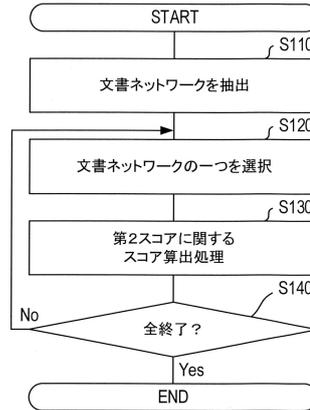


FIG. 4

【図5】

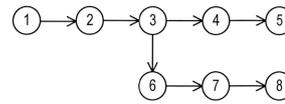


FIG. 5

【図6】

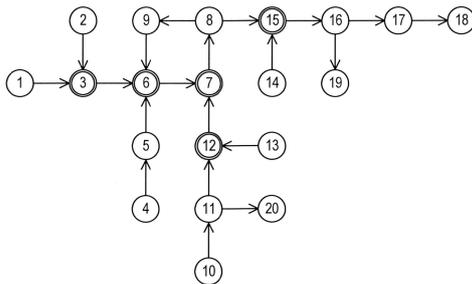


FIG. 6

【図7】

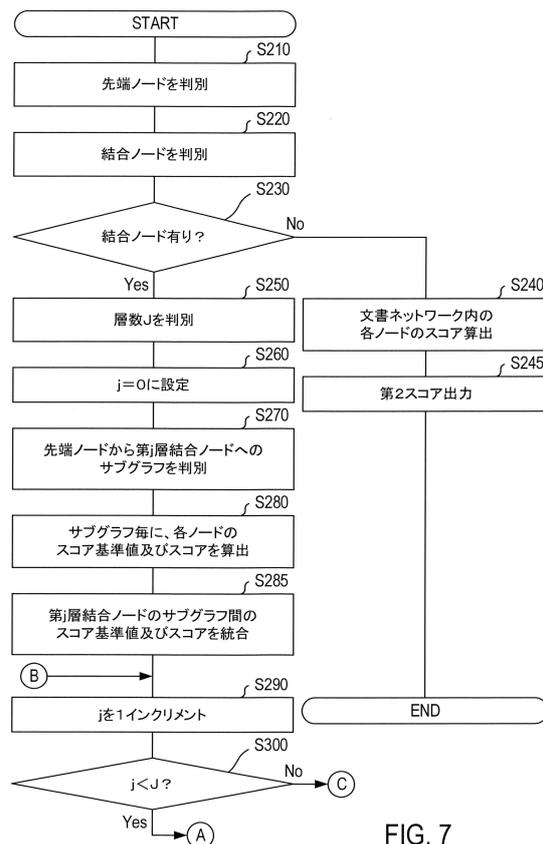


FIG. 7

【 図 8 】

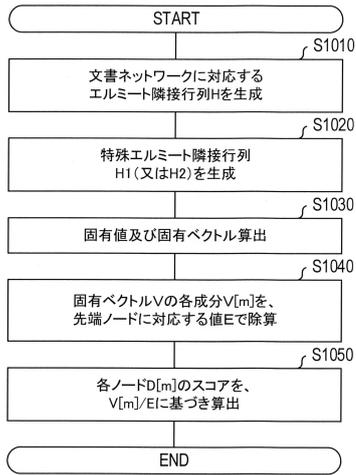


FIG. 8

【 図 9 】

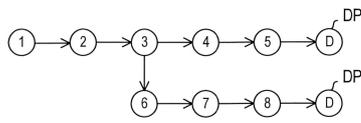


FIG. 9

【 図 1 1 】

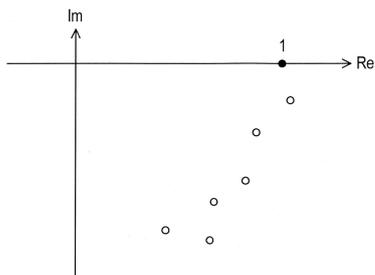


FIG. 11A

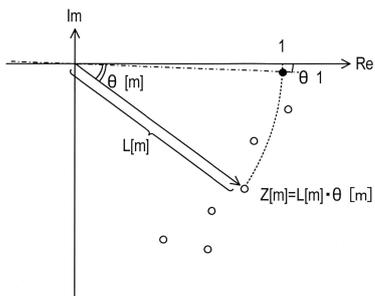


FIG. 11B

【 図 1 0 】

$$H = \begin{bmatrix} & & & & \dots & -i & -i & \dots \\ & & & & & \vdots & \vdots & \\ \dots & +i & +i & \dots & 0 & \dots & -i & -i & \dots \\ & \vdots & \vdots & & & & & & \\ \dots & +i & +i & \dots & & & & & \end{bmatrix}$$

↓

$$H1 = \begin{bmatrix} & & & & \dots & C1(C2-i)/W1 & C1(C2-i)/W1 & \dots \\ & & & & & \vdots & \vdots & \\ \dots & C1(C2+i)/W1 & C1(C2+i)/W1 & \dots & 0 & \dots & C1(C2-i)/W2 & C1(C2-i)/W2 & \dots \\ & \vdots & \vdots & & & & & & \\ \dots & C1(C2+i)/W2 & C1(C2+i)/W2 & \dots & & & & & \end{bmatrix}$$

FIG. 10

【 図 1 2 】

$$H2 = \begin{bmatrix} & & & & \dots & C1(C2-i) \frac{Z1}{W1} & C1(C2-i) \frac{Z1}{W1} & \dots \\ & & & & & \vdots & \vdots & \\ \dots & C1(C2+i) \frac{Z1}{W1} & C1(C2+i) \frac{Z1}{W1} & \dots & 0 & \dots & C1(C2-i) \frac{Z2}{W2} & C1(C2-i) \frac{Z2}{W2} & \dots \\ & \vdots & \vdots & & & & & & \\ \dots & C1(C2+i) \frac{Z2}{W2} & C1(C2+i) \frac{Z2}{W2} & \dots & & & & & \end{bmatrix}$$

FIG. 12

【 図 1 3 】

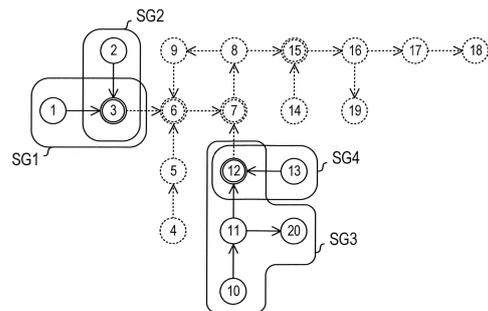


FIG. 13

【 図 1 4 】

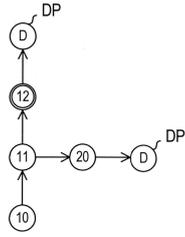


FIG. 14

【 図 1 5 】

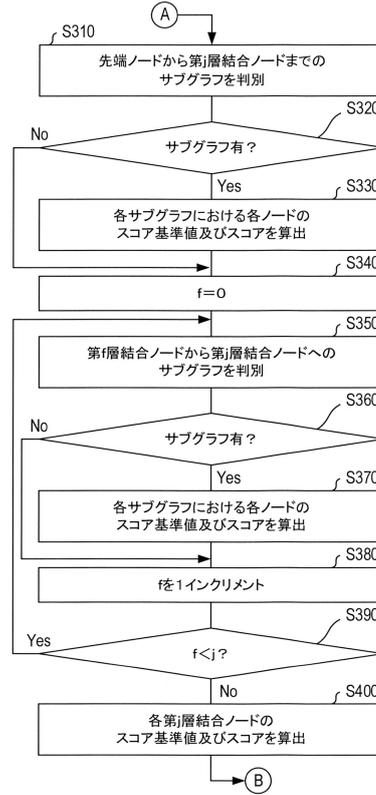


FIG. 15

【 図 1 6 】

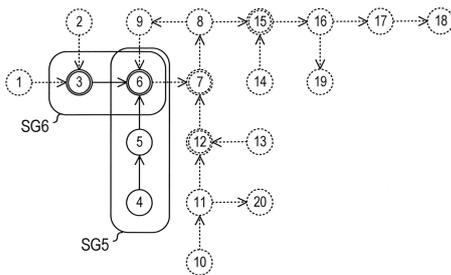


FIG. 16

【 図 1 8 】

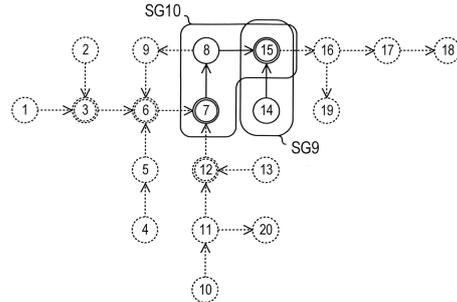


FIG. 18

【 図 1 7 】

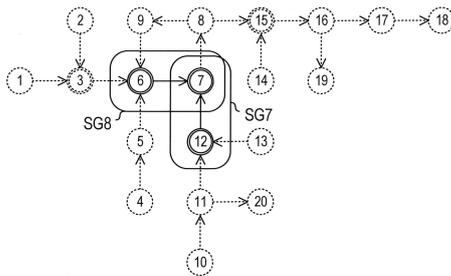


FIG. 17

【図19】

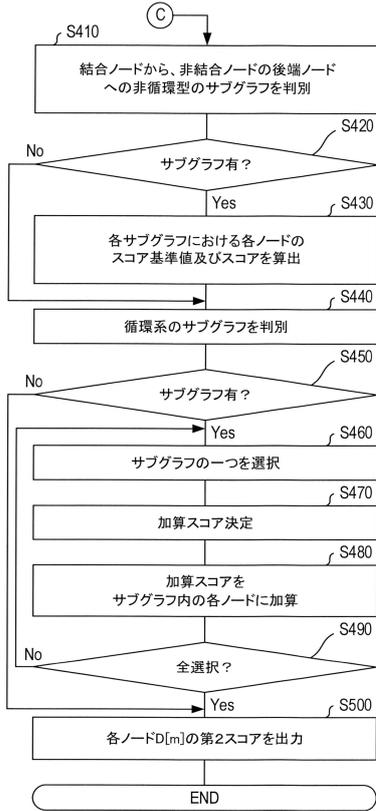


FIG. 19

【図20】

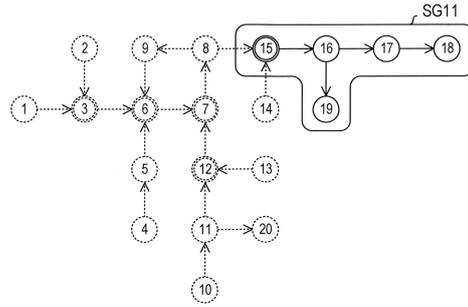


FIG. 20

【図21】

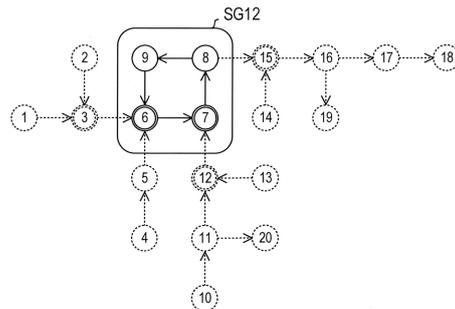


FIG. 21

【図22】

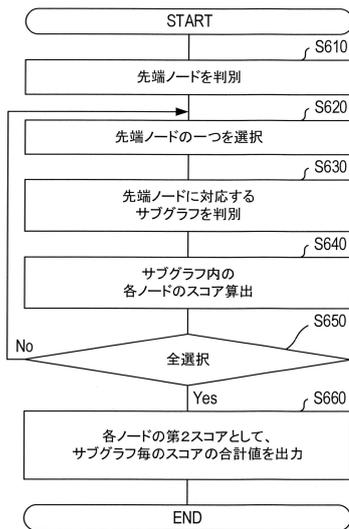


FIG. 22

【図24】

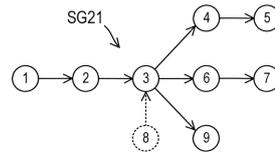


FIG. 24A

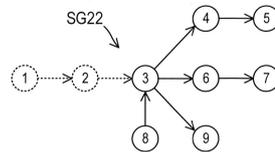


FIG. 24B

【図23】

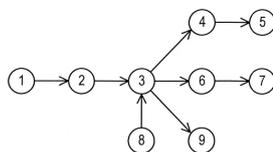


FIG. 23

【図 25】

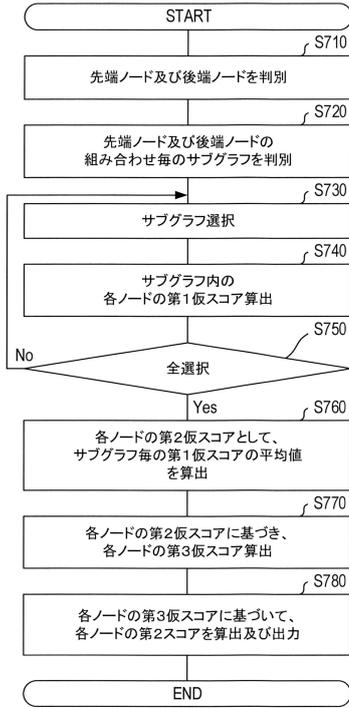


FIG. 25

【図 26】

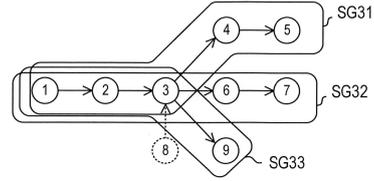


FIG. 26A

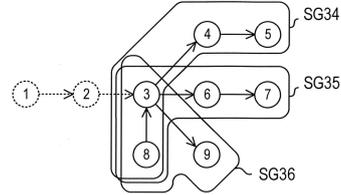


FIG. 26B

【図 27】

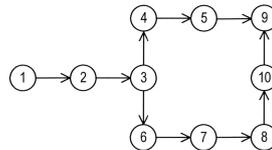


FIG. 27

【図 28】

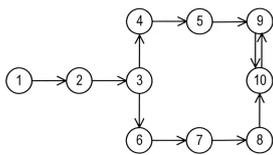


FIG. 28

【図 29】

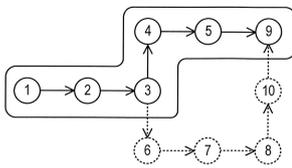


FIG. 29A

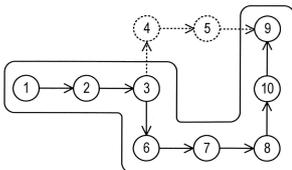


FIG. 29B

【図 30】

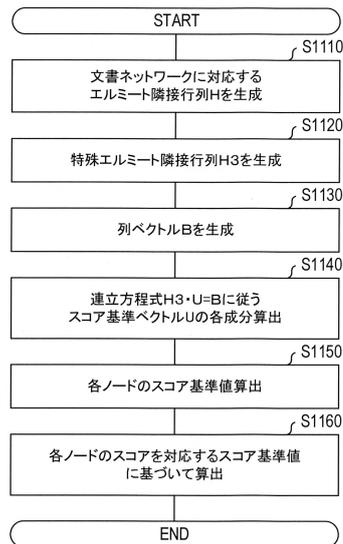


FIG. 30

【 図 3 1 】

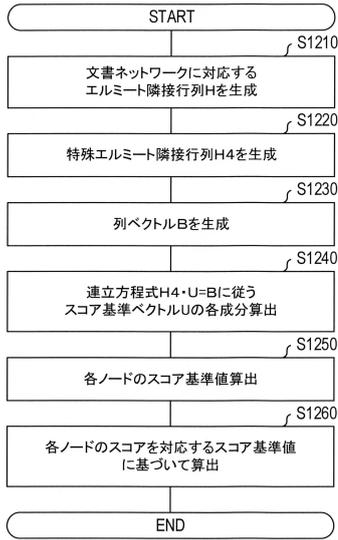


FIG. 31

【 図 3 2 】

$$H = \begin{bmatrix} 0 & +i & 0 & 0 & 0 \\ -i & 0 & +i & +i & 0 \\ 0 & -i & 0 & 0 & 0 \\ 0 & -i & 0 & 0 & +i \\ 0 & 0 & 0 & -i & 0 \end{bmatrix}$$

↓

$$H^{(1)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ C1(C2-i) & 0 & 0 & 0 & 0 \\ 0 & C1(C2-i) & 0 & 0 & 0 \\ 0 & C1(C2-i) & 0 & 0 & 0 \\ 0 & 0 & 0 & C1(C2-i) & 0 \end{bmatrix}$$

FIG. 32

【 図 3 3 】

$$H^{(2)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ C1(C2-i) & 0 & 0 & 0 & 0 \\ 0 & C1(C2-i)/2 & 0 & 0 & 0 \\ 0 & C1(C2-i)/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & C1(C2-i) & 0 \end{bmatrix}$$

↓

$$H4 = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ C1(C2-i) & -1 & 0 & 0 & 0 \\ 0 & C1(C2-i)/2 & -1 & 0 & 0 \\ 0 & C1(C2-i)/2 & 0 & -1 & 0 \\ 0 & 0 & 0 & C1(C2-i) & -1 \end{bmatrix}$$

FIG. 33

フロントページの続き

特許法第30条第2項適用 集会 第12回計算科学情報技術国際学会 発表日 令和1年12月19日
特許法第30条第2項適用 刊行物 ジャーナルオブ ソフトウェア 14巻2号第58-64頁 2019年
発行日 平成31年2月25日
特許法第30条第2項適用 刊行物 グローバル ジャーナル オブ コンピュータサイエンス アンド テク
ノロジー 19巻3号バージョン1.0 2019年 発行日令和1年10月14日
特許法第30条第2項適用 集会 東北社会学会 2019年(第66回) 年次大会 発表日 令和 1年7
月15日
特許法第30条第2項適用 ウェブサイトの掲載日 令和2年7月8日 ウェブサイトのアドレス <http://www.pfeffer.at/sunbelt/talks/285.html>
特許法第30条第2項適用 集会 2020年 サンベルト バーチャル会議 発表日 令和2年7月15日
特許法第30条第2項適用 集会 第2回コンピュータ通信とインターネットに関する国際会議 発表日 令和
2年6月28日
特許法第30条第2項適用 刊行物 南山大学紀要『アカデミア』理工学編、第20巻、1-20頁、2020
年3月 発行日 令和2年3月27日
特許法第30条第2項適用 集会 第11回複雑・情報学・サイバーネティクス国際マルチ会議 発表日 令和
2年3月13日
特許法第30条第2項適用 刊行物 電子情報通信学会2020年総合大会講演論文集、情報・システム1、第
3頁、2020年3月3日、 発行日 令和2年3月3日
特許法第30条第2項適用 集会 第17回ウェブ・グラフのためのアルゴリズムとモデルのワークショップ
発表日 令和2年9月23日

早期審査対象出願

(56)参考文献 国際公開第2019/106878(WO, A1)
特表2007-511815(JP, A)
米国特許第6285999(US, B1)

(58)調査した分野(Int.Cl., DB名)

G06F16/00-16/958