

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6058554号
(P6058554)

(45) 発行日 平成29年1月11日(2017.1.11)

(24) 登録日 平成28年12月16日(2016.12.16)

(51) Int. Cl.	F I
G06F 17/30 (2006.01)	G06F 17/30 210A
	G06F 17/30 220Z
	G06F 17/30 170A

請求項の数 13 (全 20 頁)

(21) 出願番号	特願2013-548574 (P2013-548574)	(73) 特許権者	508109737
(86) (22) 出願日	平成24年1月6日(2012.1.6)		アイエックスリビール インコーポレイテッド
(65) 公表番号	特表2014-502766 (P2014-502766A)		アメリカ合衆国, フロリダ州 32216
(43) 公表日	平成26年2月3日(2014.2.3)		, ジャクソンビル, ユニバーシティー プールバード サウス 3100, スイート 240
(86) 国際出願番号	PCT/US2012/020478	(74) 代理人	100079108
(87) 国際公開番号	W02012/094592		弁理士 稲葉 良幸
(87) 国際公開日	平成24年7月12日(2012.7.12)	(74) 代理人	100109346
審査請求日	平成27年1月5日(2015.1.5)		弁理士 大貫 敏史
(31) 優先権主張番号	61/430, 919	(72) 発明者	モーハン, レンガスワミー
(32) 優先日	平成23年1月7日(2011.1.7)		アメリカ合衆国, フロリダ州 32256
(33) 優先権主張国	米国 (US)		, ジャクソンヴィル, マウント レーニア ドライブ 7837

最終頁に続く

(54) 【発明の名称】 概念及びリンク発見システム

(57) 【特許請求の範囲】

【請求項1】

第1の概念を示す第1の選択をユーザから受信することであって、前記第1の概念が非構造化データ・オブジェクト内のテキスト・ストリングの有無又は構造化データ・オブジェクトに記憶されたデータ・コードによって定義されることと、

第2の概念を示す第2の選択をユーザから受信することであって、前記第2の概念が非構造化データ・オブジェクト内のテキスト・ストリングの有無又は構造化データ・オブジェクトに記憶されたデータ・コードによって定義されることと、

前記第1の概念と前記第2の概念との関係を決定することであって、前記関係が、複数の文書のうち、前記第1の概念及び前記第2の概念を含む文書の数に基づくものであり、前記関係が前記複数の文書のうちの第1の文書と前記複数の文書のうちの第2の文書とに基づくものであり、前記第1の文書が、前記第1の概念からの第1の非構造化データ・オブジェクト内の第1のテキスト・ストリングと、前記第2の概念からの第1の非構造化データ・オブジェクト内の第1のテキスト・ストリングとを含み、前記第2の文書が、前記第1の概念からの第2の非構造化データ・オブジェクト内の第2のテキスト・ストリングと、前記第2の概念からの第2の非構造化データ・オブジェクト内の第2のテキスト・ストリングとを含む、決定することと、

前記関係の可視表現をディスプレイに出力することとをシステムが行う、方法。

【請求項2】

前記関係が、前記複数の文書のうち、前記第2の概念から所定の近接性の範囲内に前記第1の概念を含む文書の数に基づくものである、請求項1記載の方法。

【請求項3】

前記近接性が同じパラグラフ内である、請求項2記載の方法。

【請求項4】

前記第1の概念からの前記第1の非構造化データ・オブジェクト内の前記第1のテキスト・ストリングが、前記第1の概念からの前記第2の非構造化データ・オブジェクト内の前記第2のテキスト・ストリングとは異なる、請求項1記載の方法。

【請求項5】

第3の概念を示す第3の選択をユーザから受信することであって、前記第3の概念が非構造化データ・オブジェクト内のテキスト・ストリングの有無又は構造化データ・オブジェクトに記憶されたデータ・コードによって定義されることと、

前記第1の概念と前記第3の概念との関係を決定することであって、前記関係が、前記複数の文書のうち、前記第1の概念及び前記第3の概念を含む文書の数に基づくものであることと、

前記第1の概念と前記第2の概念と前記第3の概念との関係を決定することであって、前記関係が、前記複数の文書のうち、前記第1の概念、前記第2の概念、及び前記第3の概念を含む文書の数に基づくものであることと、

前記第1の概念と前記第3の概念との前記関係の可視表現をディスプレイに出力することと、

前記第1の概念と前記第2の概念と前記第3の概念との前記関係の可視表現をディスプレイに出力すること

を更に前記システムが行う、請求項1記載の方法。

【請求項6】

前記第1の概念と前記第2の概念と前記第3の概念との前記関係が、前記複数の文書のうち、所定の近接性の範囲内に前記第1の概念、前記第2の概念、及び前記第3の概念を含む文書の数に基づくものである、請求項5記載の方法。

【請求項7】

複数の概念からの第1の概念の選択を示す少なくとも1つのユーザ入力を受信することであって、前記第1の概念が非構造化データ・オブジェクト内のテキスト・ストリングの有無又は構造化データ・オブジェクトに記憶されたデータ・コードによって定義されることと、

前記複数の概念からの第2の概念の選択を示す前記少なくとも1つのユーザ入力であって、前記第2の概念が非構造化データ・オブジェクト内のテキスト・ストリングの有無又は構造化データ・オブジェクトに記憶されたデータ・コードによって定義されることと、

前記複数の概念からの第3の概念の選択を示す前記少なくとも1つのユーザ入力であって、前記第3の概念が非構造化データ・オブジェクト内のテキスト・ストリングの有無又は構造化データ・オブジェクトに記憶されたデータ・コードによって定義されることと、

前記第1の概念と前記第2の概念と前記第3の概念とのマルチリンク関係を決定することであって、前記マルチリンク関係が、(1)前記第1の概念と前記第2の概念との関係の強度、並びに(2)前記第1の概念と前記第2の概念と前記第3の概念との関係の強度を示し、前記マルチリンク関係が複数の文書のうちの第1の文書と前記複数の文書のうちの第2の文書とに基づくものであり、前記第1の文書が、前記第1の概念からの第1の非構造化データ・オブジェクト内の第1のテキスト・ストリングと、前記第2の概念からの第1の非構造化データ・オブジェクト内の第1のテキスト・ストリングとを含み、前記第2の文書が、前記第1の概念からの第2の非構造化データ・オブジェクト内の第2のテキスト・ストリングと、前記第2の概念からの第2の非構造化データ・オブジェクト内の第2のテキスト・ストリングとを含む、決定することと、

前記マルチリンク関係を示す可視表現を表示することをシステムが行う、方法。

10

20

30

40

50

【請求項 8】

前記第 1 の概念と前記第 2 の概念との前記関係の前記強度が、複数の文書のうち、前記第 1 の概念と前記第 2 の概念の両方を含む文書の数を示す数値である、請求項 7 記載の方法。

【請求項 9】

前記第 1 の概念と前記第 2 の概念との前記関係の前記強度が、複数の文書のうち、前記第 1 の概念と前記第 2 の概念の両方を含む文書の数を示す重みを有する線として表示される、請求項 7 記載の方法。

【請求項 10】

前記第 1 の概念と前記第 2 の概念と前記第 3 の概念との前記関係の前記強度が、複数の文書のうち、前記第 1 の概念、前記第 2 の概念、及び前記第 3 の概念を含む文書の数を示すカラーの色調として表示される、請求項 7 記載の方法。

10

【請求項 11】

前記少なくとも 1 つのユーザ入力、前記第 1 の概念の前記選択を示す第 1 のユーザ入力と、前記第 2 の概念の前記選択を示す第 2 のユーザ入力と、前記第 3 の概念の前記選択を示す第 3 のユーザ入力である、請求項 7 記載の方法。

【請求項 12】

前記複数の概念に第 4 の概念を含み、前記第 4 の概念が非構造化データ・オブジェクト内の前記第 1 の概念、前記第 2 の概念、及び前記第 3 の概念の存在又は構造化データ・オブジェクトに記憶されたデータ・コードによって定義されることを更に含む、請求項 7 記載の方法。

20

【請求項 13】

前記第 1 の概念と前記第 2 の概念と前記第 3 の概念との前記関係の前記強度が、複数の文書のうち、所定の近接性の範囲内に前記第 1 の概念、前記第 2 の概念、及び前記第 3 の概念を含む文書の数に基づくものである、請求項 7 記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

[1001] 本出願は、2011年1月7日出願され、「Concepts and Link Discovery」という名称の米国仮特許出願第 61 / 430, 919 号に対する優先権及び利益を請求するものであり、同出願は参照により全体として本明細書に組み込まれる。

30

【背景技術】

【0002】

[1002] 諜報及び犯罪防止組織は、典型的に、分析技法とグラフィカル技法の組み合わせを使用して、エンティティ（例えば、人物、場所、及び物）間のリンク（即ち、関連性）を発見し表現する。このような技法は、一般に、1つ又は複数のデータセットを分析し、そこに発現するか又は述べられている様々なエンティティ間のリンクを発見する。これらの技法は「リンク解析（link analysis）」と呼ばれる場合が多い。

【0003】

[1003] このプロセス中に、発見されたリンクは、そのデータに固有の様々なエンティティ関係の理解を容易にするためにグラフィカル形式で表現することができる。例えば、エンティティ間で発見されたリンクは、一続きの線で接続された2つ以上のノードを含む線図又は図解を介して表現することができ、それぞれのノードは1つのエンティティを表し、それぞれのノード間の線はこれらの2つのノード間のリンクの存在を示している。このようなリンク解析は、典型的に、事前構造化したデータ・テーブル内に編成された共通属性に基づいて、人物、場所、及び物をリンクすることに限定される。このような構造化データ及び/又はフィールドは、リレーショナル・データベース（例えば、Oracle、IBM DB 2、Microsoft SQL Server、MySQL、又はPostgreSQLリレーショナル・データベース）など、編成された列、テーブル、スプレッドシート、又はその他のデータ構造で検出されるデータを含む。

40

50

【0004】

[1004] 現在容認されているリンク解析の方法の1つは、本質的にノード方式であり、リレーショナル・データベースのテーブルに含まれる構造化データ・フィールド間のリンクを確立することを伴うものである。この伝統的なリンク解析は、編成された構造化データについては比較的効率の良いものであり、上記のように視覚的に表現されると、ユーザについて「ドットを接続する」際に有用である可能性がある。

【0005】

[1005] しかし、この手法には多くの制限がある。第1に、伝統的なリンク解析によって定義されるリンク接続は、構造化データベースのルールに限定される。このようなルールは、分析されたデータが多面的であるか、非構造化であるか、及び/又は適正なリンクのために文脈要因を必要とする時に、非効率的である場合が多い。また、現在の解決策は、伝統的なリンク解析ツール及びシステムによって処理可能な構造化フォーマットにデータを再加工することを伴う場合も多い。換言すれば、このような解決策は、解決策をデータに適合させるのではなく、データを解決策に適合させようとしてリソースを消費する機会が多い。更に、現在の手法は、一般に、2つのエンティティのそれぞれがもう一方のエンティティのものとは一致する構造化フィールドであるか又はそのような構造化フィールドを含む時に、2つのエンティティ間にリンクが存在すると判断する。従って、これらの手法は、一般に、概念上の近接性又は文脈を説明するものではない。実際に、このような手法は、一般に、非構造化（又は「汚い（dirty）」）データ・フィールドを適切に説明するか及び/又は分析することができない。このように非構造化データを分析できないことにより、結果的にリンク定義が不正確になるだけでなく、一般に、組織の非構造化データが完全に分析されないままになる。ある組織にとって使用可能なデータの大部分が典型的に本質的に非構造化である場合、上記の欠点は現在のデータ解析技法の有効性を著しく制限する。

【0006】

[1006] 例えば、所与の警察の報告書は、データベース（又はその他のデータ構造）に記憶された任意の構造化フィールド内に含まれないテキスト（即ち、非構造化）情報内に述べられている2つ以上のエンティティ間の明確なリンクを掲載しているか又は含む可能性がある。このデータが構造化形式で編成又は提示された場合、人物と住所の両方が同じ構造化レコード又はフィールド内に発現する予定であれば、警察の報告書に述べられている人物と住所との間でリンクが発見されるだけであろう。しかし、多くの状況では、2つ以上のエンティティ間のリンクは、1つの文書又は複数文書のコーパス内の単なる同時発現以外の何かに基づいて存在する可能性がある。例えば、2つ以上のエンティティ間のリンクは、非構造化データ内の近接性に基づくか、又はそのデータのその他のプロパティ又は内容（単語、語句、文脈、或いはその他の言語学的又はコード上の特徴など）に基づく可能性がある。例えば、海軍犯罪調査サービス（NCIS）は、「ブルー（blue）」又は「オールド（Old）」（衣料品店「オールド・ネイビー（Old Navy）」による）などのその他の頻繁に相関する用語に極めて接近して発現しない「ネイビー（navy）」という単語/用語への参照を発見しようと努める可能性がある。この例では、NCISは、「ネイビー」という用語が分析データ内に明示的に発現しない場合でも、軍艦或いはその他の関連エンティティ又は概念への参照も発見したいと希望する可能性がある。伝統的なリンク解析を使用すると、この問題に対する典型的な解決策は、リンク解析システムが無視するはずの潜在的に曖昧な用語（例えば、上記のもの）の項目を含む、「ネイビー」という用語を明らかにするためのユーザ入力を必要とするであろう。しかし、非構造化データから構造を生成するためのこのような努力は、一般に効果的ではない。

【0007】

[1007] 従って、エンティティ間のリンクと、拳動、領域、出来事、活動などのより抽象的な構成体（即ち、概念）間のリンクの両方を発見する必要がある。実世界の例として、伝統的なリンク解析は、人物（例えば、下着姿の爆弾犯人アブドゥル・ムッタラブ）と、1つ又は複数の電話番号と、その人物が訪問した国と、その人物が加入していた組織（例

10

20

30

40

50

えば、イエメン内のアルカイダ・グループ)とのリンクを発見することができる。このような挙動は典型的に構造化テキストと非構造化テキストの組み合わせ(即ち、人間が作った単語又は散文)で記述されるので、典型的なリンク発見技法はそれらの間のリンクを正確かつ十分に識別することができない。例えば、海軍関係の警察逮捕に関する分析は、構造化データベース・フィールドを頼りにしており、典型的に、逮捕者の「就職先」などの列を頼りにすると思われ、この場合、これは海軍に勤務する逮捕者について海軍を示すことになるであろう。しかし、この分析に関するより深くかつ正確な諜報は、逮捕報告書の叙述情報から非構造化データを融解することによって達成されるであろう。これは、逮捕者が海軍との強い結びつきを持っているという事実にもかかわらず、構造化フィールド内の「就職先」として海軍を指定しないと思われる海軍関係の逮捕の例が数多くあるからである。例としては、少数の例を挙げれば、配偶者の逮捕、任意の数の企業のうちの1社を介して海軍に勤務する契約作業員の逮捕、その施設、船舶、又は所有物など、海軍に明確に関連する無数の引用のうちのいずれかを含む。更に、リンクは多面的なものになる可能性があり、複数の概念レベルで、並びに様々なデータ・タイプ、データ・システム、及び/又はリポジトリにわたって発現する可能性があるので、伝統的なリンク解析ではこのようなリンクを正確かつ十分に識別することができない。

10

【発明の概要】

【発明が解決しようとする課題】

【0008】

[1008] 従って、非構造化データ内に存在する概念間のリンクを発見するように構成された方法及び装置が必要である。更に、定義済みの特性及び/又は挙動の組み合わせに適合する、人物及び/又は組織間のリンクを柔軟かつ効率的に発見するように構成された方法及び装置が必要である。

20

【課題を解決するための手段】

【0009】

[1009] いくつかの実施形態の方法は、第1の概念を示す第1の選択をユーザから受信することによって、第1の概念が非構造化データ・オブジェクト内のテキスト・ストリングの有無又は構造化データ・オブジェクトに記憶されたデータ・コードによって定義されることを含む。この方法は、第2の概念を示す第2の選択をユーザから受信することによって、第2の概念が非構造化データ・オブジェクト内のテキスト・ストリングの有無又は構造化データ・オブジェクトに記憶されたデータ・コードによって定義されることを更に含む。この方法は、第1の概念と第2の概念との関係を決定することによって、その関係が、複数の文書のうち、第1の概念及び第2の概念を含む文書の数に基づくものであることを更に含む。この方法は、その関係の可視表現をディスプレイに出力することを更に含む。

30

【図面の簡単な説明】

【0010】

【図1】[1010]一実施形態による概念及びリンク発見システム (concept and link discovery system) を示す概略図である。

【図2】[1011]一実施形態によるホスト・デバイス (host device) のプロセッサを示す概略図である。

40

【図3】[1012]一実施形態による概念及びリンク発見システムを操作する方法を示すフローチャートである。

【図4A】[1013]一実施形態による概念及びリンク発見システムのグラフィカル・ユーザ・インターフェース (GUI) を示す図である。

【図4B】[1014]一実施形態による概念及びリンク発見システムのグラフィカル・ユーザ・インターフェース (GUI) を示す図である。

【図5A】[1015]伝統的なリンク解析手法によって定義されたリンク図である。

【図5B】[1016]一実施形態による概念ベースのリンク発見手法 (concept-based link discovery approach) によって定義されたリンク図である。

50

【図 6】[1017]一実施形態による概念ベースのリンク発見モジュールによって定義されたリンク図の一部である。

【図 7】[1018]一実施形態による概念及びリンク発見システムを操作する方法を示すフローチャートである。

【図 8 A】[1019]一実施形態により少なくとも 1 つの概念リンク (concept link) を含むデータセットに基づく概念ベースのリンク図である。

【図 8 B】[1020]一実施形態により少なくとも 1 つの複数概念リンク (multi-concept link) を含むデータセットに基づく概念ベースのリンク図である。

【図 8 C】[1021]一実施形態により少なくとも 1 つの複数概念リンクと少なくとも 1 つのその他の概念と少なくとも 1 つの新しい概念とを含むデータセットに基づく概念ベースのリンク図である。

10

【図 8 D】[1022]一実施形態により少なくとも 1 つの複数概念リンクと少なくとも 1 つのその他の概念と少なくとも 1 つの新しい概念とを含むデータセットに基づく概念ベースのリンク図である。

【図 8 E】[1023]一実施形態により少なくとも 1 つの複数概念リンクと少なくとも 1 つの新しい概念とを含むデータセットに基づく概念ベースのリンク図である。

【発明を実施するための形態】

【0011】

[1024] 本明細書で使用する「概念」とは、任意の実世界の観測結果の表現及び/又は考え方又は意味を伝達する 1 つ又は複数の単語又は語句の集合を指す。また、概念は、1 つ又は複数の事業ニーズ、考え方、挙動、多面的なエンティティの集合、或いはこれらの任意の組み合わせであるか及び/又はそれを含むことができる。いくつかの実施形態では、少なくとも部分的に機械学習技法の組み合わせ及び/又はユーザ入力に基づいて概念を定義することができる。概念、概念定義、及び概念発見に関する詳細は、「Concept-based Method and System for Dynamically Analyzing Unstructured Information」という名称の米国特許第 6 9 7 0 8 8 1 号及び「Method, System, and Computer Program Product for Concept-based Multi-dimensional Analysis of Unstructured Information」という名称の米国特許第 7 1 9 4 4 8 3 号に明記されており、いずれも参照により全体として本明細書に組み込まれる。

20

【0012】

[1025] また、概念は、構造化データ (コード及び数など) 及び/又は非構造化データ (人間に優しいテキストなど) も含むことができる。いくつかの実施形態では、機械又はユーザは、階層的に及び/又は正規表現の一部として又は両方の組み合わせとして、少なくとも部分的にその他の概念に基づいて 1 つ又は複数の概念を定義することができる。階層概念及び少なくとも部分的に 1 つ又は複数の正規表現に基づいて定義された概念に関する詳細については、2009 年 4 月 14 日に出願され、「Concept-based Analysis of Structured and Unstructured Data Using Concept Inheritance」という名称の同時係属米国特許出願第 12 / 4 2 3 0 2 4 号に明記されている。いくつかの実施形態では、概念は、任意選択で、様々なレベルの細分性で構造化及び非構造化データを含むことができ、それによりビジネス・ルールによって要求された通りに動的かつ連続的にデータを混合する能力を提供する。

30

40

【0013】

[1026] 概念はテキスト (典型的に、非構造化データ) で発現することができるので、非構造化データ内に存在する概念間のリンクを発見するために同時発現、近接性、及び言語学的技法を使用するように 1 つ又は複数のモジュールを構成することができる。より具体的には、このようなモジュールは、a) 同じ文書内の 2 つ以上の概念の同時発現、b) 1 つ又は複数の文書内のユーザ定義の近接性の範囲内の 2 つ以上の概念の同時発現、及び/又は c) 自然言語部分内に存在する主語 - 述語、主語 - 目的語、又は述語 - 目的語の関係の認識に基づいて、2 つ以上の概念間のリンクを発見するか及び/又は定義することができる。このように、1 つ又は複数のモジュールは、そこに存在する概念に基づいて文書又

50

はレコードを分析することができ、従って、伝統的なリンク解析技法に代わる動的な代案を提供することができる。

【 0 0 1 4 】

[1027] いくつかの実施形態では、1つ又は複数のモジュールは、1つ又は複数のデータセット、データ・ソース、又はレコードを受信し、それについて概念ベースのリンク発見を実行するように構成された1つ又は複数のハードウェア及び/又はソフトウェア・モジュール(ハードウェア内で実行される)にすることができる。例えば、これらのモジュールは、コンピュータ化テキスト及び/又はデータベース情報を参照することができる、計算装置(compute device)、ホスト・デバイス、及び/又は計算装置とホスト・デバイスを含むシステムに含めるか及び/又はそこで実行することができる。計算装置は、ネットワーク(例えば、ローカル・エリア・ネットワーク(LAN)、広域ネットワーク(WAN)、又はインターネット)、取り外し可能記憶媒体(例えば、光ディスク、フラッシュ・メモリ・ドライブなど)、又は固定記憶媒体(例えば、ハード・ディスク・ドライブ又はソリッド・ステート・ドライブ(SSD))を介してコンピュータ化テキスト及び/又はデータベース情報を受信するか及び/又はそれにアクセスすることができる。次に、1つ又は複数のモジュールは、コンピュータ化テキスト及び/又はデータベース情報に含まれる1つ又は複数の概念を発見するか及び/又は定義することができる。受信したデータ内に存在する1つ又は複数の概念を定義すると、1つ又は複数のモジュールは次に、発見された概念のうちの2つ以上の概念間に存在する1つ又は複数の概念ベースのリンクを発見し定義することができる。いくつかの実施形態では、この1つ又は複数のモジュールは次に、発見された概念及び概念ベースのリンクをメモリに記憶するか及び/又は任意選択でユーザ消費のために可視形式で(モニター又はその他の出力装置上に表示された図を介するなど)概念及び/又は概念リンクを出力することができる。

【 0 0 1 5 】

[1028] 図1は、一実施形態により、ネットワーク160を介してホスト・デバイス120と通信状態にある計算装置110を示す概略図である。ネットワーク160は、有線ネットワーク及び/又は無線ネットワークとして実現された任意のタイプのネットワーク(例えば、ローカル・エリア・ネットワーク(LAN)、広域ネットワーク(WAN)、仮想ネットワーク、電気通信ネットワーク)にすることができる。本明細書に更に詳細に記載されているように、いくつかの実施形態では、例えば、計算装置110は、インターネット・サービス・プロバイダ(ISP)及びインターネット(例えば、ネットワーク160)を介してホスト・デバイス120に接続されたパーソナル・コンピュータである。

【 0 0 1 6 】

[1029] いくつかの実施形態では、計算装置110は、中間ネットワーク及び/又は代替ネットワークを介してホスト・デバイス120及びネットワーク160と通信することができる。このような中間ネットワーク及び/又は代替ネットワークは、ネットワーク160と同じタイプ及び/又は異なるタイプのものにすることができる。このため、いくつかの実施形態では、計算装置110は、共通ネットワークを使用してホスト・デバイス120に送信される場合もあれば送信されない場合もある複数の通信モード(例えば、電子メール、テキスト・メッセージ、インスタント・メッセージ、光学パターン伝送、モバイル・デバイス・アプリケーションの使用、Webサイト経由、パーソナル・コンピュータ(PC)アプリケーションの使用、及び/又はTCP/IP伝送など)を使用して、ホスト・デバイス120にデータを送信するか及び/又はホスト・デバイス120からデータを受信することができる。

【 0 0 1 7 】

[1030] ホスト・デバイス120は、ネットワーク160により計算装置110にデータを送信するか及び/又は計算装置110からデータを受信するように構成することができる。いくつかの実施形態では、ホスト・デバイス120は、例えば、サーバ・デバイス(例えば、Webサーバ・デバイス)、ネットワーク管理装置、データ・リポジトリ、及び/又は同様の物として機能するように構成される。

10

20

30

40

50

【 0 0 1 8 】

[1031] ホスト・デバイス 1 2 0 はメモリ 1 2 4 とプロセッサ 1 2 2 とを含む。メモリ 1 2 4 は、例えば、ランダム・アクセス・メモリ (R A M)、メモリ・バッファ、ハード・ドライブ、データベース、消去可能プログラム可能読み取り専用メモリ (E P R O M)、電気消去可能読み取り専用メモリ (E E P R O M)、読み取り専用メモリ (R O M)、及び / 又は同様の物にすることができる。いくつかの実施形態では、ホスト・デバイス 1 2 0 のメモリ 1 2 4 は、1 つ又は複数の概念に関連するデータセット 1 4 0 を更新するために使用されるデータを含む。このような実施形態では、例えば、ホスト・デバイス 1 2 0 は、1 つ又は複数の通信モードを使用して計算装置 1 1 0 から受信した信号に基づいてデータセット 1 4 0 を追加、除去、改訂、及び / 又は編集するように構成される。いくつかの実施形態では、メモリ 1 2 4 は、このようなユニバーサル・リスト・システム及び / 又はサービスに関連するモジュール、プロセス、及び / 又は機能をプロセッサに実行させるための命令を記憶する。

10

【 0 0 1 9 】

[1032] ホスト・デバイス 1 2 0 のプロセッサ 1 2 2 は、概念及びリンク発見システム 1 0 0 を実行させるか及び / 又は実行するように構成された任意の適切な処理装置にすることができる。例えば、プロセッサ 1 2 2 は、本明細書に更に詳細に記載されているように、計算装置 1 1 0 から信号を受信したことに応答して、データセット 1 4 0 を更新するように構成することができる。いくつかの実施形態では、プロセッサ 1 2 2 は、汎用プロセッサ、書き換え可能ゲート・アレイ (F P G A)、特定用途向け集積回路 (A S I C)、デジタル信号プロセッサ (D S P)、及び / 又は同様の物にすることができる。

20

【 0 0 2 0 】

[1033] ホスト・デバイス 1 2 0 は動作可能にデータセット 1 4 0 に結合される。データセット 1 4 0 は、例えば、R A M、R O M、ハード・ディスク・ドライブ、光ドライブ、又はその他の取り外し可能媒体などのコンピュータ化メモリに常駐することができる。データセット 1 4 0 の構造化データ・ソース (図示せず) は、例えば、構造化照会言語 (S Q L) データベース、1 つ又は複数のコマンド区切り形式 (C S V) ファイル、1 つ又は複数のその他のパターンで区切られたファイル、或いはその他の構造化データ・フォーマット階層などのリレーショナル・データベースに編成することができる。データセット 1 4 0 の非構造化データ・オブジェクト (図示せず) は、例えば、手書き文書、タイプライターで打った文書、電子的なワードプロセッサ文書、印刷されたか又は電子的なスプレッドシート文書、印刷された用紙又はチャート、或いは電子メール、A d o b e P D F 文書、M i c r o s o f t O f f i c e 文書などのテキストを含むその他の電子文書のうちの 1 つ又は複数にすることができる。いくつかの実施形態では、構造化データ・ソースは、例えば、s t r i n g 又は v a r c h a r 型のリレーショナル・データベース列として記憶されたテキストのストリングなど、1 つ又は複数の非構造化データ要素を含むことができる。データセット 1 4 0 は、階層概念 (例えば、少なくとも 1 つ又は複数のその他の概念によって定義された概念) を含むユーザ及び / 又は機械生成概念を含むことができる。

30

【 0 0 2 1 】

[1034] 計算装置 1 1 0 は、例えば、計算エンティティ (例えば、デスクトップ・コンピュータ、ラップトップ・コンピュータなどの個人用計算装置)、携帯電話、モニター装置、携帯情報端末 (P D A)、及び / 又は同様の物にすることができる。図示されていないが、いくつかの実施形態では、計算装置 1 1 0 は、計算装置 1 1 0 をネットワーク 1 6 0 に接続するように構成された 1 つ又は複数のネットワーク・インターフェース装置 (例えば、ネットワーク・インターフェース・カード) を含むことができる。

40

【 0 0 2 2 】

[1035] 図 1 に示されているように、計算装置 1 1 0 は、プロセッサ 1 1 2 と、メモリ 1 1 4 と、ディスプレイ 1 1 4 とを有する。メモリ 1 1 4 は、例えば、ランダム・アクセス・メモリ (R A M)、メモリ・バッファ、ハード・ドライブ、及び / 又は同様の物にすることができる。ディスプレイ 1 1 6 は、例えば、液晶ディスプレイ (L C D)、陰極線管

50

ディスプレイ（ＣＲＴ）又は同様の物などの任意の適切なディスプレイにすることができる。その他の実施形態では、計算装置１１０は、ディスプレイ１１６の代わりに又はそれに加えて他の出力装置を含むことができる。例えば、計算装置１１０は、オーディオ出力装置（例えば、スピーカ）、触覚出力装置、及び／又は同様の物を含むことができる。

【００２３】

[1036] いくつかの実施形態では、ホスト・デバイス１２０の１つ又は複数の部分及び／又は計算装置１１０の１つ又は複数の部分は、ハードウェア・ベースのモジュール（例えば、デジタル信号プロセッサ（DSP）、書き換え可能ゲート・アレイ（FPGA））及び／又はソフトウェア・ベースのモジュール（例えば、メモリに記憶されたか及び／又はプロセッサで実行されたコンピュータ・コードのモジュール）を含むことができる。いくつかの実施形態では、ホスト・デバイス１２０に関連する機能のうちの１つ又は複数（例えば、プロセッサ１２２に関連する機能）は、１つ又は複数のモジュールに含めることができる（例えば、図２を参照）。いくつかの実施形態では、計算装置１１０に関連する機能のうちの１つ又は複数（例えば、プロセッサ１１２に関連する機能）は、１つ又は複数のモジュールに含めることができる。いくつかの実施形態では、計算装置１１０はホスト・デバイス１２０に関連する１つ又は複数の機能を実行するように構成することができ、逆もまた同様である。

【００２４】

[1037] 図２は、他の実施形態により概念及びリンク発見システムをホストとして処理するホスト・デバイスのプロセッサ２００を示す概略図である。いくつかの実施形態では、プロセッサ２００はホスト・デバイス１２０のプロセッサ１２２と同様の物にすることができる。より具体的には、プロセッサ２００は、複数の通信モードを使用して、データセットを更新するか及び／又は概念ベースのリンク解析を実行するように構成された任意の適切な処理装置にすることができる。

【００２５】

[1038] プロセッサ２００は、ユーザ入力モジュール２０２と、リンク発見モジュール２０４と、表示モジュール２０６とを含む。それぞれのモジュールは他のすべてのモジュールと直接通信状態にあるものとして図２に示されているが、他の実施形態では、それぞれのモジュールが他のすべてのモジュールと直接通信状態にある必要はない。その上、他の実施形態では、任意の他の数のモジュールをプロセッサ２００内に含めることができる。

【００２６】

[1039] ユーザ入力モジュール２０２は、ユーザ選択、新しい文書及び／又はデータセット、並びにその他の入力を計算装置（例えば、計算装置１１０）から受信するように構成される。具体的には、ユーザ入力モジュール２０２は、１つ又は複数の概念を示すユーザ選択を示す信号と、１つ又は複数の選択された概念間のリンク（関係）及び／又はそのリンクの強度を決定するための関連信号を受信するように構成される。いくつかの実施形態では、ユーザ入力装置は、例えば、データセット又はデータセットの一部内の１つ又は複数の概念間の近接性など、リンクに関連するパラメータを受信するように構成することができる。

【００２７】

[1040] リンク発見モジュール２０４は、本明細書に更に詳細に記載されているように、ユーザ入力モジュール２０２を介してユーザ入力を受信し、ユーザ入力に基づいて、１つ又は複数の選択された概念間のリンクを発見するように構成される。いくつかの実施形態では、リンク発見モジュール２０４は、例えば、データセット又はデータセットの一部内の１つ又は複数の選択された概念間の近接性など、ユーザ入力モジュール２０２を介して受信したパラメータに基づいて、リンクを発見するように構成することができる。いくつかの実施形態では、リンク発見モジュール２０４は、発見されたリンク及び／又は発見されたリンクの強度を示す信号を表示モジュール２０６に出力するように構成することができる。いくつかの実施形態では、リンク発見モジュールは、本明細書に更に詳細に記載されているように、発見されたリンクに基づいて新しい概念を定義するように構成するこ

10

20

30

40

50

とができ、新しい概念をデータセットに追加するか及び/又は他のモジュールに追加させるように構成することができる。

【0028】

[1041] 表示モジュール206は、発見されたリンク及び/又は発見されたリンクの強度を示す信号を受信するように構成される。表示モジュール206は、発見されたリンク及び/又は発見されたリンクの強度の可視表現をディスプレイに出力させるための信号を送信するように更に構成される。いくつかの実施形態では、表示モジュール206は、本明細書に更に詳細に記載されているように、データセットからの文書又はその他のデータ、新しい概念、その他の概念、及び/又は選択された概念、新しい概念(複数も可)、及び/又はその他の概念間のリンクをディスプレイに出力させるように構成することができる。

10

【0029】

[1042] 図3は、概念及びリンク発見のための方法300を描写するフローチャートである。方法300は、302で第1の概念を示す第1の選択をユーザから受信し、第1の概念が非構造化データ・オブジェクト内のテキスト・ストリングの有無又は構造化データ・オブジェクトに記憶されたデータ・コードによって定義されることを含む。一例として、図2に関連して説明すると、リンク発見モジュール204は、ユーザ入力モジュール202を介してユーザから第1の選択を受信することができる。方法300は、304で第2の概念を示す第2の選択をユーザから受信し、第2の概念が非構造化データ・オブジェクト内のテキスト・ストリングの有無又は構造化データ・オブジェクトに記憶されたデータ・コードによって定義されることを含む。引き続き図2に関連して説明すると、リンク発見モジュール204は、ユーザ入力モジュール202を介してユーザから第2の選択を受信することができる。方法300は、306で第1の概念と第2の概念との関係を決定し、その関係が、複数の文書のうち、第1の概念及び第2の概念を含む文書の数に基づくものであることを含む。引き続き図2に関連して説明すると、リンク発見モジュール204は、1つ又は複数のデータセット又はデータセットの一部分を分析して、第1の概念及び第2の概念を含む文書がある場合にその数を決定することができる。方法300は、308でその関係の可視表現をディスプレイに出力することを含む。引き続き図2に関連して説明すると、表示モジュール206は、リンク発見モジュール204によって発見された1つ又は複数のリンクに基づいて、可視表現を定義し、ディスプレイに出力することができる。

20

30

【0030】

[1043] 図4A及び図4Bは一実施形態による概念及びリンク発見システムのグラフィカル・ユーザ・インターフェース(GUI)を描写し、図5Aは従来技術のシステムのリンク発見出力の可視表現を描写し、図5Bは一実施形態によるリンク発見出力の可視表現を描写している。図4A~図5Bに関連して説明すると、1つの事例では、ユーザは、所与のエリア又は領域内のどの暴力団が現在衝突していて、どの暴力団が共謀しているかを複数のデータ・ソースから判断しようと努めることができる。伝統的なリンク解析では、一般に、以下のステップを含むことになるであろう。

【0031】

[1044] (1)複数のフィールドを含むデータ構造の作成とそれに続くその索引付け、並びに(2)暴力団に関する既知の情報に基づくリンクの捕捉及び発見。例えば、主要暴力団名構造化フィールドを識別して、構造化データ・フィールド間の接続を発見するプロセスを開始することができる。次にこの主要暴力団名フィールドは、関連データ(電話番号、入れ墨の記述、落書きの記述、ナンバープレートなど)によりエンティティを接続する能力をもたらし、ユーザは発見されたリンクを手動で調査して、検討中の2つの暴力団同士の関係の性質を決定するか及び/又は定義できるようになるであろう。

40

【0032】

[1045] 対照的に、いくつかの実施形態では、複数のデータ・ソース(上記の通り)について概念ベースのリンク発見を実行するように構成された概念及びリンク発見システムの

50

1つ又は複数のモジュールを使用して、同じ問題例を解決することができる。伝統的なリンク解析とは対照的に、このプロセスは、構造化データの生成又は定義、構造化フィールドの索引付け、データ浄化、或いは検討中の暴力団に関する重大な予備知識を全く必要としない可能性がある。例えば、単に「暴力団」という用語並びに調査すべき1つ又は複数の挙動関連概念（例えば、「不平」、「確執」、「連携」、「忠誠」など）を使用して概念を定義することにより、リンク発見モジュールなどの1つ又は複数のモジュールは、概念間の1つ又は複数のリンクを発見することができる。これらの概念及び概念ベースのリンクを発見し定義すると、1つ又は複数のモジュールは、ユーザ・インターフェース（UI）を提供して、ユーザが1つ又は複数の暴力団間に存在する共通リンク及び/又は関係に関連する照会を実行し、発見された概念及び/又は概念ベースのリンクに基づいて応答を受信できるように構成することができる。

10

【0033】

[1046] この例では、複数のデータ・ソースに含まれる非構造化データ内で「doo doo creek boys」という名前の暴力団を参照することができる。図4Aに示されているように、この暴力団名の複数の変形を含む、この暴力団に関する概念を定義することにより、1つ又は複数のモジュールは、複数のデータ・ソース内のこの暴力団の発現、実例、及び/又は言及の多く又はすべてを検出することができる。このように、このモジュール（複数も可）は、伝統的なリンク解析を使用して（即ち、構造化データ・フィールドのみを使用して）捕捉/発見されないと思われる、発現を捕捉し、リンクを発見することができる。図4Aに示されているように、「doo doo creek」という概念について検索すると、その結果、20個の文書がヒットする（doo doo creekについて15個、creek boysについて5個、d d cについて8個、doodooについて1個）。対照的に、図4Bに示されているように、この例では、伝統的なリンク解析により「doo doo creek」という正確な語句が発現するいくつかの文書のみが得られる。換言すれば、伝統的なリンク解析では、「doo doo creek」という語句を含む15個の文書を発見できるだけであるが、「doo doo creek」に関する概念の定義に含まれるその他の語句を含む14個の追加文書を見落とす可能性がある。他の例では、リンク発見システムのユーザは、「doo doo creek boys」と「eastside boyz」とのリンクを発見しようと試みる可能性がある。伝統的なリンク解析では、「doo doo creek」と「eastside boyz」という両方の語句を含むリンクについて検索できるだけであるので、「creek boys」と「eastside boys」という語句の両方を含む既存の文書

20

30

【0034】

[1047] 図5Aは、伝統的なリンク解析手法によって定義されたリンクの可視表現を示している。図5Aに示されているように、（伝統的なリンク解析のように）「Doo Doo Creek Boys」に関連する各データ又は概念が別個の構造化フィールドに含まれていた場合、「D D C」、「Creek Boyz」、「Creek Boys」、「DoDo」、「DooDoo」など、この暴力団名の変形に関する構造化フィールドは全く存在しないであろう。このような手法は、恐らく、暴力団のメンバーをDoo Doo Creekという暴力団に結びつける電話番号をリンクする関係表（リレーショナル・データベース・テーブルなど）を含むであろう。例えば、「John Doe」という個人は632-784-3972という電話番号に結びつけることができ、次にこの番号は「Doo Doo Creek Boys」に結びつけることができ、次にJohn Doeによって行われた強盗も「Doo Doo Creek Boys」に結びつけることができるであろう。

40

【0035】

[1048] 対照的に、図5Bは、一実施形態による概念ベースのリンク発見手法によって定義されたリンク図の可視表現を示している。概念ベースのリンク解析により、1つ又は複

50

数のモジュールは、上記で列挙した暴力団名の変形及び／又はその他の名前の変形を含む、「Doo Doo Creek Boys」という概念を定義することができる。いくつかの実施形態では、それぞれの名前の変形は、上記の電話番号例に関する下位概念（sub-concept）と、「John Doe」という名前に関する電話番号の下位概念から外れた下位概念を含むことができる。このように、1つ又は複数のモジュールは、複数の概念及び概念ベースのリンクからなる階層グループ又はバンク内で、図5Bに描写されている概念ベースのリンクを定義することができる。

【0036】

[1049] いくつかの実施形態では、リンク発見モジュールなどの1つ又は複数のモジュールは、「半構造化（semi-structured）」フィールドのデータセット内に存在する概念及び／又は概念ベースのリンクを発見するか及び／又は定義することができる。半構造化データは、1つ又は複数の拡張可能マークアップ言語（XML）規格により定義されたデータ、1つ又は複数の形式文書及び／又はスプレッドシートに含まれるデータなど、非叙述的であって完全に構造化されているわけではないデータにすることができる。多数の構造化データ・フィールドが豊富な情報を含んでいるにもかかわらず、典型的なリンク解析技法は、データセット例で言及されている企業A（「Structure Tone Company」）と企業B（「Constructors and Associates」）との間のリンクを生成又は発見することができない。この例では、企業Aは1人又は複数人の従業員、場所などに関連付けることができ、企業Aのそれぞれの従業員は1つ又は複数の名前、電子メール・アドレス、電話番号など（様々な方法で、構造化データと非構造化データの両方に入力される場合が多い）に更に
10
20
関連付けることができる。このため、図6は、一実施形態及びこの例によりリンク発見モジュールによって定義されたリンク図の一部を示している。いくつかの実施形態では、リンク発見モジュールは、企業A及び／又は企業Bに関連する構造化、非構造化、及び／又は半構造化データを受信し、それに応じて企業Aの従業員に関する概念を定義することができる（例えば、図6に描写されている通り）。

【0037】

[1050] 同じく図6に示されているように、このデータセット例では、第1の文書内の企業Aに対する言及及び第2の文書内の企業Aの従業員（「Robin Malacrea」）に対する言及により、リンクが存在する。このデータセット例では、異なる企業向けの連絡窓口（POC）としてその従業員に言及する、第3の文書内のその従業員に対する第2の言及により、第2のリンクが存在する。
30

【0038】

[1051] 同じく図6に示されているように、企業Aに関連するアドレスに関する概念は、1つ又は複数のその他の企業に関する1つ又は複数の概念にリンクされる。より具体的には、図6は、「Structure_Tone_Company_address」と他の2つの企業（「Vfinity_Company」及び「Nielsen_Company」）との間のリンクを（黒い線により）示している。図6に表されているように、このデータセット例では、このアドレスをVfinity_Companyに関連付ける1つの文書と、このアドレスをNielsen_Companyに関連付ける6つの文書が存在する。
40

【0039】

[1052] 図7は、一実施形態による概念及びリンク発見方法に関する方法700を示すフローチャートである。具体的には、方法700は一実施形態によるマルチリンク発見方法を示している。方法700は、702で複数の概念からの第1の概念の選択を示す少なくとも1つのユーザ入力を受信し、第1の概念が非構造化データ・オブジェクト内のテキスト・ストリングの有無又は構造化データ・オブジェクトに記憶されたデータ・コードによって定義されることを含む。方法700は、704で複数の概念からの第2の概念の選択を示す少なくとも1つのユーザ入力であり、第2の概念が非構造化データ・オブジェクト内のテキスト・ストリングの有無又は構造化データ・オブジェクトに記憶されたデータ・コードによって定義されることを含む。方法700は、706で複数の概念からの第3の概念の選択を示す少なくとも1つのユーザ入力であり、第3の概念が非構造化データ・オ
50

プロジェクト内のテキスト・ストリングの有無又は構造化データ・オブジェクトに記憶されたデータ・コードによって定義されることを含む。方法700は、708で第1の概念と第2の概念と第3の概念とのマルチリンク関係を決定し、そのマルチリンク関係が、(1)第1の概念と第2の概念との関係の強度、並びに(2)第1の概念と第2の概念と第3の概念との関係の強度を示すことを含む。方法700は、710でそのマルチリンク関係を示す可視表現を表示することを含む。

【0040】

[1053] 図8A～図8Dは発見されたリンクの可視表現を描写している。具体的には、図8Aは3つの選択された概念間の単一リンク関係を描写し、図8Bは図8Aに追加された3つの選択された概念間のマルチリンク関係を描写し、図8Cは図8Bに追加された2つの選択された概念間のリンクに関連する2組の追加の概念(その他の概念と学習された新しい概念)を描写し、図8Dは図8Cに追加された2組の追加の概念間のリンクを描写している。

10

【0041】

[1054] 図2及び図8A～図8Bに関連して説明すると、一例では、概念及びリンク発見システムのユーザは、3つの概念、具体的には「暴走族(Outlaw Biker Gang)」、「違法取引(Illegal Trafficking)」、及び「ナイトクラブ(Night Club)」の間の関係を発見することができる。この例では、法執行機関は、ナイトクラブに関わっていた暴走族の未知のメンバーを含む違法取引犯罪を調査している可能性がある。ユーザは計算装置を使用して、暴走族、違法取引、及びナイトクラブを選択することができる。ユーザは、例えば、グラフィカル・ユーザ・インターフェース内のラジオボタン、チェックボックス、グラブボックスなどを使用して選択を行うことができる。ユーザ入力モジュール202は計算装置から選択を受信することができる。リンク発見モジュール204は、ユーザ入力モジュール202から選択を受信ことができ、データセット及び/又はデータセットの一部を分析して、第1の概念と第2の概念との間、第1の概念と第3の概念との間、並びに第2の概念と第3の概念との間にリンクが存在するかどうかを判断することができる(例えば、図8Aを参照)。

20

【0042】

[1055] 単一文書に関しては、第1の概念に関連するテキスト・ストリングがその文書内に存在し、第2の概念に関連するテキスト・ストリングがその文書内に存在する時に、第1の概念が第2の概念とリンクされるか又は第2の概念との関係を有する。いくつかの実施形態では、複数の文書のうち、第1の概念及び第2の概念を含む文書の数は、2つの概念間のリンクの強度を示すことができる。図8Aに関連して説明すると、暴走族とナイトクラブは両方の概念を含む28個の文書を有し(最強リンク)、暴走族と違法取引は両方の概念を含む19個の文書を有し(中間リンク)、違法取引とナイトクラブは両方の概念を含む14個の文書を有する(最弱リンク)。1つの概念は複数の別個のテキスト・ストリングに関連付けることができるので、第1の概念と第2の概念は、別個のテキスト・ストリングを含む複数のリンクされた文書を含むことができる。一例として、「暴走族」という概念は「eastside boy\$ or Fairfax boy\$」という正規表現によって表すことができ、「違法取引」という概念は「drug trafficking or human trafficking」という正規表現によって表すことができる。イーストサイド・ボーイ及び麻薬取引を含む第1の文書は暴走族と違法取引とのリンクとして発見されることになり、フェアファックス・ボーイ及び人身売買を含む第2の文書も暴走族と違法取引とのリンクとして発見されるであろう。

30

40

【0043】

[1056] リンク発見モジュール204は、ユーザ入力モジュール202から選択を受信ことができ、データセット及び/又はデータセットの一部を分析して、第1の概念と第2の概念と第3の概念との間にリンクが存在するかどうかを判断することができる(例えば、図8Bを参照)。単一文書に関しては、第1の概念に関連するテキスト・ストリング、第2の概念に関連するテキスト・ストリング、並びに第3の概念に関連するテキスト・ストリングがその文書内に存在する時に、第1の概念と第2の概念と第3の概念がリン

50

クされるか又は関係を有する。図 8 B に関連して説明すると、3つの文書が第1の概念、第2の概念、及び第3の概念を含む。この関係は概念間の単一集中ノードによって表される。

【0044】

[1057] 表示モジュール206は、リンク発見モジュール204から信号を受信することができ、発見された単一リンク及びマルチリンクの可視表現を定義することができる。表示モジュール206は、計算装置のディスプレイにこの可視表現を表示させるためにユーザの計算装置に信号を送信することができる。リンクの可視表現及び/又はリンクの強度は、例えば、関係概念間の線を含むことができる。この線は、関係の絶対強度(例えば、文書の数に基づいて増分的に太くなる)又は関係の相対強度(例えば、最強関係は最も太くなるか特定のパターンを有する)を示すための重み及び/又はパターンを含むことができる。いくつかの実施形態では、ある線の重み又はパターンは概念又は文書に関する様々なデータ又はメタデータに基づくものにもできる。例えば、任意の所与の概念が接続文書を有するヒット数をカウントする代わりに、可視化では、線の太さを決定するためのアルゴリズムでヒット数(それらが接続しているという事実を超えるもの)を使用することができるであろう。他の例では、文書ではなく概念に関するデータが線の太さを決定できるであろう。その例では、いくつかの概念が重みを有することができ、これらの重み(スコア)はいくらか「より」重要な物にするものである(線の太さを決定する式においてより重い重みが付けられる)。発見されたリンクの可視表現では、概念又は文書に関するデータ又はメタデータ、或いは関係又はそれらが関係する理由の近接性、若しくはこれらの組み合わせを使用することができる。

10

20

【0045】

[1058] いくつかの実施形態では、2つ又はそれ以上の概念がリンクされると、ユーザは、概念及びリンク発見システムが、a)すでに概念バンク内に存在するか及び/又はb)機械学習を必要とする文書内のその他の概念を発見できるようにすることにより、その関係に含まれる任意の文書を調べることができる。GUIでは、ユーザは、単一又はマルチリンク関係の可視表現を操作(例えば、「ダブルクリック」など)して、新たに追加された概念に接続される線を示すことができる。図8Cに示されているように、データセット内の28個の文書が2つの概念を接続するものとして示されている場合、ユーザはこれらの文書内でその他の概念を「発見」できるであろう。他の言い方をすると、リンク発見モジュールは、その文書が記憶している任意の数の概念も含むと判断するか、又は機械学習を使用してこのように実行することにより以前は未知であった概念について学習することができる。従って、反復的に、ユーザは可視化においてこれらの新たに追加された概念と任意のその他の概念との間で、文書とのその他の接続を無期限に調べ続けることができる。他の言い方をすると、新たに発見された概念のうちの1つを接続できるその他の文書は、図8Dに示されているように、個別に又は同じ可視化におけるその他の概念間で文書に接続できるであろう。

30

【0046】

[1059] 図8C及び図8Dは2つの概念間の単一リンクからの新しい概念及びその他の概念を描写しているが、新しい概念及びその他の概念はマルチリンクに基づいて発見することもできる。図8Eに示されているように、このような実施形態では、概念及びリンク発見システムからの出力は新しい概念を含むことができる。概念及びリンク発見システムは、2つ以上の概念が所定のしきい値以上の強度の関係、一例として、2つ以上の共通文書を有する時に、新しい概念を作成するようユーザに指示するか及び/又はシステムが自動的に新しい概念を定義して記憶するように構成することができる。一例として、図8Eに関連して説明すると、3つの文書が暴走族とナイトクラブと違法取引をリンクするので、システムはこれらの3つの概念を含む新しい概念を定義するようユーザに指示することができる。この例では、新しい概念は、「暴走族取引活動(Biker Gang Trafficking Operations)」と呼ぶことができ、「Outlaw Biker Gang and Nightclubs and Illegal trafficking」という正規表現によって定義することができる。このように、ユーザは、「暴走

40

50

族取引活動」などの単一概念を後で選択することができ、3つの文書を発見することになるであろう。

【0047】

[1060] 本明細書に記載されているいくつかの実施形態は、コンピュータによって実現される様々な動作を実行するための命令又はコンピュータ・コードをその上に有する非一時的コンピュータ可読媒体（非一時的プロセッサ可読媒体と呼ぶこともできる）を備えたコンピュータ記憶製品に関する。コンピュータ可読媒体（又はプロセッサ可読媒体）は、本質的に一時的な伝搬信号（例えば、空間又はケーブルなどの伝送媒体上で情報を運ぶ伝搬電磁波）を含まないという意味で非一時的なものである。この媒体及びコンピュータ・コード（コードと呼ぶこともできる）は特定の1つ又は複数の目的のために設計され構築されたものに行うことができる。非一時的コンピュータ可読媒体の例としては、ハード・ディスク、フロッピー（登録商標）・ディスク、及び磁気テープなどの磁気記憶媒体、コンパクト・ディスク/デジタル・ビデオ・ディスク（CD/DVD）、コンパクト・ディスク読み取り専用メモリ（CD-ROM）、及びホログラフィ・デバイスなどの光記憶媒体、光ディスクなどの光磁気記憶媒体、搬送波信号処理モジュール、並びに特定用途向け集積回路（ASIC）、プログラマブル・ロジック・デバイス（PLD）、読み取り専用メモリ（ROM）、及びランダム・アクセス・メモリ（RAM）デバイスなど、プログラム・コードを記憶し実行するように特別に構成されたハードウェア・デバイスを含むが、これらに限定されない。

10

【0048】

[1061] コンピュータ・コードの例としては、マイクロコード又はマイクロ命令、コンパイラによって生成されたような機械命令、Webサービスを生成するために使用されるコード、及びインタプリタを使用してコンピュータによって実行される高レベル命令を含むファイルを含むが、これらに限定されない。例えば、諸実施形態は、Java（登録商標）、C++、又はその他のプログラミング言語（例えば、オブジェクト指向プログラミング言語）、並びに開発ツールを使用して実現することができる。コンピュータ・コードの追加の例としては、制御信号、暗号化コード、及び圧縮コードを含むが、これらに限定されない。

20

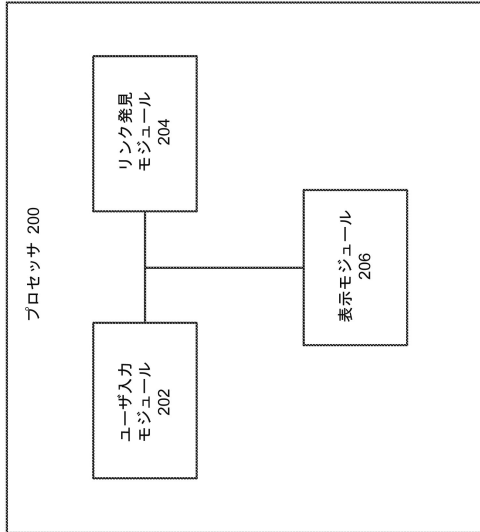
【0049】

[1062] 様々な実施形態について上記で説明してきたが、これらは限定としてではなく一例としてのみ提示されており、形式及び詳細において様々な変更が可能であることを理解されたい。本明細書に記載されている装置及び/又は方法のどの部分も、相互排他的な組み合わせを除き、任意の組み合わせで結合することができる。本明細書に記載されている諸実施形態は、記載されている異なる諸実施形態の機能、コンポーネント、及び/又は特徴の様々な組み合わせ及び/又は副次的組み合わせを含むことができる。例えば、いくつかの実施形態では、非一時的プロセッサ可読媒体は、複数の計算装置に常駐する複数のリレーショナル・データベース内に存在する概念及び/又は概念リンクを発見するように構成されたコードを記憶することができる。いくつかの実施形態では、1つのデータセットは1つの計算装置でローカルに記憶することができ、ホスト・デバイスに関連するプロセッサ、モジュール、及び/又は方法はその計算装置でローカルに含めるか及び/又は実行することができる。いくつかの実施形態では、概念及びリンク発見システムは共通データセットにアクセスする複数の計算装置を含むことができる。いくつかの実施形態では、ユーザはそれぞれの概念を個別に及び/又は実質的に同時に選択することができる。

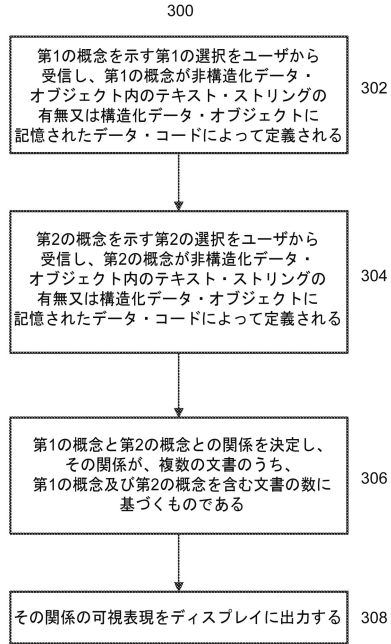
30

40

【図2】

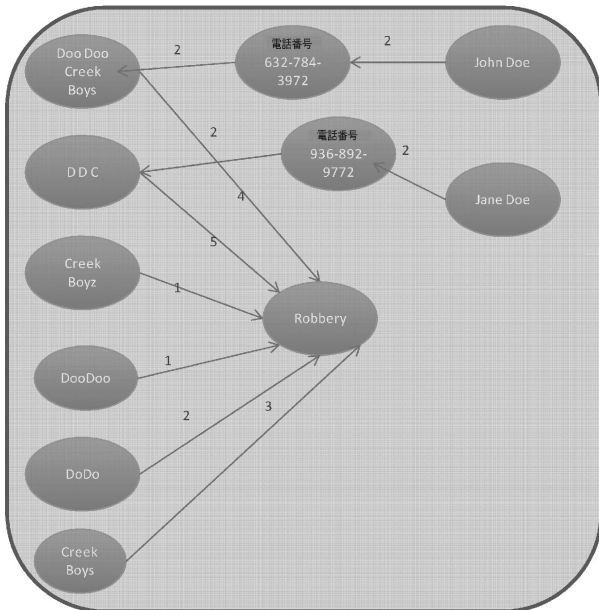


【図3】

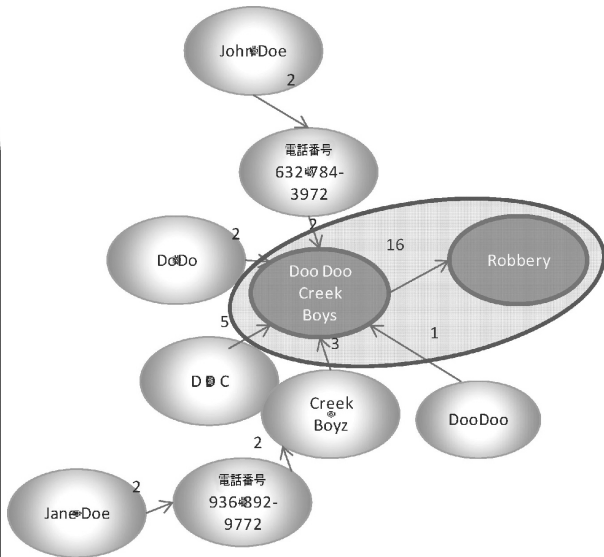


【図5A】

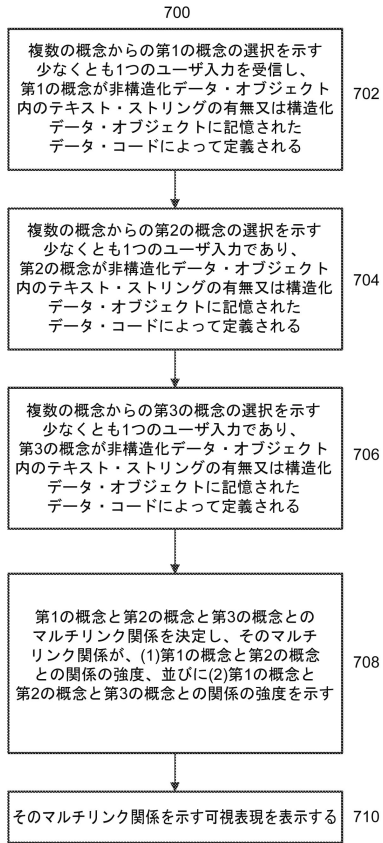
従来技術



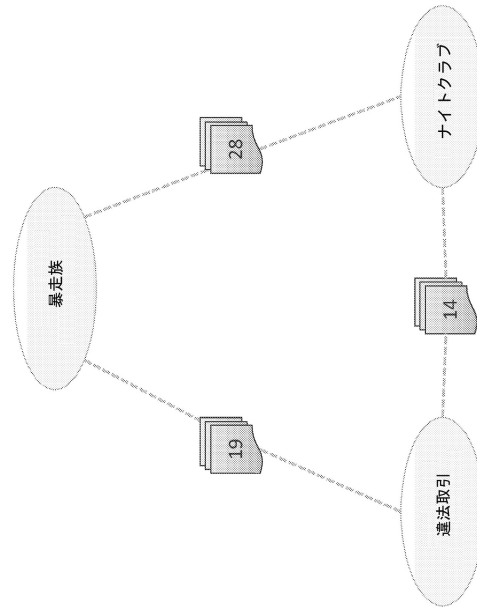
【図5B】



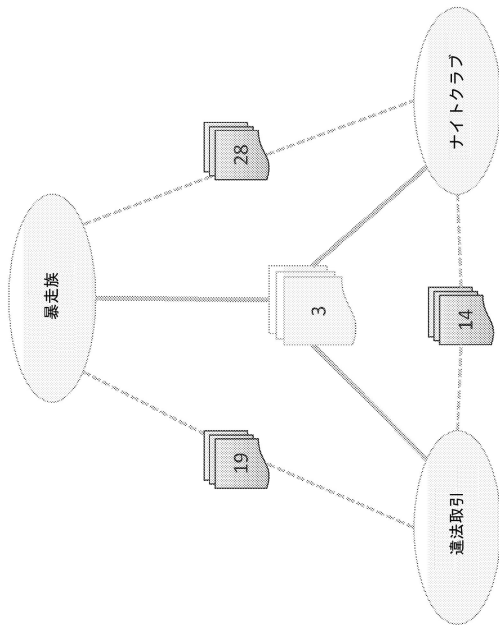
【図7】



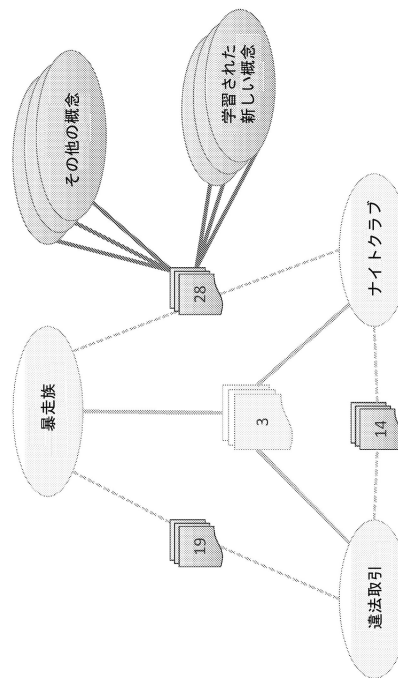
【図8A】



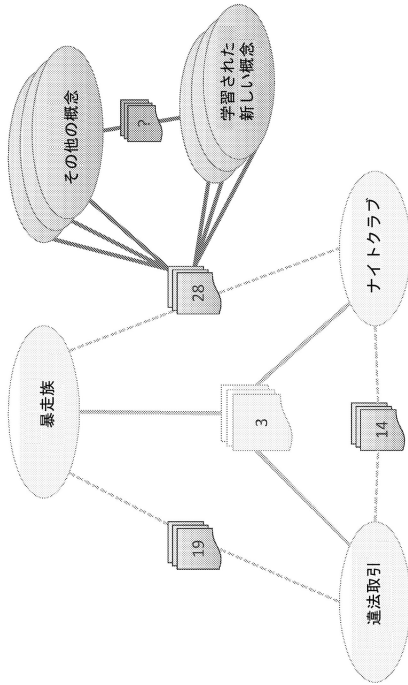
【図8B】



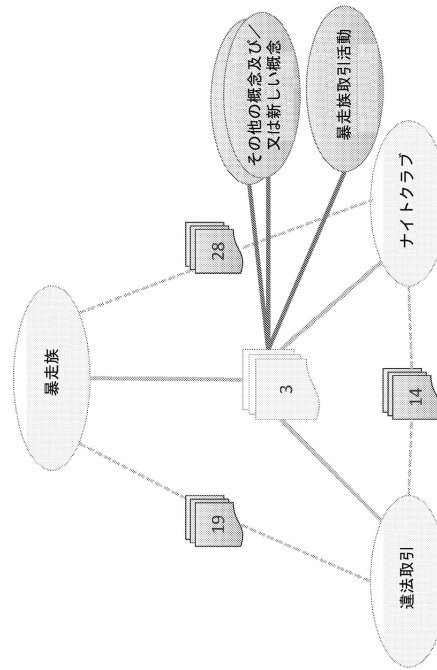
【図8C】



【 図 8 D 】



【 図 8 E 】



フロントページの続き

(72)発明者 ホホワイト, マシュー ブルース
アメリカ合衆国, フロリダ州 32250, ジャクソンヴィル, ペンマン ロード 603

審査官 吉田 誠

(56)参考文献 特表2010-529518(JP, A)
国際公開第2008/121377(WO, A1)
特開2003-263458(JP, A)

(58)調査した分野(Int.Cl., DB名)
G06F 17/30