

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5581339号  
(P5581339)

(45) 発行日 平成26年8月27日 (2014. 8. 27)

(24) 登録日 平成26年7月18日 (2014. 7. 18)

(51) Int. Cl. F 1  
**G 0 6 F 17/30 (2006.01)**  
 G 0 6 F 17/30 3 6 0 Z  
 G 0 6 F 17/30 2 1 0 A

請求項の数 16 (全 124 頁)

(21) 出願番号	特願2011-546411 (P2011-546411)	(73) 特許権者	507103802
(86) (22) 出願日	平成22年1月16日 (2010. 1. 16)		グーグル・インコーポレーテッド
(65) 公表番号	特表2012-515407 (P2012-515407A)		アメリカ合衆国・カリフォルニア・940
(43) 公表日	平成24年7月5日 (2012. 7. 5)		43・マウンテン・ビュー・アンフィシア
(86) 国際出願番号	PCT/US2010/021290		ター・パークウェイ・1600
(87) 国際公開番号	W02010/083478	(74) 代理人	100108453
(87) 国際公開日	平成22年7月22日 (2010. 7. 22)		弁理士 村山 靖彦
審査請求日	平成25年1月16日 (2013. 1. 16)	(74) 代理人	100064908
(31) 優先権主張番号	12/355, 103		弁理士 志賀 正武
(32) 優先日	平成21年1月16日 (2009. 1. 16)	(74) 代理人	100089037
(33) 優先権主張国	米国 (US)		弁理士 渡邊 隆
(31) 優先権主張番号	12/355, 459	(74) 代理人	100110364
(32) 優先日	平成21年1月16日 (2009. 1. 16)		弁理士 実広 信哉
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 非構造化電子文書コレクションからの情報の取り出しおよび表示

(57) 【特許請求の範囲】

【請求項 1】

1つまたは複数のデータ処理機にオペレーションを実行させるように動作可能な命令を格納する1つまたは複数の機械可読データ記憶媒体を備える装置であって、前記オペレーションは、

事前に存在している構造化プレゼンテーションを記述する記述データを受け取る手段と

、  
第1のウェブサイトから第1のインスタンスの識別子を引き出す手段と、

第2のウェブサイトから前記第1のインスタンスの第1の属性の第1の値を引き出す手段と

、  
第1のインスタンスの前記識別子および新規値を前記事前に存在している構造化プレゼンテーションに追加して新規構造化プレゼンテーション内に新規レコードを形成する手段と、

前記新規構造化プレゼンテーションを視覚的に提示するための命令を出力する手段とを具備し、

前記事前に存在している構造化プレゼンテーションの視覚的提示は構造化設計に適合するシステム配置で情報を視覚的に提示し、前記構造化プレゼンテーションはレコードのコレクションを含み、これらのそれぞれは前記構造化プレゼンテーションの視覚的提示でインスタンスと前記インスタンスの属性を特徴付ける値との間の関連付けを前記インスタンスの識別子と前記値の配列によって表し、

前記第1のウェブサイトから前記第1のインスタンスの前記識別子を引き出す手段は、前記事前に存在している構造化プレゼンテーションの特性を前記事前に存在している構造化プレゼンテーションのコンテンツと比較する手段を具備する装置。

【請求項2】

前記オペレーションは、ユーザーから第2のインスタンスの識別子を受け取る手段をさらに含み、

新規の前記構造化プレゼンテーションは、前記第2のインスタンスの前記第1の属性の第2の値に関連して前記第2のインスタンスを提示する第2の新規レコードを含む請求項1に記載の装置。

【請求項3】

前記オペレーションは、前記ユーザーから前記第2の値を受け取る手段をさらに具備する請求項2に記載の装置。

【請求項4】

前記オペレーションは、

前記ユーザーに候補値のコレクションを提示する手段と

前記ユーザーから前記第2の値の選択を受け取る手段とをさらに具備し、

前記コレクションは前記第2の値を含む請求項2に記載の装置。

【請求項5】

前記オペレーションは、

前記第2のインスタンスの前記第1の属性の候補値のコレクションを識別する手段と、

前記候補値のそれぞれについて、前記候補値が正しいという確かさを決定する手段とをさらに具備する請求項2に記載の装置。

【請求項6】

前記オペレーションは、前記構造化プレゼンテーションに追加する新規インスタンスのコレクションを提案する手段をさらに具備する請求項1に記載の装置。

【請求項7】

新規インスタンスの前記コレクションを提案する手段は、前記事前に存在している構造化プレゼンテーションの特性を前記第1のウェブサイトおよび前記第2のウェブサイトのコンテンツと比較する手段を具備する請求項6に記載の装置。

【請求項8】

新規インスタンスの前記コレクションを提案する手段は、機械可読検索クエリを前記第1のウェブサイトおよび前記第2のウェブサイトのコンテンツと比較する手段を具備する請求項6に記載の装置。

【請求項9】

前記第2のウェブサイトから前記第1の値を引き出す手段は、前記第2のウェブサイトがレビューを含むことを識別する手段を具備する請求項1に記載の装置。

【請求項10】

前記第1のウェブサイトから前記識別子を引き出す手段は、前記第1のウェブサイトから直接的に前記識別子を抽出する手段を具備する請求項1に記載の装置。

【請求項11】

前記第1のウェブサイトから前記識別子を引き出す手段は、前記第1のウェブサイトから抽出された情報を含む機械可読データベースから前記識別子を抽出する手段を具備する請求項1に記載の装置。

【請求項12】

前記事前に存在している構造化プレゼンテーションは、テーブルを含み、

前記レコードは、前記テーブルの行または列を含む請求項1に記載の装置。

【請求項13】

前記事前に存在している構造化プレゼンテーションは、カードのコレクションを含み、

前記レコードは、前記コレクション内に個別のカードを含む請求項1に記載の装置。

【請求項14】

前記第1のウェブサイトから前記識別子を引き出す手段は、前記第1のウェブサイトから直接的に前記識別子を抽出する手段を具備する請求項1に記載の装置。

10

20

30

40

50

前記オペレーションは、ディスプレイ画面の1つまたは複数の要素を物理的に変換する手段を含めて、前記新規構造化プレゼンテーションを前記ディスプレイ画面上に視覚的に提示する手段をさらに具備する請求項1に記載の装置。

【請求項15】

クライアントデバイスと、前記クライアントデバイスをインタラクティブに操作し、オペレーションを実行するようにプログラムされた1つまたは複数のコンピュータとを具備するシステムであって、

前記オペレーションは、

事前に存在している構造化プレゼンテーションを記述する記述データを受け取る手段と

、

第1のウェブサイトから第1のインスタンスの識別子を引き出す手段と、

第2のウェブサイトから前記第1のインスタンスの第1の属性の第1の値を引き出す手段と

、

第1のインスタンスの前記識別子および新規値を前記事前に存在している構造化プレゼンテーションに追加して新規構造化プレゼンテーション内に新規レコードを形成する手段と、

前記新規構造化プレゼンテーションを視覚的に提示するための命令を前記クライアントデバイスに出力する手段とを具備し、

前記事前に存在している構造化プレゼンテーションの視覚的提示は構造化設計に適合するシステム配置で情報を視覚的に提示し、前記構造化プレゼンテーションはレコードのコレクションを含み、これらのそれぞれは前記構造化プレゼンテーションの視覚的提示でインスタンスと前記インスタンスの属性を特徴付ける値との間の関連付けを前記インスタンスの識別子と前記値の配列によって表し、

前記第1のウェブサイトから前記第1のインスタンスの前記識別子を引き出す手段は、前記事前に存在している構造化プレゼンテーションの特性を前記事前に存在している構造化プレゼンテーションのコンテンツと比較する手段を具備するシステム。

【請求項16】

前記1つまたは複数のコンピュータは、データ通信ネットワークを通じて前記クライアントデバイスをインタラクティブに操作するように動作可能なサーバーを具備し、

前記クライアントデバイスは、クライアントとして前記サーバーをインタラクティブに操作するように動作可能である請求項15に記載のシステム。

【発明の詳細な説明】

【技術分野】

【0001】

本明細書は、非構造化電子文書コレクション(unstructured electronic document collection)から情報を取り出して表示する方法に関する。

【背景技術】

【0002】

電子文書は、機械可読データのコレクションである。電子文書は、一般的に個別のファイルであり、定義済みのフォーマット(例えば、PDF、TIFF、HTML、MS Word、PCL、PostScript、または同様のもの)に従ってフォーマットされる。電子文書は、電子的に格納され、電子的に配布されうる。いくつかの場合において、電子文書は、オーディオコンテンツ、ビジュアルコンテンツ、および他の情報、さらにはテキストおよび他の電子文書へのリンクを含む。

【0003】

電子文書は、電子文書コレクションにまとめることができる。電子文書コレクションは、構造化されていないか、または構造化されているかのいずれかとすることができる。非構造化電子文書コレクションに含まれる文書のフォーマットは、所定の構造に準じるように制約されず、多くの場合に予見できない形で発展しうる。言い換えると、非構造化電子文書コレクション内の個別の文書のフォーマットは、文書コレクション全体にわたって制

10

20

30

40

50

限的でも、または永続的でもない。さらに、非構造化電子文書コレクションでは、新規文書があるフォーマットに従うこと、またはフォーマットへの変更がすでに存在している文書に適用されることを確実にするためのメカニズムはない。したがって、非構造化電子文書コレクション内の文書は、情報の抽出の際に利用可能な共通構造を共有することを期待できない。非構造化電子文書コレクションの例として、インターネット上で利用可能な文書、レジユメのコレクション、雑誌記事のコレクション、およびニュース記事のコレクションが挙げられる。いくつかの非構造化電子文書コレクション内の文書は、そのコレクションの内部の他の文書および外部の他の文書へのリンクを含むことを禁じられる。

#### 【0004】

それとは対照的に、構造化電子文書コレクション内の文書は、一般的には、制限的であつ永続的とすることができるフォーマットに従う。構造化電子文書コレクション内の文書に適用されるフォーマットは、適用されるフォーマットが完全には適していない場合でも共通フォーマットがそれらのコレクション内の文書すべてに適用されるという点において制限的であつてよい。これらのフォーマットは、構造化電子文書コレクションを構築する当事者による特定のフォーマットへの顕著な関わりが一般的に要求されるという点で永続的であつてよい。さらに、これらのコレクションのユーザー（特に、コレクション内の文書を使用するプログラム）は、期待されるフォーマットを有する文書に依存する。その結果、フォーマットの変更は、実施しにくいことがある。構造化電子文書コレクションは、情報内容が単純で安定したカテゴリ化に役立つアプリケーションに最適である。したがって、構造化電子文書コレクション内の文書は、情報の抽出の際に利用可能な共通構造を一般的には共有する。構造化電子文書コレクションの例として、階層データモデルおよびリレーショナルデータモデルに従ってデータベース管理システム(DBMS)を通じて整理され、閲覧されるデータベース、さらには情報を矛盾なく提示するように単一のエンティティによって作成された電子文書のコレクションが挙げられる。例えば、個別の書籍に関する情報を提示するためにオンライン書店が提供するウェブページのコレクションは、構造化電子文書コレクションを形成することができる。他の例として、サーバーサイドスクリプトによって作成され、アプリケーションサーバーを通じて閲覧されるウェブページのコレクションは、構造化電子文書コレクションを形成することができる。そこで、1つまたは複数の構造化電子文書コレクションのそれぞれを、非構造化電子文書コレクションのサブセットとしてもよい。

#### 【発明の概要】

#### 【発明が解決しようとする課題】

#### 【0005】

本明細書では、非構造化電子文書コレクション、例えば、インターネット上で利用可能な電子文書から情報を取り出し、表示する段階に関係する技術について説明する。電子文書コレクションは非構造化コレクションであつてよいが、非構造化電子文書コレクションの情報内容は、構造化プレゼンテーション(structured presentation)で表示することができる。特に、非構造化電子文書コレクションの情報内容は、属性の値を決定するためだけではなく、構造化プレゼンテーションで属性およびインスタンスを識別し、選択し、名前を付けるためにも使用できる。このような構造化プレゼンテーションを使用すると、情報源の多様性にもかかわらず、情報を首尾一貫した様式でユーザーに提示することができる。構造化プレゼンテーションの例として、テーブル、およびレコードの他のコレクションが挙げられる。

#### 【課題を解決するための手段】

#### 【0006】

一般に、本明細書で説明される発明対象の一態様は、ユーザーから機械可読検索クエリを受け取るアクションと、検索クエリに関連するインスタンスの構造化プレゼンテーションをユーザーに提示するための命令で検索クエリに応答するアクションとを含む機械実装方法により具現化されうる。構造化プレゼンテーションの視覚的提示は、インスタンスとインスタンスの属性を特徴付ける値との間の関連付けを、インスタンスの識別子と値の配

10

20

30

40

50

列によって表す。インスタンスの識別子と値は、電子文書の非構造化コレクション内の2つまたはそれ以上の文書から引き出される。電子文書コレクションは、電子文書コレクション内の電子文書のフォーマットが制限的でも永続的でもないという点で非構造化コレクションである。

【0007】

この態様および他の態様は、以下の特徴のうちの1つまたは複数を含みうる。検索クエリに応答する段階は、インスタンスに関する非構造化コレクション内の電子文書の第1のコレクションを識別する段階と、電子文書の第1のコレクションからインスタンスの属性の値を抽出する段階と、構造化プレゼンテーションに2つまたはそれ以上の電子文書から抽出された値を書き込む段階とを含むことができる。検索クエリに応答する段階は、第1の電子文書から第1のインスタンスの第1の属性の第1の値を抽出する段階と、第2の電子文書から第1のインスタンスの第2の属性の第2の値を抽出する段階と、第1の値および第2の値を第1のインスタンスに構造化プレゼンテーションにおいて単一のものとして関連付ける段階とを含むことができる。第1の属性は、第2の属性と異なっていてよく、また第1の電子文書は、第2の電子文書と異なっていてよい。検索クエリに応答する段階は、第1の電子文書から第1のインスタンスの属性の第1の値を抽出する段階と、第1の電子文書から第2のインスタンスの属性の第2の値を抽出する段階と、第1の値を第1のレコード内の第1のインスタンスに関連付ける段階と、第2の値を第2のレコード内の第2のインスタンスに関連付ける段階とを含むことができる。第1のインスタンスは、第2のインスタンスと異なっていてよい。構造化プレゼンテーションは、テーブルを含み、レコードは、そのテーブルの行または列を含むことができる。構造化プレゼンテーションは、カードのコレクションを含み、レコードは、そのコレクション内の個別のカードであってよい。この方法は、新規インスタンスを構造化プレゼンテーションに追加するためのトリガーを受け取る段階、およびそのトリガーに反応して構造化プレゼンテーションに追加する新規インスタンスを提案する段階も含むことができる。この方法は、ユーザーから制約条件の指定を受け取る段階を含むこともでき、また新規インスタンスを提案する段階は、ユーザー指定制約条件を満たす新規インスタンスを提案する段階を含む。この方法は、新規属性を構造化プレゼンテーションに追加するためのトリガーを受け取る段階と、そのトリガーに反応して構造化プレゼンテーションに新規属性を追加する段階とを含むことができる。この方法は、新規属性の特質のユーザー指定を受け取る段階と、ユーザー指定の特質に基づいて構造化プレゼンテーションに属性の値を書き込む段階とを含むこともできる。非構造化電子文書コレクションは、インターネット上で利用可能な電子文書を含むことができる。構造化プレゼンテーションは、ディスプレイ画面の1つまたは複数の要素を物理的に変換する段階を含めて、ディスプレイ画面上に物理的に提示することができる。

【0008】

この態様の他の実施形態は、対応するシステム、装置、およびそれぞれこれらの方法のオペレーションを実行するように構成されているコンピュータ記憶デバイス上に記録されたコンピュータプログラムを含む。

【0009】

本明細書で説明されている発明対象の他の態様は、1つまたは複数のデータ処理機にオペレーションを実行させるように動作可能な命令を格納する1つまたは複数の機械可読データ記憶媒体を備える装置により具現化されうる。これらのオペレーションは、事前に存在している構造化プレゼンテーションを記述する記述データを受け取る段階と、第1のウェブサイトから第1のインスタンスの識別子を引き出す段階と、第2のウェブサイトから第1のインスタンスの第1の属性の第1の値を引き出す段階と、第1のインスタンスの識別子および新規値を事前に存在している構造化プレゼンテーションに追加して新規構造化プレゼンテーション内に新規レコードを形成する段階と、新規構造化プレゼンテーションを視覚的に提示するための命令を出力する段階とを含むことができる。事前に存在している構造化プレゼンテーションの視覚的提示は、構造化設計に適合するシステム配置で情報を視覚的に提示する。構造化プレゼンテーションは、構造化プレゼンテーションの視覚的提示で

10

20

30

40

50

インスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す。

【 0 0 1 0 】

この態様および他の態様は、以下の特徴のうちの1つまたは複数を含みうる。第1のウェブサイトから第1のインスタンスの識別子を引き出す段階は、事前に存在している構造化プレゼンテーションの特性を事前に存在している構造化プレゼンテーションのコンテンツと比較する段階を含むことができる。これらのオペレーションは、ユーザーから第2のインスタンスの識別子を受け取る段階を含むこともできる。新規構造化プレゼンテーションは、第2のインスタンスの第1の属性の第2の値に関連して第2のインスタンスを提示する第2の新規レコードを含むことができる。これらのオペレーションは、ユーザーから第2の値を受け取る段階を含むことができる。候補値のコレクションをユーザーに対して提示し、第2の値の選択をユーザーから受け取ることができる。候補値のコレクションは、第2の値を含む。第2のインスタンスの第1の属性の候補値のコレクションを識別することができる。候補値のそれぞれについて、候補値が正しい確かさを決定することができる。これらのオペレーションは、構造化プレゼンテーションに追加すべき新規インスタンスのコレクションを提案する段階を含むことができる。新規インスタンスのコレクションは、事前に存在している構造化プレゼンテーションの特性を第1のウェブサイトおよび第2のウェブサイトのコンテンツと比較することによって、および/または機械可読検索クエリを第1のウェブサイトおよび第2のウェブサイトのコンテンツと比較することによって提案されうる。第2のウェブサイトから第1の値を引き出す段階は、第2のウェブサイトがレビューを含むことを識別する段階、第1のウェブサイトから直接的に識別子を抽出する段階、または第1のウェブサイトから抽出された情報を含む機械可読データベースから識別子を抽出する段階を含むことができる。事前に存在している構造化プレゼンテーションは、テーブルを含み、レコードは、そのテーブルの行または列を含むことができる。事前に存在している構造化プレゼンテーションは、カードのコレクションを含み、レコードは、そのコレクション内の個別のカードであってよい。これらのオペレーションは、ディスプレイ画面の1つまたは複数の要素を物理的に変換する段階を含めて、新規構造化プレゼンテーションをディスプレイ画面上に視覚的に表示する段階を含むことができる。

【 0 0 1 1 】

この態様の他の実施形態は、対応するシステム、装置、および方法を含む。

【 0 0 1 2 】

他の態様では、システムは、クライアントデバイスおよびクライアントデバイスをインタラクティブに操作し、オペレーションを実行するようにプログラムされた1つまたは複数のコンピュータを備える。これらのオペレーションは、事前に存在している構造化プレゼンテーションを記述する記述データを受け取る段階と、第1のウェブサイトから第1のインスタンスの識別子を引き出す段階と、第2のウェブサイトから第1のインスタンスの第1の属性の第1の値を引き出す段階と、第1のインスタンスの識別子および新規値を事前に存在している構造化プレゼンテーションに追加して新規構造化プレゼンテーション内に新規レコードを形成する段階と、新規構造化プレゼンテーションを視覚的に提示するための命令をクライアントデバイスに出力する段階とを含む。事前に存在している構造化プレゼンテーションの視覚的提示は、構造化設計に適合するシステム配置で情報を視覚的に提示する。構造化プレゼンテーションはレコードのコレクションを含み、これらのそれぞれは構造化プレゼンテーションの視覚的提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す。

【 0 0 1 3 】

この態様および他の態様は、以下の特徴のうちの1つまたは複数を含みうる。1つまたは複数のコンピュータは、データ通信ネットワークを通じてクライアントデバイスをインタラクティブに操作するように動作可能なサーバーを含んでいてよく、またクライアントデバイスは、クライアントとしてサーバーをインタラクティブに操作するように動作可能である。

10

20

30

40

50

## 【0014】

この態様の他の実施形態は、対応するシステム、装置、および方法を含む。

## 【0015】

他の態様では、システムは、クライアントデバイスおよびクライアントデバイスをインタラクティブに操作し、オペレーションを実行するようにプログラムされた1つまたは複数のコンピュータを備える。これらのオペレーションは、クライアントデバイスから機械可読検索クエリを受け取る段階と、検索クエリに関連するインスタンスの構造化プレゼンテーションを提示するための命令をクライアントデバイスに送信することによって検索クエリに応答する段階とを含む。構造化プレゼンテーションの視覚的提示は、インスタンスとインスタンスの属性を特徴付ける値との間の関連付けを、インスタンスの識別子と値の配列によって表す。インスタンスの識別子と値は、電子文書の非構造化コレクション内の2つまたはそれ以上の文書から引き出される。電子文書コレクションは、電子文書コレクション内の電子文書のフォーマットが制限的でも永続的でもないという点で非構造化コレクションである。

10

## 【0016】

この態様および他の態様は、以下の特徴のうちの1つまたは複数を含みうる。1つまたは複数のコンピュータは、データ通信ネットワークを通じてクライアントデバイスをインタラクティブに操作するように動作可能なサーバーを含んでいてよく、またクライアントデバイスは、クライアントとしてサーバーをインタラクティブに操作するように動作可能である。

20

## 【0017】

この態様の他の実施形態は、対応するシステム、装置、および方法を含む。

## 【0018】

本明細書で説明される発明対象の他の態様は、事前に存在している構造化プレゼンテーションを記述する記述データを受け取るアクションと、事前に存在している構造化プレゼンテーションの特性を電子文書の非構造化コレクション内の電子文書のコンテンツと比較して事前に存在している構造化プレゼンテーションに関連する新規インスタンスを識別する電子文書を特定するアクションと、新規インスタンスの識別子を事前に存在している構造化プレゼンテーションに追加して拡張構造化プレゼンテーションを形成するアクションと、拡張構造化プレゼンテーションを提示するための命令を出力するアクションとを含む機械実装方法により具現化されうる。事前に存在している構造化プレゼンテーションの視覚的提示は、構造化設計に適合するシステム配置で情報を視覚的に提示する。構造化プレゼンテーションは、構造化プレゼンテーションの視覚的提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す。

30

## 【0019】

この態様および他の態様は、以下の特徴のうちの1つまたは複数を含みうる。新規インスタンスの識別子を追加する段階は、インスタンス提案(instance suggestions)のコレクションを構成する(formulating a collection of)段階と、インスタンス提案コレクションをユーザーに提供する段階と、新規インスタンスのユーザー選択を受け取る段階とを含むことができ、ただし、新規インスタンスはインスタンス提案のコレクションに含まれる。事前に存在している構造化プレゼンテーションの特性を電子文書のコンテンツと比較する段階は、事前に存在している構造化プレゼンテーションで識別されたインスタンスに関連する構造化コンポーネントを含む電子文書コレクション内の文書を識別する段階を含むことができる。インスタンス提案のコレクションを構成する段階は、事前に存在している構造化プレゼンテーションで識別されたインスタンスの識別子を含み、テンプレートに従って配列されている電子文書コレクション内の第1の文書を識別する段階と、テンプレートに従って配列されているが、第2のインスタンスに関連している第2の文書を識別する段階と、第2のインスタンスをインスタンス提案コレクションに含める段階とを含むことができる。事前に存在している構造化プレゼンテーションの特性を電子文書のコンテンツと

40

50

比較する段階は、事前に存在している構造化プレゼンテーション内の1つまたは複数のインスタンスに関する情報を含む電子文書コレクション内の文書を識別する段階、インスタンスと属性との関連付けの格納されているコレクション内の新規インスタンスを特定する段階、事前に存在している構造化プレゼンテーションの特性を事前に存在している構造化プレゼンテーションにおいて特徴付けられている属性と比較する段階、事前に存在している構造化プレゼンテーション内のインスタンスを特徴付けるために使用される属性を電子文書のコンテンツと比較する段階、事前に存在している構造化プレゼンテーション内のインスタンスを特徴付けるために使用される属性の値を電子文書のコンテンツと比較する段階、および事前に存在している構造化プレゼンテーション内のインスタンスを含むインスタンスのカテゴリを電子文書のコンテンツと比較する段階のうちの1つまたは複数を含む

10

【0020】

この態様の他の実施形態は、対応するシステム、装置、およびそれぞれこれらの方法のオペレーションを実行するように構成されているコンピュータ記憶デバイス上に記録されたコンピュータプログラムを含む。

【0021】

20

本明細書で説明されている発明対象の他の態様は、1つまたは複数のデータ処理機にオペレーションを実行させるように動作可能な命令を格納する1つまたは複数の機械可読データ記憶媒体を備える装置により具現化されうる。これらのオペレーションは、非構造化電子文書コレクション内の2つまたはそれ以上の文書のコンテンツに基づいてインスタンス提案のコレクションを構成する段階と、インスタンス提案コレクションをユーザーに提供する段階と、インスタンス提案のコレクション内の第1のインスタンスのユーザー選択を受け取る段階と、第1のインスタンス提案の識別子を構造化プレゼンテーションに追加する段階とを含む。電子文書コレクションは、電子文書コレクション内の電子文書のフォーマットが制限的でも永続的でもないという点で非構造化コレクションである。構造化プレゼンテーションの視覚的提示により、情報が整理された形で視覚的に提示される。構造化プレゼンテーションは、構造化プレゼンテーションの視覚的提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す。

30

【0022】

この態様および他の態様は、以下の特徴のうちの1つまたは複数を含みうる。インスタンス提案のコレクションを構成する段階は、事前に存在している構造化プレゼンテーションの特性を電子文書コレクション内の電子文書のコンテンツと比較する段階、事前に存在している構造化プレゼンテーション内で識別されているインスタンスに関係する構造化コンポーネントを含む電子文書コレクション内の文書を識別する段階、事前に存在している構造化プレゼンテーションで識別されたインスタンスに関連し、テンプレートに従って配列されている、電子文書コレクション内の第1の文書を識別する段階、テンプレートに従って配列されているが、第2のインスタンスに関連している第2の文書を識別する段階、第2のインスタンスをインスタンス提案コレクションに含める段階、事前に存在している構造化プレゼンテーション内の1つまたは複数のインスタンスの識別子を含む電子文書コレクション内の文書を識別する段階、事前に存在している構造化プレゼンテーション内のインスタンスを特徴付けるために使用される追加の属性を識別する段階、事前に存在している構造化プレゼンテーション内のインスタンスを特徴付けるために使用される属性の値をインスタンス提案の値と比較する段階、事前に存在している構造化プレゼンテーション内のインスタンスを含むインスタンスのカテゴリを識別する段階、インスタンスのそのカテゴリ内のインスタンスを使用してインスタンス提案のコレクションを構成する段階、イン

40

50

スタンスと属性との関連付けの格納されているコレクション内のインスタンス提案を識別する段階、および事前に存在している構造化プレゼンテーションにおいて特徴付けられている属性を非構造化電子文書コレクション内の文書のコンテンツと比較する段階のうちの1つまたは複数を含むことができる。電子文書のコレクションは、インターネット上で利用可能な文書を含むことができる。電子文書は、ウェブページを含むことができる。構造化プレゼンテーションは、テーブル、またはカードのコレクションを含むことができる。

【0023】

この態様の他の実施形態は、対応するシステム、装置、および方法を含む。

【0024】

本明細書で説明されている発明対象の他の態様は、クライアントデバイスおよびクライアントデバイスをインタラクティブに操作し、オペレーションを実行するようにプログラムされた1つまたは複数のコンピュータを備えるシステムにより具現化できる。これらのオペレーションは、事前に存在している構造化プレゼンテーションを記述する記述データを受け取る段階と、事前に存在している構造化プレゼンテーションの特性を電子文書の非構造化コレクション内の電子文書のコンテンツと比較して事前に存在している構造化プレゼンテーションに関連する新規インスタンスを識別する電子文書を特定する段階と、新規インスタンスの識別子を事前に存在している構造化プレゼンテーションに追加して拡張構造化プレゼンテーションを形成する段階と、クライアントデバイスとデータ通信を行うように結合されているディスプレイデバイス上に拡張構造化プレゼンテーションを提示するための命令を出力する段階とを含む。事前に存在している構造化プレゼンテーションの視覚的提示は、構造化設計に適合するシステム配置で情報を視覚的に提示する。構造化プレゼンテーションは、構造化プレゼンテーションの視覚的提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す。

【0025】

この態様の他の実施形態は、対応するコンピュータプログラム製品、装置、および方法を含む。

【0026】

本明細書で説明されている発明対象の他の態様は、クライアントデバイスおよびクライアントデバイスをインタラクティブに操作し、オペレーションを実行するようにプログラムされた1つまたは複数のコンピュータを備えるシステムにより具現化できる。これらのオペレーションは、非構造化電子文書コレクション内の2つまたはそれ以上の文書のコンテンツに基づいてインスタンス提案のコレクションを構成する段階と、クライアントデバイスを使用してインスタンス提案コレクションをユーザーに提供する段階と、インスタンス提案のコレクション内の第1のインスタンスのユーザー選択を受け取る段階と、クライアントデバイスとデータ通信を行うように結合されているディスプレイデバイス上に提示される構造化プレゼンテーションに第1のインスタンス提案の識別子を追加する段階とを含み、構造化プレゼンテーションの視覚的提示により、情報が整理された形で視覚的に提示される。電子文書コレクションは、電子文書コレクション内の電子文書のフォーマットが制限的でも永続的でもないという点で非構造化コレクションである。構造化プレゼンテーションは、構造化プレゼンテーションの視覚的提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す。

【0027】

この態様および他の態様は、以下の特徴のうちの1つまたは複数を含みうる。1つまたは複数のコンピュータは、データ通信ネットワークを通じてクライアントデバイスをインタラクティブに操作するように動作可能なサーバーを含むことができる。クライアントデバイスは、クライアントとしてサーバーをインタラクティブに操作するように動作可能であってよい。クライアントデバイスは、ウェブブラウザを実行するパーソナルコンピュータを含んでよい。パーソナルコンピュータは、ディスプレイデバイスを備えることができる。

## 【0028】

この態様の他の実施形態は、対応するコンピュータプログラム製品、装置、および方法を含む。

## 【0029】

本明細書で説明される発明対象の他の態様は、事前に存在している構造化プレゼンテーションを記述する記述データを受け取るアクションと、事前に存在している構造化プレゼンテーションの特性を電子文書の非構造化コレクション内の電子文書のコンテンツと比較して事前に存在している構造化プレゼンテーションに関連する新規属性を識別する電子文書を特定するアクションと、新規属性の識別子を事前に存在している構造化プレゼンテーションに追加して拡張構造化プレゼンテーションを形成するアクションと、拡張構造化プレゼンテーションを提示するための命令を出力するアクションとを含む機械実装方法により具現化されうる。事前に存在している構造化プレゼンテーションの視覚的提示は、構造化設計に適合するシステム配置で情報を視覚的に提示する。構造化プレゼンテーションは、構造化プレゼンテーションの視覚的提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す。

10

## 【0030】

この態様および他の態様は、以下の特徴のうちの1つまたは複数を含みうる。新規属性の識別子を追加する段階は、属性提案(attribute suggestions)のコレクションを構成する段階と、属性提案コレクションをユーザーに提供する段階と、新規属性のユーザー選択を受け取る段階とを含むことができる。新規属性はインスタンス提案のコレクション内にあってもよい。事前に存在している構造化プレゼンテーションの特性を電子文書のコンテンツと比較する段階は、事前に存在している構造化プレゼンテーションで識別されたインスタンスに関する構造化コンポーネントを含む電子文書コレクション内の文書を識別する段階を含むことができる。属性提案コレクションを構成する段階は、事前に存在している構造化プレゼンテーションで識別されたインスタンスに関連し、テンプレートに従って配列されている、電子文書コレクション内の第1の文書を識別する段階と、属性提案コレクション内のインスタンスを特徴付けるために第1の文書内で使用される属性を追加する段階とを含むことができる。事前に存在している構造化プレゼンテーションの特性を電子文書のコンテンツと比較する段階は、事前に存在している構造化プレゼンテーション内の1つまたは複数のインスタンスに関する情報を含む電子文書コレクション内の文書を識別する段階、インスタンスと属性との関連付けの格納されているコレクション内の新規属性を識別する段階、事前に存在している構造化プレゼンテーションにおいて特徴付けられているインスタンスを電子文書のコンテンツと比較する段階、事前に存在している構造化プレゼンテーションにおいて識別されているインスタンスに関する追加のインスタンスを識別する段階、事前に存在している構造化プレゼンテーション内のインスタンスを特徴付けるために使用される属性または属性の値を電子文書のコンテンツと比較する段階、事前に存在している構造化プレゼンテーション内のインスタンスを含むインスタンスのカテゴリを電子文書のコンテンツと比較する段階のうちの1つまたは複数を含むことができる。電子文書のコレクションは、インターネット上で利用可能な電子文書を含むことができ、電子文書は、ウェブページを含むことができる。拡張構造化プレゼンテーションは、テーブル、またはカードのコレクションを含むことができる。この方法は、ディスプレイ画面の1つまたは複数の要素を物理的に変換する段階を含めて、拡張構造化プレゼンテーションをディスプレイ画面上に視覚的に提示する段階を含むことができる。

20

30

40

## 【0031】

この態様の他の実施形態は、対応するシステム、装置、およびコンピュータプログラム製品を含む。

## 【0032】

本明細書で説明されている発明対象の他の態様は、1つまたは複数のデータ処理機にオペレーションを実行させるように動作可能な命令を格納する1つまたは複数の機械可読データ記憶媒体を備える装置により具現化されうる。これらのオペレーションは、非構造化

50

電子文書コレクション内の2つまたはそれ以上の文書のコンテンツに基づいて属性提案のコレクションを構成する段階と、属性提案コレクションをユーザーに提供する段階と、属性提案のコレクション内の第1の属性のユーザー選択を受け取る段階と、第1の属性提案の識別子を構造化プレゼンテーションに追加する段階とを含むことができる。電子文書コレクションは、電子文書コレクション内の電子文書のフォーマットが制限的でも永続的でもないという点で非構造化コレクションである。構造化プレゼンテーションの視覚的提示により、情報が整理された形で視覚的に提示される。構造化プレゼンテーションは、構造化プレゼンテーションの提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す。

【0033】

この態様および他の態様は、以下の特徴のうちの1つまたは複数を含みうる。属性提案のコレクションを構成する段階は、事前に存在している構造化プレゼンテーションの特性を電子文書コレクション内の電子文書のコンテンツと比較する段階、事前に存在している構造化プレゼンテーション内で識別されているインスタンスを特徴付ける構造化コンポーネントを含む電子文書コレクション内の文書を識別する段階、事前に存在している構造化プレゼンテーションで識別されたインスタンスに関連し、テンプレートに従って配列されている、電子文書コレクション内の第1の文書を識別する段階と、属性提案コレクション内のインスタンスを特徴付けるために使用される属性を含める段階、事前に存在している構造化プレゼンテーション内の1つまたは複数のインスタンスに関する情報を含む電子文書コレクション内の文書を識別する段階のうちの1つまたは複数を含むことができる。事前に存在している構造化プレゼンテーションの特性を電子文書のコンテンツと比較する段階は、事前に存在している構造化プレゼンテーションにおいて識別されているインスタンスを電子文書のコンテンツと比較する段階、事前に存在している構造化プレゼンテーション内のインスタンスを特徴付けるために使用される属性または属性の値を電子文書のコンテンツと比較する段階のうちの1つまたは複数を含むことができる。属性提案のコレクションを構成する段階は、事前に存在している構造化プレゼンテーション内のインスタンスを含むインスタンスのカテゴリを識別する段階と、インスタンスのそのカテゴリ内のインスタンスを特徴付けるために使用される属性から属性提案のコレクションを構成する段階とを含むことができる。属性提案のコレクションは、インスタンスと属性との関連付けの格納されているコレクション内の属性提案を識別することによって構成することもできる。電子文書のコレクションは、インターネット上で利用可能な電子文書を含むことができ、電子文書は、ウェブページを含むことができる。構造化プレゼンテーションは、テーブル、またはカードのコレクションを含むことができる。これらのオペレーションは、ディスプレイ画面の1つまたは複数の要素を物理的に変換する段階を含めて、構造化プレゼンテーションをディスプレイ画面上に視覚的に提示する段階を含むこともできる。

【0034】

この態様の他の実施形態は、対応するシステム、装置、およびコンピュータプログラム製品を含む。

【0035】

本明細書で説明されている発明対象の他の態様は、ディスプレイ画面を備えるクライアントデバイスおよびクライアントデバイスをインタラクティブに操作し、オペレーションを実行するようにプログラムされた1つまたは複数のコンピュータを備えるシステムにより具現化できる。これらのオペレーションは、事前に存在している構造化プレゼンテーションを記述する記述データを受け取る段階と、事前に存在している構造化プレゼンテーションの特性を電子文書の非構造化コレクション内の電子文書のコンテンツと比較して事前に存在している構造化プレゼンテーションに関連する新規属性を識別する電子文書を特定する段階と、新規属性の識別子を事前に存在している構造化プレゼンテーションに追加して拡張構造化プレゼンテーションを形成する段階と、拡張構造化プレゼンテーションをディスプレイ画面上に提示するための命令を出力する段階とを含む。事前に存在している構造化プレゼンテーションの視覚的提示は、構造化設計に適合するシステム配置で情報を視

10

20

30

40

50

覚的に提示する。構造化プレゼンテーションは、構造化プレゼンテーションの視覚的提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す。

【0036】

この態様の他の実施形態は、対応するコンピュータプログラム製品、装置、および方法を含む。

【0037】

本明細書で説明されている発明対象の他の態様は、ディスプレイ画面を備えるクライアントデバイスおよびクライアントデバイスをインタラクティブに操作し、オペレーションを実行するようにプログラムされた1つまたは複数のコンピュータを備えるシステムにより具現化できる。これらのオペレーションは、非構造化電子文書コレクション内の2つまたはそれ以上の文書のコンテンツに基づいて属性提案のコレクションを構成する段階と、属性提案コレクションをクライアントデバイスに提供する段階と、クライアントデバイスから属性提案のコレクション内の第1の属性の選択を受け取る段階と、第1の属性提案の識別子をディスプレイ画面上に提示される構造化プレゼンテーションに追加する段階とを含む。電子文書コレクションは、電子文書コレクション内の電子文書のフォーマットが制限的でも永続的でもないという点で非構造化コレクションである。構造化プレゼンテーションの視覚的提示により、情報が整理された形で視覚的に提示される。構造化プレゼンテーションは、構造化プレゼンテーションの提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す。

10

20

【0038】

この態様の他の実施形態は、対応するコンピュータプログラム製品、装置、および方法を含む。

【0039】

本明細書で説明される発明対象の他の態様は、事前に存在している構造化プレゼンテーションを記述する記述データを受け取るアクションと、事前に存在している構造化プレゼンテーションの特性を電子文書の非構造化コレクション内の電子文書のコンテンツと比較して事前に存在している構造化プレゼンテーションに関連する新規値を識別する電子文書を特定するアクションと、新規値を事前に存在している構造化プレゼンテーションに追加して新規構造化プレゼンテーションを形成するアクションと、新規構造化プレゼンテーションを視覚的に提示するための命令を出力するアクションとを含む機械実装方法により具現化される。事前に存在している構造化プレゼンテーションの視覚的提示は、構造化設計に適合するシステム配置で情報を視覚的に提示する。構造化プレゼンテーションは、構造化プレゼンテーションの視覚的提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す。

30

【0040】

この態様および他の態様は、以下の特徴のうちの1つまたは複数を含みうる。第1の電子文書において構造化プレゼンテーション内に出現する第1のインスタンスの識別子を特定し、新規値を第1の電子文書から抽出することができる。新規値を追加する段階は、第1のインスタンスの第1の属性の値のコレクションを識別する段階と、識別された値のうちの1つまたは複数の値のサブセットを、第1のインスタンスの第1の属性を適切に特徴付けるものとして確定する段階とを含むことができる。値のサブセットを適切なものとして確定する段階は、コレクション内の値をいくつかのグループに分ける段階、サブセット内の値の個数に少なくとも一部は基づいてサブセットを選択する段階、ユーザー指定制約条件を満たすサブセット内の値に少なくとも一部は基づいてサブセットを選択する段階、サブセット内の値が高品質文書から引き出される段階に少なくとも一部は基づいてサブセットを選択する段階、サブセット内の値が事前に存在している構造化プレゼンテーション内の他のインスタンスに関連する文書から引き出される段階に少なくとも一部は基づいてサブセット

40

50

を選択する段階のうちの1つまたは複数を含むことができる。電子文書のコレクションは、インターネットであってもよく、電子文書は、ウェブページであってもよい。事前に存在している構造化プレゼンテーションは、テーブル、またはカードのコレクションを含むことができる。この方法は、ディスプレイ画面の1つまたは複数の要素を物理的に変換する段階を含めて、新規構造化プレゼンテーションをディスプレイ画面上に視覚的に提示する段階を含むことができる。

【0041】

この態様の他の実施形態は、対応するシステム、装置、およびコンピュータプログラム製品を含む。

【0042】

本明細書で説明されている発明対象の他の態様は、1つまたは複数のデータ処理機にオペレーションを実行させるように動作可能な命令を格納する1つまたは複数の機械可読データ記憶媒体を備える装置により具現化されうる。これらのオペレーションは、第1のインスタンス、第2のインスタンス、および第1の属性を記述する記述データを受け取る段階と、非構造化電子文書コレクションの2つまたはそれ以上の文書から第1のインスタンスの第1の属性の値の第1のコレクションを抽出する段階と、非構造化電子文書コレクションの2つまたはそれ以上の文書から第2のインスタンスの第1の属性の値の第2のコレクションを抽出する段階と、値の第1のコレクションの第1のサブセットを、第1のインスタンスの第1の属性を適切に特徴付けるものとして確定する段階と、値の第2のコレクションの第2のサブセットを、第2のインスタンスの第1の属性を適切に特徴付けるものとして確定する段階と、第1のサブセットの第1の値および第2のサブセットの第2の値を含む構造化プレゼンテーションを表示するための機械可読命令を生成する段階とを含むことができる。構造化プレゼンテーションは、インスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す。

【0043】

この態様および他の態様は、以下の特徴のうちの1つまたは複数を含みうる。値の第1のサブセットは、第1のコレクション内の値をいくつかのグループに分けることによって適切なものとして確定することができる。ただし、それぞれのグループは値の第1のコレクションのサブセットを含む。値の第1のサブセットは、第1のサブセット内の値の個数に少なくとも一部に基づき第1のサブセットを選択することによって適切なものとして確定することができる。値の第1のサブセットは、第1のサブセット内の値をそれらの値に対するユーザー指定制約条件と比較することによって適切なものとして確定することができる。値の第1のサブセットは、第1のサブセット内の値が高品質文書から引き出されたと判定することによって適切なものとして確定することができる。値の第1のサブセットは、第1のサブセット内の値が第2のインスタンスに関連する文書から引き出されたと判定することによって適切なものとして確定することができる。値の第1のサブセットは、第1のサブセット内の値が第1のインスタンスおよび第2のインスタンスの両方を特徴付ける他の属性に関連する文書から引き出されたと判定することによって適切なものとして確定することができる。第1のインスタンスの記述は、事前に存在している構造化プレゼンテーション内に出現する第1のインスタンスの識別子を含むことができる。第2のインスタンスの記述は、事前に存在している構造化プレゼンテーション内に出現する第2のインスタンスの識別子を含むことができる。第1の属性の記述は、事前に存在している構造化プレゼンテーションに追加される新規属性の記述を含むことができる。非構造化電子文書コレクションは、インターネット上で利用可能な電子文書を含むことができる。構造化プレゼンテーションは、テーブル、またはカードのコレクションであってもよい。構造化プレゼンテーションは、ディスプレイ画面の1つまたは複数の要素を物理的に変換する段階を含めて、ディスプレイ画面上に視覚的に提示することができる。

【0044】

この態様の他の実施形態は、対応するシステム、装置、および方法を含む。

【0045】

10

20

30

40

50

本明細書で説明されている発明対象の他の態様は、デバイスおよびデバイスをインタラクティブに操作し、オペレーションを実行するようにプログラムされた1つまたは複数のコンピュータを備えるシステムにより具現化されうる。これらのオペレーションは、事前に存在している構造化プレゼンテーションを記述する記述データを受け取る段階と、事前に存在している構造化プレゼンテーションの特性を電子文書の非構造化コレクション内の電子文書のコンテンツと比較して事前に存在している構造化プレゼンテーションに関連する新規値を識別する電子文書を特定する段階と、新規値を事前に存在している構造化プレゼンテーションに追加して新規構造化プレゼンテーションを形成する段階と、新規構造化プレゼンテーションをデバイス上に視覚的に提示するための命令を出力する段階とを含む。事前に存在している構造化プレゼンテーションの視覚的提示は、構造化設計に適合するシステム配置で情報を視覚的に提示する。構造化プレゼンテーションは、構造化プレゼンテーションの視覚的提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す。

10

**【0046】**

この態様の他の実施形態は、対応するコンピュータプログラム製品、装置、および方法を含む。

**【0047】**

本明細書で説明されている発明対象の他の態様は、デバイスおよびデバイスをインタラクティブに操作し、オペレーションを実行するようにプログラムされた1つまたは複数のコンピュータを備えるシステムにより具現化されうる。これらのオペレーションは、第1のインスタンス、第2のインスタンス、および第1の属性を記述する記述データを受け取る段階と、非構造化電子文書コレクションの2つまたはそれ以上の文書から第1のインスタンスの第1の属性の値の第1のコレクションを抽出する段階と、非構造化電子文書コレクションの2つまたはそれ以上の文書から第2のインスタンスの第1の属性の値の第2のコレクションを抽出する段階と、値の第1のコレクションの第1のサブセットを、第1のインスタンスの第1の属性を適切に特徴付けるものとして確定する段階と、値の第2のコレクションの第2のサブセットを、第2のインスタンスの第1の属性を適切に特徴付けるものとして確定する段階と、第1のサブセットの第1の値および第2のサブセットの第2の値を含む構造化プレゼンテーションを表示するための機械可読命令を生成する段階であって、構造化プレゼンテーションは、インスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す、段階と、機械可読命令をデバイスに送信する段階とを含む。

20

30

**【0048】**

この態様の他の実施形態は、対応するコンピュータプログラム製品、装置、および方法を含む。

**【0049】**

本明細書で説明される発明対象の他の態様は、構造化プレゼンテーションをディスプレイデバイス上に表示するアクションと、表示される構造化プレゼンテーションに対するユーザーインタラクションを特徴付ける、構造化プレゼンテーションの第1のインスタンスおよび第1の属性の指定を含むデータを受け取るアクションと、データを受け取ったことに応答してディスプレイデバイス上に以前には隠されていた検索インターフェイスを表示するアクションとを含む機械実装方法によって具現化されうる。構造化プレゼンテーションは、構造化設計に適合する組織的な構造化された配列で情報を視覚的に提示する。構造化プレゼンテーションは、構造化プレゼンテーションの視覚的提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す。検索インターフェイスは、電子文書コレクション内の第1のインスタンスの第1の属性を特徴付ける第1の値の位置を識別する情報もしくはインタラクティブ要素を含む。

40

**【0050】**

この態様および他の態様は、以下の特徴のうちの1つまたは複数を含む。表示される構

50

造化プレゼンテーションに対するユーザーインタラクションを特徴付けるデータを受け取る段階は、構造化プレゼンテーション内のセルに関連付けられている第1のインスタンスおよび第1の属性の手動によるユーザー指定を受け取る段階、または構造化プレゼンテーション内のセルに対するユーザーインタラクションを特徴付けるデータを受け取る段階を含むことができる。セルは、構造化プレゼンテーション内の第1のインスタンスおよび第1の属性の識別子に関するセルの配列を用いて第1のインスタンスおよび第1の属性に関連付けることができる。セルに対するユーザーインタラクションを特徴付けるデータを受け取る段階は、空のセルに対するユーザーインタラクションを特徴付けるデータを受け取る段階も含みうる。以前には隠されていた検索インターフェイスを表示する段階は、電子文書コレクションの検索をトリガーし、第1の値を特定するためにユーザー側で選択できるインタラクティブ要素を表示する段階、第1のインスタンスの第1の属性を特徴付ける値を指定するためにユーザー側で選択できるインタラクティブ値入力要素を表示する段階、電子文書コレクションの第1の文書内の第1の値のコンテキストを特徴付けるスニペットを表示する段階、および第1の値を特定するために電子文書コレクションの事前検索の結果を表示する段階のうちの1つまたは複数を含むことができる。第1の値は、第1のインスタンスの第1の属性を特徴付ける値として構造化プレゼンテーション内に出現しうる。以前には隠されていた検索インターフェイスを表示する段階は、電子文書コレクション内の第1の電子文書の識別子を表示する段階も含むことができ、第1の値は、第1の電子文書から引き出される。この方法は、第1の電子文書が機能せず第1の値を提供することができないと判定する段階と、第1の文書が機能しないことを示す視覚的指示を表示する段階も含むことができる。ユーザーに対して、第1の文書と第1のインスタンスおよび第1の属性との関連性の変化に関係なく第1の文書から矛盾なく第1の値を選択するオプション、または第1のインスタンスおよび第1の属性に最も関連している第1の文書から第1の値を選択するオプションを提示することができる。この方法は、検索インターフェイスに対するユーザーインタラクションにตอบสนองして第1の値を特定するために電子文書の非構造化コレクションを検索する段階と、第1の値を構造化プレゼンテーションに追加する段階も含むことができる。第1のインスタンスおよび第1の属性の指定を受け取る段階は、属性のコレクションまたはインスタンスのコレクションの指定を受け取る段階を含むことができる。この方法は、時間の経過に応じて構造化プレゼンテーションの表示を更新する段階も含みうる。

**【0051】**

この態様の他の実施形態は、対応するシステム、装置、およびコンピュータプログラム製品を含む。

**【0052】**

本明細書で説明されている発明対象の他の態様は、クライアントデバイスをインタラクティブに操作し、オペレーションを実行するようにプログラムされた1つまたは複数のコンピュータを備えるシステムにより具現化されうる。これらのオペレーションは、ディスプレイデバイス上に表示される構造化プレゼンテーションの第1のセルを指定するユーザーインタラクションを特徴付けるデータを受け取る段階と、第1のセルに第1の値を書き込むために事前検索が実施されたと判定する段階と、事前検索が実施されたと判定したことに応答して、ディスプレイデバイス上に事前検索を特徴付ける情報を表示する段階とを含む。構造化プレゼンテーションは、構造化設計に適合する組織的な構造化された配列で情報を視覚的に提示する。構造化プレゼンテーションは、インスタンスとインスタンスの属性を特徴付ける値との間の関連付けをセル内の値の配列によって表す。

**【0053】**

この態様および他の態様は、以下の特徴のうちの1つまたは複数を含む。第1のセルを指定するユーザーインタラクションを特徴付けるデータは、第1のセルに関連付けられている第1のインスタンスおよび第1の属性の手動によるユーザー指定を含むことができる。事前検索を特徴付ける情報は、第1の値の引き出し元となる電子文書を識別する情報を含むことができる。事前検索を特徴付ける情報は、第1の値の引き出し元になった可能性のある電子文書のコレクション、第1の値の引き出し元となる電子文書コレクション内の第1の

10

20

30

40

50

電子文書を識別する情報、電子文書コレクションの第1の文書内の第1の値のコンテキストを特徴付けるスニペットのうちの1つまたは複数を含むことができる。事前検索を特徴付ける情報は、例えば以前には隠されていた検索インターフェイスの表示要素内に、表示することができる。これらのオペレーションは、第1の電子文書が操作不能であり第1の値を提供することができないと判定する段階と、第1の文書が操作不能であることを示す視覚的指示を表示する段階も含むことができる。これらのオペレーションは、ユーザーインタラクションにตอบสนองして構造化プレゼンテーションの第1のセル内の値の表示を更新する段階も含みうる。電子文書のコレクションは、インターネット上で利用可能な電子文書を含むことができる。電子文書は、ウェブページを含むことができる。構造化プレゼンテーションは、カードのコレクションであってよい。

10

**【0054】**

この態様の他の実施形態は、対応するコンピュータプログラム製品、装置、およびコンピュータプログラム製品を含む。

**【0055】**

本明細書で説明されている発明対象の他の態様は、ディスプレイデバイスを備えるクライアントデバイスをインタラクティブに操作し、オペレーションを実行するようにプログラムされた1つまたは複数のコンピュータを備えるシステムにより具現化されうる。これらのオペレーションは、構造化プレゼンテーションをディスプレイデバイス上に表示する段階と、表示される構造化プレゼンテーションに対するユーザーインタラクションを特徴付けるデータを受け取る段階と、データを受け取ったことにตอบสนองしてディスプレイデバイス上に以前には隠されていた検索インターフェイスを表示する段階とを含む。構造化プレゼンテーションは、構造化設計に適合する組織的な構造化された配列で情報を視覚的に提示する。構造化プレゼンテーションは、構造化プレゼンテーションの視覚的提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す。データは、構造化プレゼンテーションの第1のインスタンスおよび第1の属性の指定を含む。検索インターフェイスは、電子文書コレクション内の第1のインスタンスの第1の属性を特徴付ける第1の値の位置を識別する情報もしくはインタラクティブ要素を含む。

20

**【0056】**

この態様の他の実施形態は、対応するコンピュータプログラム製品、装置、およびコンピュータプログラム製品を含む。

30

**【0057】**

1つまたは複数の実施形態の詳細は、付属の図面および以下の説明で述べられる。他の特徴および利点は、説明および図面、ならびに請求項から明らかになるであろう。

**【図面の簡単な説明】****【0058】**

【図1】電子文書コレクションからの情報が構造化プレゼンテーションでユーザーに提示されるシステムを示す略図である。

【図2】電子文書コレクションからの情報が構造化プレゼンテーションでユーザーに提示される他のシステムの一実装を示す略図である。

40

【図3】例示的な構造化プレゼンテーションの略図である。

【図4】例示的な構造化プレゼンテーションの略図である。

【図5】例示的な構造化プレゼンテーションの略図である。

【図6】電子文書コレクションからの情報を構造化プレゼンテーションでユーザーに提示するための例示的なプロセスの流れ図である。

【図7】電子文書コレクション内の2つまたはそれ以上の関連する文書を識別するための例示的なプロセスの流れ図である。

【図8】電子文書コレクション内の2つまたはそれ以上の関連する文書を識別するための例示的なプロセスの流れ図である。

【図9】新規インスタンスを構造化プレゼンテーションに提案および/または追加するた

50

めのプロセスの流れ図である。

【図10】構造化プレゼンテーションの修正を指定するユーザー入力を受け取るためのユーザーインターフェイスコンポーネントの略図である。

【図11】新規インスタンスを構造化プレゼンテーションに追加するための技術を指定するユーザー入力を受け取るためのユーザーインターフェイスコンポーネントの略図である。

【図12】新規インスタンスを構造化プレゼンテーションに追加するためにユーザー指定制約条件オプションで使用される制約条件を指定するユーザー入力を受け取るためのユーザーインターフェイスコンポーネントの略図である。

【図13】新規属性を構造化プレゼンテーションに追加するための例示的なプロセスの流れ図である。

【図14】新規属性を構造化プレゼンテーションに追加するためのユーザーインターフェイスコンポーネントの略図である。

【図15】新規属性値を構造化プレゼンテーションに追加するための例示的なプロセスの流れ図である。

【図16】新規属性値を構造化プレゼンテーションに追加するための例示的なプロセスの流れ図である。

【図17】構造化プレゼンテーションに追加すべき候補値を選択するためのユーザーインターフェイスコンポーネントの略図である。

【図18】中に提示される属性値の不足のハイライトを含む構造化プレゼンテーションの略図である。

【図19】構造化プレゼンテーションに追加すべき候補属性を選択するためのユーザーインターフェイスコンポーネントの略図である。

【図20】構造化プレゼンテーションに追加すべき候補インスタンスを選択するためのユーザーインターフェイスコンポーネントの略図である。

【図21】事前に存在している構造化プレゼンテーションを拡張するために新規インスタンスを追加することができるプロセスの略図である。

【図22】電子文書コレクション内の文書のコンテンツに基づいてインスタンスを構造化プレゼンテーションに追加するための例示的なプロセスの流れ図である。

【図23】事前に存在している構造化プレゼンテーションの特性に基づいて電子文書コレクション内の電子文書からインスタンス提案を構成するための例示的なプロセスの流れ図である。

【図24】事前に存在している構造化プレゼンテーションの特性に基づく電子文書コレクション内の電子文書からのインスタンス提案の構成を表す図である。

【図25】事前に存在している構造化プレゼンテーションの特性に基づいて電子文書コレクション内の電子文書からインスタンス提案を構成するための例示的なプロセスの流れ図である。

【図26】動画を記述するためのパターンとして使用されるハイパーテキストマークアップ言語のテンプレートの一部を表す図である。

【図27】事前に存在している構造化プレゼンテーション内の情報に基づいて新規インスタンス提案のコレクションを構成することができるプロセスの略図である。

【図28】電子文書コレクション内の属性およびインスタンスを関連付けるテーブルの略図である。

【図29】事前に存在している構造化プレゼンテーションの特性に基づいてインスタンスおよび属性のコレクションからインスタンス提案を構成するためのプロセスの流れ図である。

【図30】事前に存在している構造化プレゼンテーション内の情報に基づいて新規インスタンス提案のコレクションを構成するためのプロセスの流れ図である。

【図31】事前に存在している構造化プレゼンテーション内の情報に基づいて新規インスタンス提案のコレクションを構成するためのプロセスの流れ図である。

10

20

30

40

50

【図32】データコレクション内の属性、インスタンス、およびそれらの値を関連付けるテーブルの略図である。

【図33】事前に存在している構造化プレゼンテーション内の情報に基づいて新規インスタンス提案のコレクションを構成するためのプロセスの流れ図である。

【図34】事前に存在している構造化プレゼンテーションの特性に基づく電子文書コレクション内の電子文書からのインスタンス提案の構成を表す図である。

【図35】事前に存在している構造化プレゼンテーション内の情報に基づいて新規インスタンス提案のコレクションを構成するために使用されうるプロセスのコレクションの略図である。

【図36】事前に存在している構造化プレゼンテーション内の情報に基づいて新規インスタンス提案のコレクションを構成するためのプロセスの流れ図である。

10

【図37】事前に存在している構造化プレゼンテーションを拡張するために新規属性を追加することができるプロセスの略図である。

【図38】電子文書コレクション内の文書のコンテンツに基づいて属性を構造化プレゼンテーションに追加するための例示的なプロセスの流れ図である。

【図39】事前に存在している構造化プレゼンテーションの特性に基づいて電子文書コレクション内の電子文書から属性提案を構成するための例示的なプロセスの流れ図である。

【図40】事前に存在している構造化プレゼンテーションの特性に基づく電子文書コレクション内の電子文書からの属性提案の構成を表す図である。

【図41】事前に存在している構造化プレゼンテーションの特性に基づいて電子文書コレクション内の電子文書から属性提案を構成するための例示的なプロセスの流れ図である。

20

【図42】動画を記述するためのパターンとして使用されるハイパーテキストマークアップ言語 (HTML) のテンプレートの一部を表す図である。

【図43】事前に存在している構造化プレゼンテーション内の情報に基づいて新規属性提案のコレクションを構成することができるプロセスの略図である。

【図44】電子文書コレクション内の属性およびインスタンスを関連付けるテーブルの略図である。

【図45】事前に存在している構造化プレゼンテーションの特性に基づいてインスタンスおよび属性のコレクションから属性提案を構成するためのプロセスの流れ図である。

【図46】事前に存在している構造化プレゼンテーション内の情報に基づいて新規属性提案のコレクションを構成するためのプロセスの流れ図である。

30

【図47】事前に存在している構造化プレゼンテーション内の情報に基づいて属性提案を構成する際に使用する関係するインスタンスを識別するためのプロセスの流れ図である。

【図48】事前に存在している構造化プレゼンテーション内の情報に基づいて新規属性提案のコレクションを構成するためのプロセスの流れ図である。

【図49】事前に存在している構造化プレゼンテーションの特性に基づく電子文書コレクション内の電子文書からの属性提案の構成を表す図である。

【図50】事前に存在している構造化プレゼンテーション内の情報に基づいて新規属性提案のコレクションを構成するために使用されうるプロセスのコレクションの略図である。

【図51】事前に存在している構造化プレゼンテーション内の情報に基づいて新規属性提案のコレクションを構成するためのプロセスの流れ図である。

40

【図52】電子文書コレクション内の2つまたはそれ以上の電子文書から引き出される属性値が構造化プレゼンテーションでユーザーに提示されるシステムを示す略図である。

【図53】電子文書コレクション内の2つまたはそれ以上の電子文書から引き出される属性値が構造化プレゼンテーションでユーザーに提示されるシステムの一実装を示す略図である。

【図54】電子文書コレクション内の属性、値、およびインスタンスを関連付けることができるテーブルの略図である。

【図55】電子文書コレクション内の2つまたはそれ以上の電子文書から引き出される属性値を構造化プレゼンテーションでユーザーに提示するための例示的なプロセスの流れ図

50

である。

【図56】構造化プレゼンテーションで提示する1つまたは複数の値を選択するためのプロセスの流れ図である。

【図57】構造化プレゼンテーションで提示する1つまたは複数の値を選択するためのプロセスの流れ図である。

【図58】構造化プレゼンテーションで提示する1つまたは複数の値を選択するための例示的なプロセスの流れ図である。

【図59】電子文書コレクション内の電子文書から引き出される属性値が構造化プレゼンテーションでユーザーに提示される状況を示す略図である。

【図60】属性と属性値の両方が電子文書コレクション内の電子文書から引き出され、構造化プレゼンテーションでユーザーに提示されるプロセスを示す略図である。

【図61】電子文書コレクション内の文書のコンテンツに基づいて値を構造化プレゼンテーションに追加するためのプロセスの流れ図である。

【図62】検索インターフェイスが隠されている構造化プレゼンテーションの略図である。

【図63】検索インターフェイスが隠されている構造化プレゼンテーションの略図である。

【図64】検索インターフェイスが隠されている構造化プレゼンテーションの略図である。

【図65】以前には隠されていた検索インターフェイスが提示される表示要素を例示する図である。

【図66】以前には隠されていた検索インターフェイスが提示される表示要素を例示する図である。

【図67】以前には隠されていた検索インターフェイスが提示される表示要素を例示する図である。

【図68】以前には隠されていた検索インターフェイスが提示される表示要素を例示する図である。

【図69】以前には隠されていた検索インターフェイスが提示される表示要素を例示する図である。

【図70】以前には隠されていた検索インターフェイスが提示される表示要素を例示する図である。

【図71】電子文書コレクション内の文書のコンテンツから値を引き出すことによって値を構造化プレゼンテーションに追加するためのプロセスの流れ図である。

【図72】以前には隠されていた検索インターフェイスが提示される表示要素を例示する図である。

【図73】以前には隠されていた検索インターフェイスが提示される表示要素を例示する図である。

【図74】以前には隠されていた検索インターフェイスが提示される表示要素を例示する図である。

【図75】電子文書コレクション内の文書のコンテンツに基づいて値を構造化プレゼンテーションに追加するためのプロセスの流れ図である。

【発明を実施するための形態】

【0059】

さまざまな図面内の類似の参照記号は、類似の要素を示す。

【0060】

図1は、非構造化電子文書コレクション102からの情報が構造化プレゼンテーション106でユーザーに提示されるシステム100を示す略図である。電子文書コレクション102に加えて、システム100は、ディスプレイ画面104およびデータ通信インフラストラクチャ108を備える。動作中、システム100は、電子文書102の非構造化コレクションから情報を抽出し、その抽出された情報を構造化プレゼンテーション106でディスプレイ画面104に提示する

10

20

30

40

50

## 【 0 0 6 1 】

電子文書コレクション102は、電子文書コレクション102内の個別の文書内の情報の編成が情報の抽出に際して利用することができる所定の構造に適合する必要はないという点で非構造化コレクションである。例えば、電子文書コレクション102内に3つの電子文書、つまり、電子文書110、112、114を考える。各電子文書のコンテンツを別々に編成する3人の異なるユーザーによって文書110、112、114がコレクション102に追加された。ユーザーは、文書110、112、114内の情報が、特定のフォーマットであることを保証するために共同作業する必要はない。さらに、一方のユーザーが文書110のフォーマットの変更を望んでいる場合、そのユーザーは他のユーザーによって追加された文書のフォーマットと無関係にそのようにすることができる。ユーザーが他のユーザーにその変更を通知する必要はない。実際、文書は、自動車メーカー3社、またはデジタルカメラの販売会社3社など、共同作業できないだけでなく、お互い敵同士である競合会社でもあるエンティティによってコレクション102に追加される場合もある。文書をコレクション102に追加するエンティティ同士の特定の連携があっても、文書内の情報がそれらの文書内において同様に編成されていることを保証する公式のメカニズムはない。さらに、コレクション102内のそれぞれの文書のそれぞれにおける情報の編成に変更がないことを保証する公式のメカニズムもない。

10

## 【 0 0 6 2 】

それとは対照的に、構造化プレゼンテーション106は、構造化されており、整理されたシステム配置のコレクション102内の文書から引き出された情報を提示する。したがって、構造化プレゼンテーション106内の情報のグループ分け、セグメント分割、および配列は、その中の情報がコレクション102内の多様な一組の文書における異なるコンテキストから引き出されるとしても、構造化設計に従う。さらに、構造化プレゼンテーション106の設計の一態様に加えられた変更は、構造化プレゼンテーション106全体に伝搬しうる。

20

## 【 0 0 6 3 】

構造化プレゼンテーションの例としては、スプレッドシートテーブル、カードまたは他のレコードのコレクション、および他の構造化プレゼンテーションフォーマットが挙げられる。このような構造化プレゼンテーションは、ディスプレイにおける情報の空間的配列、構造化プレゼンテーションの編成および情報に関するさまざまな態様(例えば、列見出し、行見出し、単位識別子、および同様のもの)の位置決めおよび識別、値のグラフィック表現、および他の特性を指定する規則に従うものとして行うことができる。

30

## 【 0 0 6 4 】

構造化プレゼンテーション内の情報の構造化により、一般的に、閲覧者による情報の理解がしやすくなる。例えば、閲覧者は、見出しを読み取ることによって構造化プレゼンテーション内に格納されている情報の性質を認識することができる。閲覧者は、構造化プレゼンテーション内に記述されている値を、ディスプレイに示されているそれらの値の配列および位置決めに基づいて容易に識別し、比較することができる。例えば、ユーザーは、構造化プレゼンテーション内のいくつかの値がすべて、異なる自動車の属性(つまり、特性)に関係することを簡単に確認し、それらの値を容易に比較することができる。

40

## 【 0 0 6 5 】

システム100は、コレクション102内の文書から引き出された値を構造化プレゼンテーション106内に単に書き込むことに制限されない。その代わりに、多くの実装において、システム100は、構造化プレゼンテーション106内に記述されるエンティティ(つまり、「インスタンス」)、それらのインスタンスの属性を特徴付ける値、さらには構造化プレゼンテーション106の適切な構造化を決定することができる。このような決定は、特定のフォーマット、永続的フォーマット、またはその両方を有することに制限されていないコレクション102内の異なる文書から引き出された情報に基づいて行うことができる。例えば、構造化プレゼンテーション106内に出現する属性は、以下でさらに詳しく説明するように、いくつかのインスタンスを特徴付けるためにコレクション102内の文書で使用される属

50

性に基づくものとすることができる。他の例として、構造化プレゼンテーション106内に出現する値の単位(例えば、メートル、フィート、インチ、マイル)は、コレクション102内の文書内に出現する値の単位に基づくものとするすることができる。他の例として、構造化プレゼンテーション106内に出現するインスタンスは、コレクション102内の文書内に出現するインスタンスのコレクションに基づいて決定されうる。

**【 0 0 6 6 】**

さらに、多くの実装において、そのような情報は、コレクション102内の以前に指定されていない文書から引き出すことができる。例えば、検索クエリを使用して、コレクション102内の文書を識別し、それらの文書から情報を引き出すことができる。情報の引き出し元となりうる文書の識別または種類に対し事前に存在する制限がある必要はない。例えば、識別された文書は、特定の個人のアカウントに関連付けられるもの、または特定の小売店に由来するものに制限される必要はない。その代わりに、情報は、以前に指定されていない文書から引き出すことができる。

10

**【 0 0 6 7 】**

システム100は、こうして、さまざまな異なる方法でコレクション102内の文書の多様な情報内容を利用して、構造化プレゼンテーションをユーザーに提示することができる。電子文書コレクション102が多数の文書を含む場合、利用できる情報の量は、非常に大きくなる可能性がある。さらに、これは、以下でさらに詳しく説明するように、多くの場合、自動的に、または人によるインタラクティブな操作を比較的わずかに抑えて実行することができる。

20

**【 0 0 6 8 】**

図2は、非構造化電子文書コレクション102からの情報が構造化プレゼンテーション106でユーザーに提示されるシステム200の一実装を示す略図である。システム200では、データ通信インフラストラクチャ108が、電子文書コレクション102と、ディスプレイ画面104と、検索エンジン202、クローラー204、データセンター208、ならびに、文書圧縮、インデックス付け、および順位付けモジュール210を含むデータ記憶要素および処理要素のコレクションとを相互接続する。

**【 0 0 6 9 】**

検索エンジン202は、非構造化電子文書コレクション102を検索するように、1つまたは複数の機械可読命令セットによりプログラムされる。検索エンジン202は、1つまたは複数の地理的位置に配備されている1つまたは複数のコンピュータ上に実装することができる。

30

**【 0 0 7 0 】**

クローラー204は、非構造化電子文書コレクション102に対しクロールを実行するように、1つまたは複数の機械可読命令セットによりプログラムされる。クローラー204は、1つまたは複数の地理的位置に配備されている1つまたは複数のコンピュータ上に実装することができる。

**【 0 0 7 1 】**

圧縮、インデックス付け、および順位付けモジュール210は、コレクション102内の文書の圧縮、インデックス付け、および順位付けを行うように、1つまたは複数の機械可読命令セットによりプログラムされる。圧縮、インデックス付け、および順位付けモジュール210は、1つまたは複数の地理的位置に配備されている1つまたは複数のコンピュータ上に実装することができる。

40

**【 0 0 7 2 】**

データセンター208は、電子文書コレクション102内の電子文書の特徴付ける情報を格納する。そのような電子文書の特徴付ける情報は、インデックス付きキーワードおよびキーワードを見つけることができるコレクション102内の文書の位置を含むインデックス付きデータベースの形態で格納されうる。インデックス付きデータベースは、例えば、クローラー204によって形成されうる。

**【 0 0 7 3 】**

50

いくつかの実装では、データセンター208内に格納されている情報はそれ自体、構造化プレゼンテーション106をユーザーに提示しやすいように編成されうる。例えば、情報は、いくつかのトピックに関連する構造化プレゼンテーション106を提示することが必要であることを予想してクローラー204ならびに圧縮、インデックス付け、および順位付けモジュール210によって編成されうる。データセンター208内の情報の構造は、構造化プレゼンテーション106内の情報のグループ分け、セグメント分割、および配列を行いやすくするものであってよい。この編成は、さまざまな異なるファクタに基づいて行うことができる。例えば、オントロジーを使用して、データセンター208内に格納されている情報を整理することができる。他の例として、前の構造化プレゼンテーション106の履歴レコードを使用して、データセンター208内に格納されている情報を整理することができる。他の例として、本明細書で説明されているデータテーブルを使用して、データセンター208内に格納されている情報を整理することができる。

10

**【0074】**

図示されているように、システム200は、機械可読命令に従って構造化プレゼンテーションを提示することができる複数のディスプレイ画面104を備える。ディスプレイ画面104としては、例えば、陰極線(CRT)、発光ダイオード(LED)画面、液晶ディスプレイ(LCD)、ガスプラズマディスプレイ、および同様のものが挙げられる。ディスプレイ画面104は、自給式データ処理システム、例えば、パーソナルデータアシスタント(PDA)215、デスクトップコンピュータ217、または携帯電話の欠くことのできない部分であるものとして行うことができる。一般に、構造化プレゼンテーションを提示するための命令は、そのような自給式データ処理システムが受け取った後にディスプレイ画面104の仕様に合わせて修正される。しかし、これは必ずしもそうではない。例えば、ディスプレイ画面104は、構造化プレゼンテーションを提示するための命令の処理がそれらの命令がディスプレイ画面104に届く前に完了している、より分散度の高いシステムの一部であってもよい。例えば、ディスプレイ画面104は、「ダム」デバイス、例えば、テレビ受像機またはコンピュータのモニタに組み込むことができ、これらはローカルもしくはリモートのソースから構造化プレゼンテーション106を提示するための命令を受け取る。

20

**【0075】**

動作中、システム200は、コレクション102内の非構造化情報を、閲覧者に提示される構造化プレゼンテーション106に変換することができる。このような変換は、検索エンジンがコレクション102内の電子文書から抽出された情報に基づいて情報要求を受け取り、その情報要求に応答するウェブ検索の状況において実行されうる。

30

**【0076】**

例えば、パーソナルデータアシスタント(PDA)215またはデスクトップコンピュータ217は、ユーザーと情報をやり取りし、これにより、例えば、ウェブブラウザアプリケーションを使って検索クエリを受け取ることができる。クエリの記述212は、無線データリンク219および/または有線データリンク221を介して検索エンジン202に伝送されうる。これに対する応答として、検索エンジン202は、クエリ記述212を使用して、ディスプレイ画面104上に構造化プレゼンテーション106を提示する際に使用できるデータセンター208内の情報を識別することができる。識別された情報は、非構造化電子文書コレクション102内の2つまたはそれ以上の指定されていない電子文書から引き出すことができる。いくつかの場合において、クエリ記述212は、情報を取り出して構造化プレゼンテーション106をユーザーに提示するために検索エンジン202によって使用される検索語を含むことができる。例えば、クエリ記述212内の検索語を使用することで、データセンター208内において、関係するインスタンス、そのようなインスタンスを特徴付ける属性、個別のインスタンスを特徴付ける値、および/または構造化プレゼンテーション106の他の態様からなるコレクションを識別することができる。

40

**【0077】**

検索エンジン202は、クエリ記述212への応答214を生成することもできる。応答214は、ユーザーに対し構造化プレゼンテーション106を提示するために使用されうる。一般に、

50

応答214は、構造化プレゼンテーション106を提示するためにシステム215、217内のデータ処理デバイスによって解釈できる機械可読命令を含む。例えば、応答214は、構造化プレゼンテーション106の特性およびコンテンツを指定するためにHTMLでコード化することができる。他の実装では、応答214は、構造化プレゼンテーション106を提示するために使用されるデータセンター208からのテキストスニペットまたは他の情報を含むことができる。例えば、応答214は、以下でさらに詳しく説明するように、値、新規属性の名前、構造化プレゼンテーション106で表示する値が正しい確率の推定からなるコレクションを含むことができる。

**【 0 0 7 8 】**

多くの場合において、システム200は、データセンター208内に格納されている情報を使用して、クエリ記述212において記述されているクエリに関連する1つまたは複数の文書の位置を識別する。例えば、検索エンジン202は、クエリ記述212内のキーワードをデータセンター208内に格納されているキーワードのインデックスと比較することができる。この比較結果を使用して、クエリ記述212に関連するコレクション102内の文書を識別することができる。そのような識別された文書の位置は、応答214内に、例えば、記述されたクエリに回答する文書へのハイパーリンクとして含めることができる。

**【 0 0 7 9 】**

いくつかの実装では、システム200は、構造化プレゼンテーション106内の情報のグループ分け、セグメント分割、および配列を行いやすいように属性および/またはそれらの各値を格納することができる。例えば、システム215、217などのクライアントシステムをユーザーがインタラクティブに操作することによって修正し、変更するとき、インスタンス、それらの属性、およびそれらの値からなるコレクションを構造化プレゼンテーション106としてデータセンター208内に格納することができる。例えば、第1の閲覧者に提示される一方の構造化プレゼンテーション106内のインスタンス、属性、および値をデータセンター208内に格納し、他の閲覧者にその後の構造化プレゼンテーション106を提供する際に使用することができる。

**【 0 0 8 0 】**

図3は、例示的な構造化プレゼンテーション106、つまり、テーブル300を含むものの略図である。テーブル300は、インスタンスの1つまたは複数の識別子、さらにはそれらのインスタンスの特定の属性の値の整理されたシステム配置である。インスタンスは、個別に識別可能なエンティティであり、一般的に、少なくともいくつかの共通の属性を共有する。属性は、エンティティのプロパティ、特徴、または特性である。例えば、Tom、Dick、およびHarryは、個人のインスタンスである。それぞれのそのような個人は、名前、身長、体重、および同様のものなどの属性を有する。他の例として、都市インスタンスはそれぞれ、地理的位置、市長、および人口を有する。さらに他の例として、製品インスタンスは、モデル名、製造会社、および年を有するものとしてよい。

**【 0 0 8 1 】**

インスタンスの属性は、値によって特徴付けることができる。したがって、特定のインスタンスの特定の属性の値は、その特定のインスタンスを特徴付ける。例えば、個人の名前は「Tom」という値を有し、市の人口は値「400万」を有し、製品のモデル名は値「Wrangler」を有することができる。いくつかの実装では、テーブル300などの構造化プレゼンテーションは、属性の識別子、さらには値を表す際に使用する単位の識別子も含むことができる。

**【 0 0 8 2 】**

テーブル300内の情報のグループ分け、セグメント分割、および配列は、ユーザーによって情報を理解しやすいように選択されうる。この点で、テーブル300は、行302のコレクションを含む。それぞれの行302は、インスタンス識別子306と関連付けられている属性値307のコレクションとを含む。したがって、行302内の属性値307およびインスタンス識別子306の配列および位置決めは、それらの間の関連付けをグラフィックで表す。例えば、ユーザーは、属性値307と同じ行302内にあるインスタンス識別子306との間の関連付けを

10

20

30

40

50

認識することができる。

【0083】

テーブル300は、列304のコレクションも含む。それぞれの列304は、属性識別子308と関連付けられている属性値307のコレクションとを含む。したがって、列304内の属性値307および属性識別子308の配列および位置決めは、それらの間の関連付けをグラフィックで表す。例えば、ユーザーは、その連携に基づいて属性値307と同じ列304内にある属性識別子308との間の関連付けを認識することができる。

【0084】

それぞれの行302は、それぞれの行302が単一のインスタンス識別子306を関連付けられている属性値307のコレクションに関連付けるという点で構造化レコード310である。さらに、一方の構造化レコード310内のそれらの関連付けを表すために使用される配列および位置決めは、他方の構造化レコード310内(つまり、他の行302内)に複製される。実際、多くの場合において、構造化プレゼンテーション106内の構造化レコード310はすべて、情報の同じ配列および位置決めを有することに制限される。例えば、属性「ATTR\_2」の値307は、すべての行302における同じ列304に出現することに制限される。他の例として、属性識別子308はすべて、同じ列304内に出現する値307と同じ空間的関係を有する。さらに、一方の構造化レコード310内の情報の配列および位置決めに加えられた変更は、一般的に、構造化プレゼンテーション106内の他の構造化レコード310に伝搬する。例えば、新規属性(例えば、「ATTR\_2<sup>3</sup>/<sub>4</sub>」)を特徴付ける新規属性値307が一方の構造化レコード310に追加される場合、すべてのインスタンスの属性「ATTR\_2<sup>3</sup>/<sub>4</sub>」の値が構造化プレゼンテーション106に追加できるように新規列304が構造化プレゼンテーション106に追加される。

【0085】

いくつかの実装では、テーブル300内の値307は、いくつかの測定単位で提示することができる。測定単位の例としては、フィート、ヤード、インチ、マイル、秒、ガロン、リットル、および同様の単位が挙げられる。いくつかの場合に、値307を提示する際に使用する測定単位は、単位識別子309で示される。単位識別子309は、例えば、値307の横に、および/または関連する属性識別子308の横に表示されうる。単位識別子309と測定単位が示される値307との間の関連付けは、そのような配置によって閲覧者に対し示される。多くの場合において、単一の属性に関連付けられている値307のすべて(例えば、単一の列304内の値307のすべて)が、同じ測定単位での提示に制限される。

【0086】

システム100、200によって電子文書コレクション102から抽出された情報は、ユーザーへのテーブル300の提示にさまざまな形で影響を及ぼしうる。例えば、電子文書コレクション102から抽出された情報を使用して、テーブル300に書き込む値307を決定することができる。他の例として、電子文書コレクション102から抽出された情報を使用することにより、テーブル300に追加する新規属性および/または新規インスタンスを提案することができる。

【0087】

いくつかの実装では、インスタンス識別子306を1つまたは複数の検索文字列に基づいて選択することができる。例えば、検索エンジン202がユーザーから検索文字列「hybrid vehicles」を受け取ると、システム200などのシステムが、検索文字列を使用して電子文書コレクション102から抽出された情報に基づいてテーブル300を生成し、テーブル300に書き込むことができる。例えば、システム200は、データセンター208にアクセスし、検索文字列に関連する電子文書内のインスタンス識別子306を識別し、識別されたインスタンスに対する一組の共通属性(さらにはそれらの属性の識別子308およびそれらの属性に対する値307)を決定することができる。実際、システム200は、受け取った検索文字列に基づいてインスタンス識別子306、属性識別子308、さらには関連付けられている値307を決定することができる。

【0088】

いくつかの実装では、1つまたは複数の属性識別子308、インスタンス識別子306、およ

10

20

30

40

50

び/または値307を、テーブル300の表示を見るユーザーから受け取ることができる。以下でさらに詳しく説明するように、システム200などのシステムは、1つまたは複数の受け取った属性識別子308、インスタンス識別子306、および/または値307を使用し電子文書コレクション102から抽出された情報に基づいてテーブル300を生成し、テーブル300に書き込むことができる。実際、システム200は、受け取った属性識別子308、インスタンス識別子306、および/または値307に基づいて新規インスタンス識別子306、属性識別子308、さらには関連付けられている値307を構成することができる。

#### 【0089】

図4は、構造化プレゼンテーション、つまり、テーブル400を含むものの他の実装の略図である。行302および列304内に編成されている属性識別子308、インスタンス識別子306、値307、単位識別子309を含むことに加えて、テーブル400は、ユーザーが情報をやり取りするための多数のインタラクティブ要素も含む。特に、テーブル400は、インスタンス選択ウィジェット(instance selection widgets)405のコレクション、アクショントリガー(action triggers)410のコレクション、列アクショントリガーウィジェット(column action trigger widgets)415のコレクション、および注釈列(notes column)420を含む。

#### 【0090】

インスタンス選択ウィジェット405は、ユーザーがテーブル400内の構造化レコード310を選択することを可能にするユーザーインターフェイスコンポーネントである。例えば、インスタンス選択ウィジェット405は、その構造化レコード310に関係する配列および位置決めを用いて特定の構造化レコード310に関連付けられているクリック可能なチェックボックスのコレクションであってよい。インスタンス選択ウィジェット405は、ユーザーがマウス(例えば、コンポーネントの上でホバリングし、特定のマウスボタンをクリックする)、スタイラス(例えば、スタイラスでタッチスクリーン上に表示されているユーザーインターフェイスコンポーネントを押す)、キーボード、またはそのコンポーネントが備える機能と呼び出す他の入力デバイスを使用してウィジェット405をインタラクティブに操作することができるという点で「クリック可能」である。

#### 【0091】

アクショントリガー410は、ユーザーがインスタンス選択ウィジェット405を使用して選択されたテーブル400内の1つまたは複数の構造化レコード310に対するアクションの実行をトリガーすることを可能にするユーザーインターフェイスコンポーネントである。例えば、アクショントリガー410はクリック可能なテキストフレーズとすることができ、それぞれ、フレーズ内に記述されたアクションをトリガーするためにユーザーが使用できる。例えば、「keep and remove others」アクショントリガー410は、テーブル400の表示からのインスタンス選択ウィジェット405を使用して選択されていない構造化レコード310の削除のトリガーとなる。他の例として、「remove selected」アクショントリガー410は、テーブル400の表示からのインスタンス選択ウィジェット405を使用して選択されている構造化レコード310の削除のトリガーとなる。さらに他の例として、「show on map」アクショントリガー410は、地図上でインスタンス選択ウィジェット405を使用して選択されている構造化レコード310の位置の表示のトリガーとなる。例えば、選択されたインスタンスが自動車である場合、選択された自動車を販売しているカーディーラーの所在地が、地図上に表示されうる。他の例として、選択されたインスタンスが春休みの旅行先である場合、それらの旅行先を地図上に表示することができる。

#### 【0092】

列アクショントリガーウィジェット415は、ユーザーが単一の列304内のセルのすべてにアクションを適用することを可能にするユーザーインターフェイスコンポーネントである。ユーザーがクリック可能な「+」記号をインタラクティブに操作したときに、他のユーザーインターフェイスコンポーネントが表示され、ユーザーに、実行すべき可能な一組のアクションが与えられる。この一組のアクションとしては、例えば、構造化プレゼンテーション400から列304全体を削除する段階、または現在ブランクになっている列304内のすべてのセルに対する値を見つける検索が挙げられる。

10

20

30

40

50

## 【 0 0 9 3 】

注釈列420は、ユーザーが情報をインスタンス識別子306に関連付けることを可能にするユーザーインターフェイスコンポーネントである。特に、注釈列420は、その構造化レコード310に關係する配列および位置決めを用いて構造化レコード310にそれぞれ関連付けられている1つまたは複数の注釈425を含む。注釈425の情報内容は、列304とは異なり、注釈425が特定の属性の値であると主張されないという点で無制限である。その代わりに、注釈425に含まれる情報は、構造化レコード310で識別されているインスタンスの無關係の態様を特徴付けることができる。

## 【 0 0 9 4 】

いくつかの実装では、テーブル400は、特定の属性の値以外の追加情報を含みうる。例えば、テーブル400は、その構造化レコード310に關係する配列および位置決めを用いて構造化レコード310において識別されているインスタンスに関連付けられているイメージ430のコレクションを含むことができる。他の例として、テーブル400は、コレクション102内の電子文書から抽出されたテキストスニペット435のコレクションを含むことができる。スニペットのソースは、インスタンス識別子306を検索文字列として使用して実行される検索の高順位の結果としてよい。テキストスニペット435は、その構造化レコード310に關係する配列および位置決めを用いて構造化レコード310において識別されているインスタンスに関連付けられる。

## 【 0 0 9 5 】

他の例として、テーブル400は、コレクション102内の個別の電子文書への1つまたは複数のハイパーテキストリンク440を含むことができる。例えば、リンク付き文書は、インスタンス識別子306を検索文字列として使用して実行される検索の高順位の結果としてよい。他の例として、リンク付き文書は、テーブル400に書き込むために抽出された値307のソースであってよい。いくつかの場合において、ハイパーテキストリンク440のインタラクティブな操作によって、ハイパーテキスト440に埋め込まれた情報(例えば、ウェブサイトアドレス)に基づいてソースの電子文書へのナビゲーションがトリガーされうる。

## 【 0 0 9 6 】

図5は、構造化プレゼンテーション、つまり、カードのコレクション500を含むもの他の実装の略図である。カードコレクション500は、インスタンスの1つまたは複数の識別子、さらにはそれらのインスタンスの特定の属性の値の整理されたシステム配置である。インスタンスの属性は、値で指定することができる。さらには、カードコレクション500は、一般的に、属性の識別子、さらには適宜、値を表す際に使用する単位の識別子も含む。

## 【 0 0 9 7 】

カードコレクション500内の情報のグループ分け、セグメント分割、および配列は、ユーザーによって情報を理解しやすいように選択されうる。この点で、カードコレクション500は、カード502のコレクションを含む。それぞれのカード502は、インスタンス識別子306と関連付けられている属性値307のコレクションとを含む。したがって、カード502内の属性値307およびインスタンス識別子306の配列および位置決めは、それらの間の関連付けをグラフィックで表す。例えば、ユーザーは、属性値307と同じカード502内にあるインスタンス識別子306との間の関連付けを認識することができる。

## 【 0 0 9 8 】

例示されている実装において、カードコレクション500内のカード502は、属性識別子308のコレクションも含む。属性識別子308は、列504内に編成され、属性値307は、列506内に編成される。列504、506は、互いに隣接する位置に置かれ、個別の属性識別子308がその識別された属性を特徴付ける属性値307の隣に配置されるように整列される。この位置決めおよび配列を使用することで、閲覧者は属性識別子308とそれらの属性を特徴付ける属性値307との間の関連付けを認識することができる。

## 【 0 0 9 9 】

それぞれのカード502は、それぞれのカード502が単一のインスタンス識別子306を関連付けられている属性値307のコレクションに関連付けるという点で構造化レコード310であ

10

20

30

40

50

る。さらに、一方のカード502内のそれらの関連付けを表すために使用される配列および位置決めは、他方のカード502内に複製される。実際、多くの場合において、カード502のすべてが、情報の同じ配列および位置決めを有することに制限される。例えば、属性「ATTR\_1」を特徴付ける値307は、すべてのカード502においてインスタンス識別子306との空間的關係が同じになるように制限される。他の例として、カード502のすべてにおける属性識別子308の順序および位置決めは同じである。

#### 【0100】

さらに、一方のカード502内の情報の配列および位置決めに加えられた変更は、一般的に、カードコレクション500内の他のカード502に伝搬する。例えば、新規属性(例えば、「ATTR\_1 3/4」)を特徴付ける新規属性値307が一方のカード502内の属性値「value\_1\_1」と「value\_2\_1」との間に挿入された場合、他方のカード502内の対応する属性値307の位置決めも同様に変更される。

10

#### 【0101】

いくつかの実装では、カードコレクション500内のカード502は、他の特徴を備えることもできる。例えば、カード502は、ユーザーと情報をやり取りするためのインタラクティブ要素、例えば、インスタンス選択ウィジェット、アクショントリガー、属性選択ウィジェット、注釈入力機能、および同様のものを備えることができる。他の例として、カードコレクション500内のカード502は、特定の属性の値以外の追加の情報、例えば、イメージおよび/または識別されたインスタンスに関連付けられているテキストニペットを含むことができる。他の例として、カードコレクション500内のカード502は、コレクション102内の個別の電子文書への1つまたは複数のハイパーテキストリンクを含むことができる。そのような特徴は、そのインスタンスを識別するインスタンス識別子306を含むカード502上に表示することによって特定のインスタンスに関連付けられうる。

20

#### 【0102】

動作中、閲覧者は、カードコレクション500を提示するシステムをインタラクティブに操作して1つまたは複数のカード502の表示を変更することができる。例えば、閲覧者は、それらのカード上で識別されている特定のインスタンスの比較がしやすいようにカード502のうちの2つまたはそれ以上のカードの隣り合わせで並べた表示をトリガーすることができる。他の例として、閲覧者は、カード502の並べ替え、特定のカード502の表示の終了、または同様の操作をトリガーすることができる。他の例として、閲覧者は、カード502内に表示される属性および/またはインスタンスの選択、変更、追加、および/または削除をトリガーすることができる。さらに別の例として、閲覧者は、例えばカード内の属性値307の値に応じてカードをソートして複数の山に分ける操作をトリガーすることができる。

30

#### 【0103】

いくつかの実装では、カード502は、「両面」で表示される。例えば、第1の側は、インスタンス識別子306によって識別されるインスタンスのグラフィック表現を含み、第2の側は、インスタンス識別子306および値307を含むことができる。これは、例えば、ユーザーがカードのコレクション500内の特定のカードを検索する場合に有用であり、ユーザーはカード502の第1の側にあるグラフィック表現をざっと見るだけで特定のカードを識別することができる。

40

#### 【0104】

図6は、電子文書コレクションからの情報を構造化プレゼンテーションでユーザーに提示するための例示的なプロセス600の流れ図である。プロセス600は、1つまたは複数の機械可読命令セットを実行することによってオペレーションを実行する1つまたは複数のコンピュータによって実行されうる。例えば、プロセス600は、システム200内の検索エンジン202によって実行されうる。いくつかの実装では、プロセス600は、トリガー、例えば、ユーザー要求を受け取ったことに応答して実行され、これにより、構造化プレゼンテーションを作成または変更することができる。

#### 【0105】

プロセス600を実行するシステムは、電子文書コレクション内の2つまたはそれ以上の応

50

答電子文書(responsive electronic documents)を識別することができる(ステップ605)。応答文書は、多くの異なる方法で識別されうる。いくつかの場合において、文書は、閲覧者から受け取った「新規」情報(例えば、新規検索クエリ)に基づいて識別される。例えば、システムは、新たに受け取った検索クエリを、文字列比較を用いて電子文書コレクション内の電子文書のコンテンツと比較することができる。他の例として、システムは、データセンター208などのデータセンターにアクセスして、検索クエリ内の検索語をキーワードのインデックスと比較し、応答電子文書の位置を識別することができる。

**【0106】**

いくつかの場合において、文書は、構造化プレゼンテーション内にすでに見ついている「旧」情報に基づいて識別される。構造化プレゼンテーション内に見つかる情報には、インスタンス、属性、値、および値を表す単位の識別が含まれる。プロセス600を実行するシステムは、この旧情報を使用して電子文書コレクション内の応答電子文書を識別することができる。例えば、構造化プレゼンテーション内にすでに見ついているインスタンスを含む文書は、応答文書であるとして識別されうる。他の例として、構造化プレゼンテーション内にすでに見ついている属性を使用してインスタンスを特徴付ける文書は、応答文書であるとして識別されうる。このような識別の他の例を以下で説明する。

10

**【0107】**

プロセス600を実行するシステムは、識別された電子文書から情報を収集することもできる(ステップ610)。収集された情報は、1つまたは複数のインスタンス、属性、および/または値に関するものとしてよい。プロセス600を実行するシステムは、この情報を電子文書コレクション内の文書から直接的に、または電子文書コレクション内の電子文書を特徴付ける情報のすでに集められているコレクションから収集することができる。例えば、システム200(図2)の背景状況において、プロセス600を実行するシステムは、コレクション102内の文書を特定し、特定された文書にアクセスし、コレクション102内の元の文書から直接的に情報を抽出することができる。システム200(図2)の背景状況における他の例として、プロセス600を実行するシステムは、データセンター208内の情報のコレクションにアクセスし、例えばキーワードのインデックスおよびそれらのキーワードを含む文書の場所、オントロジー、および/またはコレクション102内の文書から抽出された情報を使用して提示された以前の構造化プレゼンテーションの履歴レコードを格納しているデータベースから情報を収集することができる。

20

30

**【0108】**

プロセス600を実行するシステムは、収集された情報を使用し、その収集された情報に基づいて構造化プレゼンテーションを提示するための命令を供給することができる(ステップ615)。例えば、プロセス600を実行するシステムは、構造化プレゼンテーション、例えば、テーブル300、400またはカードのコレクション500を提示するための機械可読命令を生成することができる。

**【0109】**

図7は、電子文書コレクション内の応答文書を識別するための例示的なプロセス700の流れ図である。プロセス700は、他のデータ処理活動から独立して、または連携して実行されうる。例えば、プロセス700は、プロセス600におけるステップ605で実行されうる(図6)。

40

**【0110】**

プロセス700を実行するシステムは、検索クエリを受け取る(ステップ705)。例えば、システムは、ユーザーから1つまたは複数の検索文字列(例えば、「hybrid vehicles」)を受け取ることができる。他の例として、システムは、他のプロセスもしくはシステムから検索文字列を受け取ることができる。いくつかの実装では、アプリケーションプログラミングインターフェイス(API)、共通ゲートウェイインターフェイス(CGI)スクリプト、または他のプログラミングインターフェイスを通じて検索文字列を受け取る。他の実装では、ウェブポータル、ウェブページ、もしくはウェブサイト、または同様のものを通じて検索文字列を受け取る。

50

## 【0111】

それに応答して、プロセス700を実行するシステムは、検索クエリに応答するインスタンス、属性、および/または値を含む2つまたはそれ以上の文書を識別する(ステップ710)。これらの文書は、検索クエリ内の検索語が構造化プレゼンテーションにおいて果たすべき役割を分類することによって識別されうる。例えば、検索クエリ内の検索語は、例えば、検索クエリ内の特定の検索語、検索クエリの検索語の分類の仕方のユーザーによる明示的な指示、および/または検索の背景状況に基づいて構造化プレゼンテーション内に出現すべきインスタンスのカテゴリ化として分類されうる。例えば、検索クエリ「cities in California」内の検索語は、複数形の語「cities」が属性によって特徴付けられている、つまり、「in California」で修飾されているため「San Diego」、「Los Angeles」、および「Bakersfield」などのインスタンスのカテゴリ化として分類することができる。他の例として、検索クエリ「Ivy League schools」内の検索語は、複数形の語「schools」が属性「Ivy League」によって特徴付けられているためインスタンス(「Cornell」、「Columbia」、および「Brown」など)のカテゴリ化として分類することができる。

10

## 【0112】

いくつかの場合において、検索クエリ内の検索語を分類するために追加の情報を使用しなければならない。例えば、検索クエリ「Ivy League」は、「school」のインスタンスのカテゴリ化として、または「Atlantic Coast Conference」および「PAC-10」などのインスタンスを含むカテゴリ「athletic conferences」の例示的なインスタンスとしてみても妥当である。このような場合、これらの検索語は、例えば、分類の仕方に関するユーザーによる明示的な指示または検索セッションにおける検索語の背景状況に基づいて分類することができる。例えば、ユーザーがフレーズ「Atlantic Coast Conference」および「PAC-10」を検索クエリとして以前に入力したことがある場合、検索クエリ「Ivy League」は、他のインスタンスとともに構造化プレゼンテーション内に出現する例示的なインスタンスとみなすことができる。

20

## 【0113】

これらの文書は、電子文書コレクション102において直接的に、または電子データセンター208内の情報に基づいて間接的に識別されうる。このような識別情報は、例えば、クロールして最後に見つかった文書のURLを含むことができる。

## 【0114】

図8は、電子文書コレクション内の2つまたはそれ以上の応答文書を識別するための例示的なプロセス800の他の流れ図である。プロセス800は、他のデータ処理活動から独立して、または連携して実行されうる。例えば、プロセス800は、プロセス600(図6)におけるステップ605で実行されうる。他の例として、プロセス800は、プロセス600(図6)におけるステップ605でプロセス700と連携して実行されうる。例えば、プロセス700、800は、反復的なインタラクティブプロセスの一部とすることができ、このプロセスにおいて検索クエリを受け取り、これを使用して応答文書の第1のコレクションを識別し、識別された文書から引き出されたコンテンツを含む第1の構造化プレゼンテーションをユーザーに提示し、ユーザー修正を受け取り、修正された構造化プレゼンテーションの記述を使用して関連する文書の第2のコレクションを識別する。いくつかの実装では、プロセス800は複数回実行できる。いくつかの実装では、プロセス800は、例えばシステム200(図2)におけるクローラー204によって、ユーザー入力なしで実行することができる。

30

40

## 【0115】

プロセス800を実行するシステムは、構造化プレゼンテーションの既存のコンテンツの記述を受け取る(ステップ805)。特に、システムは、インスタンス、属性、値、および/または値を既存の構造化プレゼンテーションにおいて提示する単位の記述を受け取ることができる。記述は、例えば、インスタンスおよび属性の識別子および/または属性の値の範囲を含むことができる。記述は、インスタンスおよび/または属性のカテゴリ化も含むことができる。このようなカテゴリ化は、例えば、オントロジーを使用して、または閲覧者によって構造化プレゼンテーションに割り当てられたカテゴリ化に基づいて決定されうる

50

。例えば、ユーザーが構造化プレゼンテーションに「Ivy League Schools」とタイトルを付けた場合、このタイトルは、その構造化プレゼンテーション内のインスタンスのカテゴリ化とみなせる。

【0116】

それに応答して、プロセス800を実行するシステムは、既存のコンテンツに関連するインスタンス、属性、および/または値を含む1つまたは複数の文書を識別することができる(ステップ810)。例えば、システムは、インスタンスおよび/または属性の識別子をインデックス付きキーワードと比較して、特定の文書が構造化プレゼンテーションの既存のコンテンツ内にすでに出現しているインスタンスおよび/または属性のうちの1つまたは複数のインスタンスおよび/または属性を含むかどうかを判定することができる。他の例として、システムは、新規インスタンス、その属性、およびそのような属性の値をそのような文書から識別し、それらの値を構造化プレゼンテーションの既存のコンテンツ内にすでに出現している値と比較し、新規インスタンスが構造化プレゼンテーションの既存のコンテンツに潜在的に関連しているかどうかを判定することができる。

10

【0117】

これらの文書は、電子文書コレクション102において直接的に、または電子データセンター208内の識別情報を使用して識別されうる。このような識別情報は、例えば、クロールリングして最後に見つかった文書のメモリロケーションを含むことができる。

【0118】

図9は、新規インスタンスを構造化プレゼンテーションに提案および/または追加するためのプロセス900の流れ図である。プロセス900は、1つまたは複数の機械可読命令セットを実行することによってオペレーションを実行する1つまたは複数のコンピュータによって実行されうる。これらのデジタルデータ処理デバイスは、キーボード、マウス、タッチスクリーン、ディスプレイ画面、および同様のものなどの入力および出力デバイス上でユーザーと情報をやり取りすることができる。例えば、システム200(図2)の背景状況において、プロセス900におけるユーザーインタラクションをPDA215またはデスクトップコンピュータ217などのクライアント側で実行できる。

20

【0119】

プロセス900は、単独で、または他のデータ処理活動と連携して実行されうる。例えば、以下でさらに詳しく説明するように、プロセス900は、事前に存在している構造化プレゼンテーションに加えてインスタンス提案を構成するためにさまざまなプロセスと連携して実行されうる。このような構成プロセスの例は、図21~26および関連する本文において説明されている。一般に、プロセス900は、複数のデジタルデータ処理デバイスによって実行される。例えば、システム200(図2)の背景状況において、インスタンス提案を構成するための活動を検索エンジン202側で実行することができ、その一方で、ユーザーインタラクションはPDA215またはデスクトップコンピュータ217(図2)などのクライアント側で実行されうる。

30

【0120】

プロセス900を実行するシステムは、新規インスタンストリガーを受け取ることができる(ステップ905)。新規インスタンスは、構造化プレゼンテーション、例えば、構造化プレゼンテーション106(図1)に現在表示されていないインスタンスである。新規インスタンストリガーは、新規インスタンスを構造化プレゼンテーションに追加するためのプロセスをアクティベートする。例えば、新規インスタンスは、マウス、スタイラス、キーボード、または同様のものを介して受け取ったユーザー入力によってトリガーされうる。他の実装では、新規インスタンスは、他のプロセスまたはシステムによってトリガーされうる。新規インスタンストリガーは、2つほど例を挙げると、プロセス間通信またはアプリケーションのメッセージハンドラを通じて受け取ることができる。

40

【0121】

プロセス900を実行するシステムは、ユーザーに対して、新規インスタンスを構造化プレゼンテーションに追加するためのオプションを提示することができる(ステップ910)。

50

オプションは、新規インスタンスを追加するための代替アプローチである。オプションの例として、完全自動オプション、ユーザー指定制約条件を含む自動オプション、および手動オプションが挙げられる。これらのオプションについて、以下でさらに詳しく説明する。

#### 【0122】

プロセス900を実行するシステムは、ユーザーに対して、ユーザーインターフェイスデバイス、例えば、ディスプレイ画面を使用するオプションを提示することができる。多くの場合において、オプションを提示するディスプレイ画面は、インスタンスを追加する構造化プレゼンテーションを提示する同じディスプレイ画面であってもよい。例えば、オプションは、ディスプレイ画面104(図1)を使用してユーザーに提示することができる。

10

#### 【0123】

プロセス900を実行するシステムは、オプションのユーザー選択を受け取ることができる(ステップ915)。ユーザー選択は、1つまたは複数の入力デバイス、例えば、キーボード、タッチパッド、またはタッチスクリーンを使用して受け取ることができる。システムは、ユーザーによって選択されたオプションの性質を判定することもできる(ステップ920)。

#### 【0124】

プロセス900を実行するシステムが、ユーザーが「自動オプション」を選択したと判定した場合、システムは、追加のインスタンスを提案し、および/またはユーザーとの情報のやり取りなしで、その追加のインスタンスを構造化プレゼンテーションに追加することができる。

20

#### 【0125】

ユーザー指定の自動オプションの一実装では、構造化プレゼンテーションの特性に基づいて新規インスタンスを提案し、および/または追加することができる(ステップ925)。そのような特性の例として、構造化プレゼンテーションにおいてすでに指定されているインスタンスの性質、それらのインスタンスのカテゴリ化、およびそれらのインスタンスの属性が挙げられる。このような特性に基づいて新規インスタンスを構成するためのアプローチは、図21~26および関連する本文において説明されている。例えば、そこで説明されているように、検索クエリは、事前に存在している構造化プレゼンテーションから引き出された属性識別子、事前に存在している構造化プレゼンテーションから引き出された属性値、および/またはそれらの組み合わせを使用して構築することができる。これらの検索クエリを使用し、文字列比較結果または他のマッチング技術を使用して構造化プレゼンテーションに追加するインスタンスを識別することができる。

30

#### 【0126】

プロセス900を実行するシステムが、ユーザーが「ユーザー指定制約条件」オプションを選択したと判定した場合、システムは、追加のインスタンスの性質に対するユーザー指定制約条件に基づいて追加のインスタンスを提案し、および/または自動的にその追加のインスタンスを構造化プレゼンテーションに追加することができる。これらの制約条件は、提案された、および/または追加されたインスタンスを特徴付ける1つまたは複数のパラメータとして表すことができる。例えば、これらの制約条件は、インスタンスの属性の許容可能な値として、または属性の許容可能な値の範囲として表すことができる。

40

#### 【0127】

ユーザー指定制約条件オプションの一実装では、プロセス900を実行するシステムは、新規インスタンスの属性の値を制約するためのオプションをユーザーに対して提示する(ステップ930)。例えば、システムは、構造化プレゼンテーション内のインスタンスを特徴付ける属性のリスト、さらにはユーザーがそれらの属性の値に対する制約条件を入力することを可能にする入力フィールドを表示することができる。多くの場合、そのようなリスト内の属性は、新規インスタンスの追加先の構造化プレゼンテーション内にも出現する。しかし、いくつかの実装では、そのようなリスト内の属性は、電子文書コレクションの文書内など、他の場所にあるインスタンスを特徴付けるために使用される属性に基づいて構

50

成することができる。このような属性を構成するための例示的なアプローチは、図37～51および関連する本文において説明されている。

【0128】

プロセス900を実行するシステムは、新規インスタンスの属性の値に対する1つまたは複数の制約条件のユーザー指定を受け取ることもできる(ステップ935)。上で説明されているように、これらの制約条件は、1つまたは複数の属性の値を特定の値もしくは値の範囲に制限することができる。例えば、自動車の特徴付ける1つの属性は、「気筒数」である。この属性の値のユーザー指定制約条件は、新規自動車インスタンスの気筒数を特定の値(例えば、「6」)または値の範囲(例えば、「6から8」または「6超」)に制限することができる。

10

【0129】

プロセス900を実行するシステムは、ユーザー指定制約条件および構造化プレゼンテーションの特性に基づいて新規インスタンスを提案し、および/または追加することもできる(ステップ940)。構造化プレゼンテーションの特性の例として、構造化プレゼンテーションにおいてすでに指定されているインスタンスの性質、それらのインスタンスのカテゴリ化、およびそれらのインスタンスの属性が挙げられる。このような特性に基づいて新規インスタンスを構成するためのアプローチは、図37～51および関連する本文において説明されている。他の例として、検索クエリは、事前に存在している構造化プレゼンテーションから引き出された属性識別子、事前に存在している構造化プレゼンテーションから引き出された属性値、および/またはそれらの組み合わせ、さらにはユーザーによって指定された制約条件を使用して構築することができる。これらの検索クエリを使用し、文字列比較結果または他のマッチング技術を使用してインスタンスを識別することができる。次いで、識別されたインスタンスが提案され、および/または構造化プレゼンテーションに追加されうる。

20

【0130】

プロセス900を実行するシステムが、ユーザーが「手動オプション」を選択したと判定した場合、システムは、ユーザーの指示に従って追加のインスタンスを構造化プレゼンテーションに追加することができる。

【0131】

手動オプションの一実装では、プロセス900を実行するシステムは、ユーザーから新規インスタンスを受け取ることができる(ステップ945)。例えば、ユーザーは、キーボードまたは他のユーザー入力デバイスを使用してインスタンス名を入力することができる。プロセス900を実行するシステムは、新規インスタンスを構造化プレゼンテーションに追加することができる(ステップ950)。一般に、新規インスタンスの名前は、新規構造化レコード310内のインスタンス識別子306として構造化プレゼンテーションに直接追加することができる。いくつかの実装では、新規構造化レコード310は、新規行302(図3、4)または新規カード502(図5)とすることができる。

30

【0132】

いくつかの実施形態では、プロセス900を実行するシステムは、受け取った新規インスタンスに基づいて追加のオペレーションを実行することもできる。例えば、このシステムは、新規インスタンスを使用して、その一組の提案されたインスタンスまたは一組の提案された属性を改善することができる。

40

【0133】

図10は、構造化プレゼンテーションの修正を指定するユーザー入力を受け取るためのユーザーインターフェイスコンポーネント1000の略図である。例えば、プロセス900(図9)のステップ905において、ユーザーインターフェイスコンポーネント1000を使用して、新規インスタストリガーを受け取ることができる。

【0134】

ユーザーインターフェイスコンポーネント1000は、属性修正領域1005およびインスタンス修正領域1010を備える。属性修正領域1005は、ヘッダ1015、それぞれが属性識別子選択

50

ウィジェット1030に関連付けられている属性識別子1025のコレクション1020、および新規属性追加トリガー1035を備える。

【0135】

ヘッダ1015は、属性修正領域1005とのユーザーインタラクションがユーザーが属性を修正することを実際に可能にすることを識別するテキストまたは他の情報を含む。属性識別子1025は、構造化プレゼンテーションに含まれる属性を識別するテキストまたは他の情報である。例えば、属性識別子1025は、構造化プレゼンテーション300、400、500(図3、4、5)内に属性識別子308として出現する同じテキストであってよい。属性識別子選択ウィジェット1030は、ユーザーが構造化プレゼンテーションで表示するための属性の選択および選択解除を行うことを可能にするインタラクティブ表示要素である。例えば、コレクション1020において、それぞれの属性識別子選択ウィジェット1030は、互いに隣接するその配列および位置決めを用いて単一の属性識別子1025に関連付けられる。属性識別子選択ウィジェット1030は、1つまたは複数のグラフィック表示、例えば、示されているチェックマークおよび色あいを使用して属性識別子1025を表示のために選択または選択解除するかどうかを示すことができる。例えば、ユーザーが属性識別子1025「Attribute\_1」に関連付けられているチェックされた属性識別子選択ウィジェット1030をインタラクティブに操作する場合、属性識別子選択ウィジェット1030内の色およびチェックされたステータスが変更され、構造化プレゼンテーションからの「Attribute\_1」に関連付けられている属性識別子(さらには「Attribute\_1」に対応する値とともに)の削除がトリガーされる。

【0136】

新規属性追加トリガー1035は、構造化プレゼンテーションへの新規属性の追加をトリガーするためにユーザーが使用することができるインタラクティブ表示要素である。追加する新規属性の構成方法は、図37~51および関連する本文において説明されている。新規属性の追加も、以下で、例えば、図13~15において、さらに詳しく説明される。

【0137】

インスタンス修正領域1010は、新規インスタンス追加トリガー1040およびインスタンスフィルタートリガー1045を備える。新規インスタンス追加トリガー1040は、構造化プレゼンテーションへの新規インスタンスの追加をトリガーするためにユーザーが使用することができるインタラクティブ表示要素である。例えば、新規インスタンス追加トリガー1040は、プロセス900(図9)におけるステップ905で使用されうる。

【0138】

インスタンスフィルタートリガー1045は、構造化プレゼンテーション内のインスタンスのフィルター処理をトリガーするためにユーザーが使用できるインタラクティブ表示要素である。インスタンスのフィルター処理により、1つまたは複数の基準を満たすインスタンスのコレクションが生成される。例えば、フィルター処理により、特定のいくつかの値、または指定された範囲内のいくつかの値を持つインスタンスのコレクションを生成することができる。したがって、フィルター処理により、構造化プレゼンテーションに含まれるインスタンスの個数を減らすことができる。

【0139】

インスタンスフィルタートリガー1045によってトリガーされるフィルター処理は、ユーザーが1つまたは複数のフィルター処理基準を指定することを可能にするユーザーインターフェイスコンポーネントを提示することと、基準を満たすことができないインスタンスが表示されないように構造化プレゼンテーションを修正することを含むことができる。

【0140】

いくつかの実装では、ユーザーインターフェイスコンポーネント1000は、ユーザーインターフェイスコンポーネント1000を使用して、または他の何らかの方法でユーザーが行った修正に動的に応答することができる。例えば、ユーザーがトリガーを実行し、新規属性を構造化プレゼンテーションに追加する場合、その新規属性の識別子をコレクション1020に追加し、ユーザーインターフェイスコンポーネント1000内に提示することができる。例えば、ユーザーが「Attribute\_9」を構造化プレゼンテーションに追加する場合、属性識

10

20

30

40

50

別子「Attribute\_9」を、関連するアクショントリガー1030でユーザーインターフェイスコンポーネント1000に追加することができる。

【0141】

図11は、新規インスタンスを構造化プレゼンテーションに追加するための技術を指定するユーザー入力を受け取るためのユーザーインターフェイスコンポーネント1100の略図である。例えば、ユーザーインターフェイスコンポーネント1100を使用することで、ステップ910において、新規インスタンスを構造化プレゼンテーションに追加するためのオプションを提示し、プロセス900(図9)のステップ915において、オプションのユーザー選択を受け取ることができる。

【0142】

ユーザーインターフェイスコンポーネント1100は、ヘッダ1105、プロンプト1110、それぞれ選択ウィジェット1130、1135、1140に関連付けられている構造化プレゼンテーション1115、1120、1125に新規インスタンスを追加するための技術の記述のコレクションを含む。

【0143】

ヘッダ1105は、ユーザーインターフェイスコンポーネント1100とのユーザーインタラクションがユーザーが新規インスタンスを追加するための技術を指定することを実際に可能にすることを識別するテキストまたは他の情報を含む。プロンプト1110により、ユーザーはユーザーインターフェイスコンポーネント1100をインタラクティブに操作して新規インスタンスを追加するための技術を指定することを促される。

【0144】

記述1115には、この技術をユーザーが指定した結果としてユーザー指定制約条件オプションによって新規インスタンスが追加されることが記述される。選択ウィジェット1130のユーザーインタラクションにより、ユーザーが記述1115によって記述されているユーザー指定制約条件オプションを指定することが可能になる。

【0145】

記述1120には、この技術をユーザーが指定した結果としてユーザー指定制約条件オプションによって新規インスタンスが追加されることが記述される。記述1120は、制約条件追加ウィジェット1145および制約条件クリアウィジェット1150を含む。制約条件追加ウィジェット1145に対するユーザーインタラクションにより、ユーザー指定制約条件オプションで使用される新規制約条件の追加がトリガーされる。制約条件クリアウィジェット1150に対するユーザーインタラクションにより、すべての現在の制約条件がクリアされる。選択ウィジェット1135のユーザーインタラクションにより、ユーザーが記述1120によって記述されているユーザー指定制約条件オプションを指定することが可能になる。

【0146】

記述1125には、この技術をユーザーが指定した結果として手動オプションによって新規インスタンスが追加されることが記述される。記述1125は、新規インスタンス識別子入力フィールド1155を含む。新規インスタンス識別子入力フィールド1155に対するユーザーインタラクションにより、ユーザーは新規インスタンスを、例えば、名前で識別することができる。選択ウィジェット1140のユーザーインタラクションにより、ユーザーが記述1125によって記述されている手動オプションを指定することが可能になる。

【0147】

図12は、新規インスタンスを構造化プレゼンテーションに追加するためにユーザー指定制約条件オプションで使用される制約条件を指定するユーザー入力を受け取るためのユーザーインターフェイスコンポーネント1200の略図である。ユーザーインターフェイスコンポーネント1200は、他のユーザーインターフェイスコンポーネントから独立して(例えば、専用ウィンドウまたはポータル上で)、または他のユーザーインターフェイスコンポーネントと連携して使用することができる。例えば、ユーザーインターフェイスコンポーネント1200は、技術記述1120(図11)のすぐ下のユーザーインターフェイス1100内に挿入することができる。例えば、ユーザーインターフェイスコンポーネント1200は、ステップ930

10

20

30

40

50

において構造化プレゼンテーションに追加される新規インスタンスの属性の値を指定するためのオプションを提示し、プロセス900(図9)のステップ935で属性のそのような値のユーザー指定を受け取るために使用することができる。

**【0148】**

ユーザーインターフェイスコンポーネント1200は、それぞれが値指定領域1215、1220に関連付けられている1つまたは複数の属性選択ウィジェット1205のコレクションを含む。属性選択ウィジェット1205、1210は、値が制約される属性をユーザーが選択することを可能にするインタラクティブ表示要素である。例示されている実装では、それぞれの属性選択ウィジェット1205、1210は、属性の識別子をリストするドロップダウンボックスウィジェットである。いくつかの実装では、リストされている属性識別子は、新規インスタンスが追加される構造化プレゼンテーション内の属性識別子308と同一のものとすることができる。

10

**【0149】**

値指定領域1215、1220は、ユーザーが各属性選択ウィジェット1205、1210において識別された属性の値に対し1つまたは複数の制約条件を指定することを可能にするインタラクティブ表示要素である。例示されている実装では、値指定領域1215は、ユーザーが属性選択ウィジェット1205において識別された属性の値の許容可能な範囲を指定することを可能にする一対のテキスト入力フィールド1225を備える。値指定領域1220は、ユーザーが属性選択ウィジェット1210において識別された属性の許容可能な値を指定することを可能にするインタラクティブチェックボックス1230のコレクションを備える。

20

**【0150】**

動作中、ユーザーが属性選択ウィジェット1205、1210を使用して特定の属性識別子を選択すると、関連付けられている値指定領域1215、1220における変更がトリガーされうる。例えば、インタラクティブ要素の性質および関連付けられている値指定領域1215、1220において指定されうる値および/または範囲を変更することができる。いくつかの実装では、これらの変更は、新規インスタンスが追加される構造化プレゼンテーション内のそのような属性の値の分布に基づくものとすることができる。例えば、属性「maker」の4つの値のみが構造化プレゼンテーション内に出現する場合、これらの同じ4つの値は、関連付けられている値指定領域における指定のために提示することができる。他の実装では、関連付けられている値指定領域1215、1220に加えらる変更は、電子文書コレクション102内の類似のインスタンスを特徴付ける属性の値に基づくものとすることができる。例えば、自動車のインスタンスの属性「maker」は、より多様な値を使用して電子文書コレクション102内の文書において特徴付けることができる。これらの値は、関連付けられている値指定領域内の指定について識別し、提示することができる。

30

**【0151】**

図13は、新規属性を構造化プレゼンテーションに追加するための例示的なプロセス1300の流れ図である。プロセス1300は、1つまたは複数の機械可読命令セットを実行することによってオペレーションを実行する1つまたは複数のコンピュータによって実行されうる。これらのデジタルデータ処理デバイスは、入力および出力デバイス、例えば、キーボード、マウス、タッチスクリーン、ディスプレイ画面、および同様のものを介してユーザーと情報をやり取りすることができる。例えば、システム200(図2)の背景状況において、プロセス1300におけるユーザーインタラクティブ性をPDA215またはデスクトップコンピュータ217などのクライアント側で実行できる。

40

**【0152】**

プロセス1300は、単独で、または他のデータ処理活動と連携して実行されうる。例えば、以下でさらに詳しく説明するように、プロセス1300は、事前に存在している構造化プレゼンテーションに加えて属性提案を構成するためにさまざまなプロセスと連携して実行されうる。このような構成プロセスの例は、図37~51および関連する本文において、また図21~26および関連する本文において説明されている。一般に、プロセス1300は、複数のデジタルデータ処理デバイスによって実行される。例えば、システム200(図2)の背景状況に

50

において、属性提案を構成するための活動を検索エンジン202側で実行することができ、その一方で、ユーザーインタラクションはPDA215またはデスクトップコンピュータ217(図2)などのクライアント側で実行されうる。

【0153】

プロセス1300を実行するシステムは、新規属性トリガーを受け取ることができる(ステップ1305)。新規属性は、構造化プレゼンテーション、例えば、構造化プレゼンテーション106(図1)に現在表示されていない属性である。新規属性トリガーは、新規属性を構造化プレゼンテーションに追加するためのプロセスをアクティベートする。例えば、新規属性は、マウス、スタイラス、キーボード、または同様のものを介して受け取ったユーザー入力によってトリガーされうる。他の実装では、新規属性は、他のプロセスまたはシステムによってトリガーされうる。新規属性トリガーは、2つほど例を挙げると、プロセス間通信またはアプリケーションのメッセージハンドラを通じて受け取ることができる。例えば、いくつかの実装では、システムは、新規属性追加トリガー1035(図10)のユーザー選択を通じてユーザーインターフェイスコンポーネント1000から新規属性トリガーを受け取ることができる。

10

【0154】

プロセス1300を実行するシステムは、新規属性を指定するためのオプションを提示することができる(ステップ1310)。例えば、システムは、構造化プレゼンテーション内のインスタンスを特徴付けるために使用される新規属性のリスト、さらにはユーザーがそれらの属性のうちの1つまたは複数の属性を選択することを可能にするインタラクティブ表示要素を表示することができる。いくつかの実装では、そのようなリスト内の属性は、電子文書コレクションの文書内など、他の場所にあるインスタンスを特徴付けるために使用される属性に基づいて構成することができる。このような属性を構成するための例示的なアプローチは、図37~51および関連する本文において説明されている。

20

【0155】

プロセス1300を実行するシステムは、ユーザーから新規属性の指定を受け取ることができる(ステップ1315)。属性の指定では、例えば、新規属性の名前もしくは他の識別子を含む新規属性の特質もしくは特性、新規属性に関連付けられているキーワード、新規属性に関する信頼できる情報源、および同様のものを特徴付けることができる。属性の指定は、1つまたは複数の入力デバイス、例えば、キーボード、タッチパッド、またはタッチスクリーンを介してユーザーから受け取ることができる。

30

【0156】

プロセス1300を実行するシステムは、指定された新規属性を構造化プレゼンテーションに追加することができる(ステップ1320)。例えば、プロセス1300を実行するシステムは、新規属性識別子308および列304をテーブル300、400に追加することができる(図3、4)。他の例として、システムは、カードコレクション500(図5)の列506内の対応する属性値307とともに、新規属性識別子308を列504内に追加することができる。いくつかの実装では、プロセス1300を実行するシステムは、新規属性を構造化プレゼンテーションに追加だけでなく、構造化プレゼンテーションの修正を指定するユーザー入力を受け取るためにユーザーインターフェイスコンポーネントにも追加することができる。例えば、システムは、新規属性をユーザーインターフェイスコンポーネント1000(図10)の属性修正領域1005に追加することができる。

40

【0157】

プロセス1300を実行するシステムは、ユーザー指定に少なくとも一部は基づいて属性値を書き込むことができる(ステップ1325)。システムは、以下でさらに詳しく説明するように、さまざまな技術を使用して属性値を書き込むことができる。

【0158】

図14は、新規属性を構造化プレゼンテーションに追加するためのユーザーインターフェイスコンポーネント1400の略図である。ユーザーインターフェイスコンポーネント1400は、新規属性の1つまたは複数の特質もしくは特性の指定に関してユーザーと情報をやり取

50

りすることができる。これらの特質または特性は、例えば、新規属性および属性値を構造化プレゼンテーションに追加する際に使用することができる。例えば、ユーザーインターフェイスコンポーネント1400を使用することで、ステップ1310において新規属性クラスを構造化プレゼンテーションに追加するためオプションを提示し、プロセス1300(図13)のステップ1315において新規属性のユーザー指定を受け取ることができる。

**【0159】**

ユーザーインターフェイスコンポーネント1400は、ヘッダ1405および新規属性を特徴付ける特質を識別する特質識別子1410、1415、1420、1425のコレクションを含む。それぞれの特質識別子1410、1415、1420、1425は、特質指定ウィジェット1430、1435、1440、1445に関連付けられ、そのウィジェットに対するユーザーインタラクションによって指定される特質を識別する。ヘッダ1405は、ユーザーインターフェイスコンポーネント1400とのユーザーインタラクションが、ユーザーが新規インスタンスを構造化プレゼンテーションに追加することを実際に可能にすることを識別するテキストまたは他の情報を含む。

10

**【0160】**

特質識別子1410は、ユーザーが特質指定ウィジェット1430をインタラクティブに操作することによって構造化プレゼンテーションに追加される属性のクラスを指定することができることを識別する。属性のクラスは、属性およびその値がどのように識別されるかを示す。例えば、属性クラスにより、電子文書コレクションにおいて属性およびその値を識別するために使用する技術を指定することができる。属性クラスの例としては、「auto-find values」、「search results」、「review」、および「note」クラスが挙げられる。これらの属性クラスに関する詳細を以下でさらに説明する。特質指定ウィジェット1430は、ユーザーが構造化プレゼンテーションに追加される属性のクラスを指定することを可能にするインタラクティブ表示要素である。例示されている実装では、特質指定ウィジェット1430は、ドロップダウンボックスウィジェットである。

20

**【0161】**

特質識別子1415は、ユーザーが特質指定ウィジェット1435をインタラクティブに操作することによって新規属性の名前または他の識別子を指定することができることを識別する。特質指定ウィジェット1435は、ユーザーが構造化プレゼンテーションに追加される新規属性の名前または他の識別子を指定することを可能にするインタラクティブ表示要素である。例示されている実装では、特質指定ウィジェット1435は、テキスト入力フィールドを含む。一般に、特質識別子1415において識別される属性識別子は、属性識別子308として構造化プレゼンテーション内に直接追加することができる。

30

**【0162】**

特質識別子1420は、ユーザーが特質指定ウィジェット1440をインタラクティブに操作することによって新規属性を特徴付けるキーワードを指定することができることを識別する。特質指定ウィジェット1440は、ユーザーが構造化プレゼンテーションに追加される属性を特徴付ける1つまたは複数のキーワードを指定することを可能にするインタラクティブ表示要素である。例示されている実装では、特質指定ウィジェット1440は、1つまたは複数のキーワードを入力することができるテキスト入力フィールドを含む。キーワードは、例えば、属性識別子の背景状況の特徴付ける属性識別子または検索語の同義語を含んでもよい。例えば、属性識別子が「bank」である場合、特質指定ウィジェット1440において識別されるキーワードは、「NASCAR」および「speedway」を含み、その属性が金融機関とは反対にレーストラックの「bank」を指すことを示すことができる。

40

**【0163】**

動作中、特質指定ウィジェット1440で指定されたキーワードは、電子文書コレクションの検索の際にインスタンス、属性、および/または属性値を識別するために使用される。例えば、キーワードは、図21~26および関連する本文ならびに図37~51および関連する本文において説明されているような新規属性および/または新規インスタンスを構成するときに使用できる。

**【0164】**

50

特質識別子1425は、ユーザーが特質指定ウィジェット1445をインタラクティブに操作することによって新規属性を特徴付ける「favorite sites」を指定することができることを識別する。「favorite sites」は、電子文書コレクション内の文書である。文書の「favorite sites」としてのユーザー指定は、文書のコンテンツが新規属性に関連すると同時に、真である可能性が高いとユーザーがみなすことを示す。したがって、例えば、(以下でさらに説明するように)事前に存在している構造化プレゼンテーションに追加するために新規インスタンスおよび新規属性を構成する際に、「favorite sites」のコンテンツに高い信頼値を割り当てることができる。文書の「favorite sites」としてのユーザー指定は、文書のコンテンツが構造化プレゼンテーションに書き込む属性値に関して信頼できるものであること示す指標として使用することもできる。

10

## 【0165】

特質指定ウィジェット1445は、ユーザーが電子文書コレクション内の1つまたは複数の文書を「favorite sites」として指定することを可能にするインタラクティブ表示要素である。例示されている実装では、特質指定ウィジェット1445は、例えば1つまたは複数のドメイン名または電子文書の他のロケーションを入力することができるテキスト入力フィールドを含む。

## 【0166】

いくつかの実装では、特質「指定解除」ウィジェットは、ユーザーが電子文書コレクション内の1つまたは複数の文書が「disfavored」サイトであると識別することを可能にする。文書の「disfavored site」としてのユーザー指定は、ユーザーがその文書を属性値のソースとして信頼していないことを示す。このような特質指定解除ウィジェットは、例えば1つまたは複数のドメイン名または電子文書の他のロケーションを入力することができるテキスト入力フィールドを含むことができる。

20

## 【0167】

図15は、新規属性値を構造化プレゼンテーションに追加するための例示的なプロセス1500の流れ図である。プロセス1500は、1つまたは複数の機械可読命令セットを実行することによってオペレーションを実行する1つまたは複数のコンピュータによって実行される。プロセス1500は、単独で、または他のデータ処理活動と連携して実行される。例えば、以下でさらに詳しく説明するように、プロセス1500は、新規属性を構造化プレゼンテーションに追加するためのさまざまなプロセス、例えば、プロセス1300(図13)と連携して実行される。

30

## 【0168】

プロセス1500を実行するシステムは、新規属性のクラスのユーザー指定を受け取ることができる(ステップ1505)。上述のように、属性のクラスは、属性およびその値がどのように識別されるかを示す。新規属性のクラスを受け取ることは、プロセス1300(図13)のステップ1315における新規属性の指定を受け取ることの一部とすることができる。いくつかの実装では、新規属性のクラスのユーザー指定は、ユーザーインターフェイスコンポーネント1400(図14)において特質指定ウィジェット1430を介して受け取ることができる。

## 【0169】

プロセス1500を実行するシステムは、どのクラスが新規属性に対して指定されるかを決定することができる(ステップ1510)。指定されたクラスに基づき、プロセス1500を実行するシステムは、新規属性値を構造化プレゼンテーションに追加するためのさまざまなサブプロセスのうちのどれが実行されるかを決定することができる。例えば、システムは、「note」クラスに関連付けられているサブプロセス、「reviews」クラスに関連付けられているサブプロセス、「search results」クラスに関連付けられているサブプロセス、または「already found」クラスに関連付けられているサブプロセスに従って属性値を追加することを決定することができる。

40

## 【0170】

プロセス1500を実行するシステムが、「note」クラスに関連付けられているサブプロセスを使用して新規属性値を追加することを決定する場合、システムは、属性値にユーザー

50

から受け取った注釈を書き込むことができる(ステップ1515)。例えば、図4の背景状況において、テーブル400内の注釈列420内の値をユーザーから受け取り、これを使用して新規属性の値を書き込むことができる。

【0171】

プロセス1500を実行するシステムが、「reviews」クラスに関連付けられているサブプロセスを使用して新規属性値を追加することを決定する場合、システムは、レビューを含む電子文書を検索し、識別することができる(ステップ1520)。レビューは、新規属性によって特徴付けられている1つまたは複数のインスタンスの厳しい評価である。いくつかの場合において、レビューは、批評家などの、インスタンスを評価する専門知識を有する誰かが作成することができる。レビューは、例えば、ラベル、またはそれらのレビューをレビューとして識別する他のテキストに基づいて識別されうる。例えば、特定のドメイン名(例えば、<http://www.google.com/prdhp>、<http://www.epinions.com/>、<http://www.amazon.com/>)を使用して、レビューを含む電子文書を識別することができる。レビューを含む電子文書は、電子文書コレクション、例えば、コレクション102内で見つけることができる。

10

【0172】

プロセス1500を実行するシステムは、識別されたレビューからのコンテンツを使用して属性値を書き込むことができる(ステップ1525)。例えば、システムは、1つまたは複数のテキストもしくはテーブルベースの抽出パターンを使用してレビューから値を抽出し、それらの抽出された値を構造化プレゼンテーションで提示することができる。これらの抽出パターンは、「感情集中(sentiment focused)」であるレビュー文書のセグメントを優先的に選択することができる。感情集中セグメントは、特定の主題に関して、肯定的なまたは否定的な声に出す強い感情として識別される。例えば、レストランのレビューは、「the food is exceptionally good」(その食べ物は並外れて良い)および「the service was very poor indeed」(サービスは全く悪かった)などの感情集中セグメントを含むことも可能である。構造化プレゼンテーションにおけるこれらの抽出された値の提示は、プロセス1300(図13)のステップ1325における構造化プレゼンテーションの書き込みの一部とすることができる。

20

【0173】

プロセス1500を実行するシステムが、「search results」クラスに関連付けられているサブプロセスを使用して新規属性値を追加することを決定する場合、システムは、電子文書コレクション、例えば、コレクション102から検索結果のコレクションを生成することができる(ステップ1530)。この検索は、レビューに限定されない結果集合を生成することができるが、むしろ、さまざまな電子文書を含みうる。電子文書は、電子文書コレクション、例えば、コレクション102内で見つけることができる。

30

【0174】

検索結果は、新規属性の識別子、さらにはその属性によって特徴付けられたインスタンスの識別子に基づく検索によって生成されうる。いくつかの実装では、新規属性に関連付けられている追加のキーワードを使用して、検索結果、例えば、ユーザーインターフェイスコンポーネント1400(図14)の特質指定ウィジェット1440を介してユーザーから受け取ったキーワードを改善することができる。

40

【0175】

プロセス1500を実行するシステムは、検索結果集合からのコンテンツを構造化プレゼンテーション内の属性値に書き込むことができる(ステップ1535)。例えば、システムは、1つまたは複数のテキストもしくはテーブルベースの抽出パターンを使用して検索結果集合から1つまたは複数の値を抽出し、それらの抽出された値を構造化プレゼンテーションで提示することができる。これらの属性値に検索結果集合のコンテンツを書き込むことは、プロセス1300(図13)のステップ1325における構造化プレゼンテーションの書き込みの一部とすることができる。

【0176】

50

プロセス1500を実行するシステムが、「already found」クラスに関連付けられているサブプロセスを使用して新規属性値を追加することを決定する場合、システムは、すでに見つかっていて、電子文書コレクション、例えば、電子文書コレクション102から抽出されている値を識別することができる(ステップ1540)。「already found」値は、例えば、電子文書の特徴付ける情報のコレクション、例えば、システム200(図2)内のデータセンター208に格納することができる。いくつかの実装では、情報のそのようなコレクションは、前の構造化プレゼンテーションの履歴レコードを含むことができる。プロセス1500を実行するシステムは、すでに抽出されている値を構造化プレゼンテーションの属性値に書き込むことができる(ステップ1545)。これらの属性値に検索結果集合のコンテンツを書き込むことは、プロセス1300(図13)のステップ1325における構造化プレゼンテーションの書き込みの一部とすることができる。

10

**【0177】**

図16は、新規属性値を構造化プレゼンテーションに追加するための例示的なプロセス1600の流れ図である。特に、プロセス1600は、構造化プレゼンテーションの属性値を書き込む際に使用する属性値を選択することに関連する。プロセス1600は、1つまたは複数の機械可読命令セットを実行することによってオペレーションを実行する1つまたは複数のコンピュータによって実行されうる。プロセス1600は、単独で、または他のデータ処理活動と連携して実行されうる。例えば、プロセス1600は、プロセス1300(図13)のステップ1325において、プロセス1500(図15)のステップ1525において、プロセス1500(図15)のステップ1535において、および/またはプロセス1500(図15)のステップ1545において実行されうる。

20

**【0178】**

プロセス1600を実行するシステムは、属性値の候補を識別することができる(ステップ1605)。属性値の候補は、例えば、コンテンツ(レビューまたは電子文書コレクション内の他の文書など)から直接抽出されるか、またはすでに抽出されている属性値のコレクションから識別されうる。例えば、図2の背景状況において、システムは、データセンター208にアクセスして、1つまたは複数の格納されている属性値を抽出することができる。

**【0179】**

プロセス1600を実行するシステムは、識別された候補値における確かさを決定することができる(ステップ1610)。候補値における確かさは、候補値がインスタンスの属性を正しく特徴付ける確度を特徴付けるものであるべきである。値の正確さの確かさは、例えば、インスタンスの属性を特徴付けるためにその値が使用される回数、インスタンスの属性を特徴付けるために使用される値の出所の文書の品質、および同様のものに基づいて決定することができる。

30

**【0180】**

プロセス1600を実行するシステムは、候補値のうちのいくつかの確かさが低、中、または高であるかを判定することができる(ステップ1615)。属性値の確かさが低の場合、これは、候補値がインスタンスの属性を正しく特徴付ける可能性が低いことを示す。属性値の確かさが高の場合、これは、候補値がインスタンスの属性を正しく特徴付ける可能性が高いことを示す。

40

**【0181】**

プロセス1600を実行するシステムが、候補値のうちのいくつかの確かさが高であると判定した場合、システムは、構造化プレゼンテーションにおける属性値に抽出した値を書き込むことができる(ステップ1545)。これは、自動的に、つまり、ユーザー入力なしで、実行できる。

**【0182】**

プロセス1600を実行するシステムが、候補値のうちのいくつかの確かさが中であると判定した場合、システムは、その候補値をユーザーに提供することができる(ステップ1625)。例えば、システムは、候補値をそれらの候補値によって潜在的に特徴付けられるインスタンスおよび属性の識別子に関連して提示するユーザーインターフェイスコンポーネント

50

を生成することができる。

【0183】

プロセス1600を実行するシステムは、提示される値のうちのいくつかの値のユーザー選択を受け取ることができる(ステップ1630)。ユーザー選択は、1つまたは複数のユーザー入力として受け取ることができる。例えば、候補値を提示するユーザーインターフェイスコンポーネントは、ユーザーが構造化プレゼンテーションに書き込むための候補値を選択することを可能にする1つまたは複数の選択ウィジェットを備えることができる。この選択は、マウス、キーボード、または他のユーザー入力デバイスを使用してユーザーから受け取ることができる。

【0184】

プロセス1600を実行するシステムは、選択された値を属性値に書き込むことができる(ステップ1635)。例えば、プロセス1600を実行するシステムは、構造化プレゼンテーションにおいて選択された値を提示することができる。

【0185】

いくつかの実装では、選択された属性値は、構造化プレゼンテーションで提示される属性、値、および/またはインスタンスをさらに改善するために使用できる。例えば、ユーザーがインスタンスの属性の値が数千ドルであると指定した場合、その値の大きさを使用して、構造化プレゼンテーションから著しく異なる大きさの値を除外することができる。他の例として、ユーザーがインスタンスの属性の値が数千ドルであると指定した場合、その値の大きさを使用して、大きさが著しく異なるその属性の値を有するインスタンスを除外することができる。

【0186】

プロセス1600を実行するシステムが、候補値のうちのいくつかの確かさが低であると判定した場合、プロセス1600を実行するシステムは、構造化プレゼンテーションにおけるそのような不足をハイライト表示することができる(ステップ1640)。これらの不足は、例えば、未確定の入力項目を残すか、または着色もしくは他の印を使用して低い信頼値をハイライト表示することによってハイライト表示することができる。システムは、インタラクティブ要素、例えば、未確定入力項目内のテキストフィールドまたは不足のある入力項目に隣接する注釈セルをインタラクティブに操作するユーザーからそれらの不足を是正する候補値を受け取ることでもできる場合がある。

【0187】

図17は、構造化プレゼンテーションに追加すべき候補値を選択するためのユーザーインターフェイスコンポーネント1700の略図である。ユーザーインターフェイスコンポーネント1700は、ユーザーと情報をやり取りして、構造化プレゼンテーションにおいて新規属性を特徴付ける値を選択することができる。例えば、ユーザーインターフェイスコンポーネント1700は、ステップ1625においてユーザーに対して提示され、プロセス1600(図16)のステップ1630においてユーザー選択を受け取ることができる。

【0188】

ユーザーインターフェイス1700は、ヘッダ1705およびテーブル1710を含む。ヘッダ1705は、ユーザーインターフェイスコンポーネント1700とのユーザーインタラクションがユーザーがインスタンスの属性の値を構造化プレゼンテーションにおいて表示するために選択することを可能にすることを識別するテキストまたは他の情報を含む。テーブル1710は、列1715、1720、1725に編成された候補値情報のコレクション、さらには行選択ウィジェット1730のコレクションを含む。

【0189】

特に、列1715は、列ヘッダ1735とともに候補値識別子のコレクションをも含む。候補値識別子は、電子文書コレクション102の文書から直接的に、またはデータセンター208を介して間接的に抽出されている可能性がある。いくつかの実装では、値は、特定の値307に対する測定単位を指定する単位識別子309も含むことができる。列ヘッダ1735で、候補値識別子が列1715内に見つかることが識別される。

10

20

30

40

50

## 【 0 1 9 0 】

列1720は、列ヘッダ1740とともに信頼値のコレクションをも含む。信頼値は、列1715で識別されている候補値が正しい可能性を示す。信頼値は、数値または単語で表すことができる。例えば、信頼値は、例えば、値が正しい値であるか、または数字目盛上にあるかを確率としてパーセンテージで表すことができる。列ヘッダ1740で、信頼値が列1720内に見つかることが識別される。

## 【 0 1 9 1 】

列1725は、列ヘッダ1745とともにソース識別子のコレクションをも含む。ソース識別子は、列1715において識別された候補値の1つまたは複数のソースを識別する。ソースは、例えば、電子文書の表題、ドメイン名、作成者の名前、または同様のものを使用して識別することができる。いくつかの実装では、ソース識別子は、列1715において識別された候補値を含むテキストスニペットを含むことができる。列ヘッダ1745で、ソース識別子が列1720内に見つかることが識別される。

10

## 【 0 1 9 2 】

選択ウィジェットコレクション1730は、ユーザーから入力を受け取るために1つまたは複数のユーザーインタラクティブ要素を備える。ユーザー入力は、列1715で識別された候補値が構造化プレゼンテーションに追加されることを識別することができる。

## 【 0 1 9 3 】

いくつかの実装では、ユーザーインターフェイスコンポーネント1700は、信頼値に基づく順序で候補値を提示することができる。例えば、最高の信頼値を持つ候補値は、列1715の一番上に提示され、最低の信頼値を持つ候補値は、列1715の一番下に提示されうる。

20

## 【 0 1 9 4 】

いくつかの実装では、ユーザーインターフェイスコンポーネント1700は、列1725で識別されている特定のソース内の属性および値の周りのテキストのスニペットも含みうる。このようなスニペットを使用することで、ユーザーは背景状況における値を見ることができる。

## 【 0 1 9 5 】

図18は、中に提示される属性値の不足のハイライト1802を含む構造化プレゼンテーション1800の略図である。示されている例では、インスタンス「INSTANCE\_1」の属性「ATTR\_1」および「ATTRIBUTE\_N」を特徴付けるための候補となる値における確かさは、インスタンス「INSTANCE\_2」の属性「ATTR\_2」を特徴付けるための候補となる値における確かさと同様に、低い。インスタンス「INSTANCE\_1」の属性「ATTR\_1」の場合、この信頼の欠如は、空のセル1804によってハイライト表示される。インスタンス「INSTANCE\_1」の属性「ATTRIBUTE\_N」およびインスタンス「INSTANCE\_2」の属性「ATTR\_2」の場合、この信頼の欠如は、カラーの印1806によってハイライト表示される。このようなハイライト表示は、特定の属性値の性質に関するフィードバックの直観的な形態をもたらす。つまり、ユーザーは、テーブル300を閲覧し、その直後に、正しさが疑わしいのはどの値かを判定することができる。システムは、ハイライト表示された不足のうちの1つまたは複数に是正するユーザー入力を受け取ることができる。例えば、システムは、手動で入力された属性値、追加の制約条件、またはシステムが追加の属性値を確実に識別するために使用できるこの指定に記述されている他のユーザー入力を受け取ることができる。

30

40

## 【 0 1 9 6 】

いくつかの実装では、不足がハイライト表示されているセルに対するユーザーインタラクションは、不足を是正することを対象とする検索をトリガーすることができる。例えば、空のセル1804に対するユーザーインタラクションは、検索をトリガーしうる。検索では、例えば、ディスプレイ内のインスタンスのカテゴリ、新規値によって特徴付けられるインスタンスの識別子、および/または新規値によって特徴付けられる属性の識別子に基づくカスタマイズ可能なクエリを使用することができる。検索結果の集合を返した後、システムは、不足を是正する値を指定するインタラクションをさらに受けることができる。いくつかの実装では、返される検索結果の集合は、潜在的値を区別するテキストスニペット

50

における属性特有のハイライト表示を含むことができる。

【0197】

図19は、構造化プレゼンテーションに追加すべき候補属性を選択するためのユーザーインターフェイスコンポーネント1900の略図である。ユーザーインターフェイスコンポーネント1900は、ユーザーと情報をやり取りして、構造化プレゼンテーションにおいてインスタンスを特徴付ける属性を選択することができる。例えば、プロセス1300(図13)のステップ1320において構造化された表示に追加する属性を選択できるようにユーザーインターフェイスコンポーネント1900をユーザーに提示することができる。

【0198】

ユーザーインターフェイス1900は、ヘッダ1905およびテーブル1910を含む。ヘッダ1905は、ユーザーインターフェイスコンポーネント1900とのユーザーインタラクションがユーザーがインスタンスの属性を構造化プレゼンテーションにおいて表示するために選択することを可能にすることを識別するテキストまたは他の情報を含む。テーブル1910は、列1915、1920、1925に編成された候補属性情報のコレクション、さらには行選択ウィジェット1930のコレクションを含む。

10

【0199】

特に、列1915は、列ヘッダ1935とともに候補属性識別子のコレクションをも含む。候補属性識別子は、電子文書コレクション102の文書から直接的に、またはデータセンター208を介して間接的に抽出されている可能性がある。いくつかの実装では、属性は、型変換される候補属性の値の測定単位を指定する単位識別子309も含むことができる。列ヘッダ1935で、候補属性識別子が列1915内に見つかることが識別される。

20

【0200】

列1920は、列ヘッダ1940とともに信頼値のコレクションをも含む。信頼値は、列1915で識別されている候補属性が正しい可能性を示す。信頼値は、数値または単語で表すことができる。例えば、信頼値は、例えば、属性が正しい属性であるか、または数字目盛上にあるかを確率としてパーセンテージで表すことができる。列ヘッダ1940で、信頼値が列1920内に見つかることが識別される。

【0201】

列1925は、列ヘッダ1945とともにソース識別子のコレクションをも含む。ソース識別子は、列1915において識別された候補属性の1つまたは複数のソースを識別するものである。ソースは、例えば、電子文書の表題、ドメイン名、作成者の名前、または同様のものを使用して識別することができる。いくつかの実装では、ソース識別子は、列1915において識別された候補属性を含むテキストスニペットを含むことができる。列ヘッダ1945で、ソース識別子が列1920内に見つかることが識別される。

30

【0202】

選択ウィジェットコレクション1930は、ユーザーから入力を受け取るために1つまたは複数のユーザーインタラクティブ要素を備える。ユーザー入力は、列1915で識別された候補属性が構造化プレゼンテーションに追加されることを識別することができる。

【0203】

いくつかの実装では、ユーザーインターフェイスコンポーネント1900は、信頼値に基づく順序で候補属性を提示することができる。例えば、最高の信頼値を持つ候補属性は、列1915の一番上に提示され、最低の信頼値を持つ候補属性は、列1915の一番下に提示される。

40

【0204】

いくつかの実装では、ユーザーインターフェイスコンポーネント1900は、列1925で識別されている特定のソース内のインスタンスおよび属性の周りのテキストのスニペットも含みうる。このようなスニペットを使用することで、ユーザーは背景状況における属性を見ることができる。

【0205】

図20は、構造化プレゼンテーションに追加すべき候補インスタンスを選択するためのユ

50

ーザーインターフェイスコンポーネント2000の略図である。ユーザーインターフェイスコンポーネント2000は、ユーザーと情報をやり取りして、構造化プレゼンテーションに追加するインスタンスを選択することができる。例えば、プロセス900(図9)のステップ925、940において構造化された表示に追加するインスタンスを選択できるようにユーザーインターフェイスコンポーネント2000をユーザーに提示することができる。

【0206】

ユーザーインターフェイス2000は、ヘッダ2005およびテーブル2010を含む。ヘッダ2005は、ユーザーインターフェイスコンポーネント2000とのユーザーインタラクションがユーザーがインスタンスを構造化プレゼンテーションにおいて表示するために選択することを可能にすることを識別するテキストまたは他の情報を含む。テーブル2010は、列2015、2020、2025に編成された候補インスタンス情報のコレクション、さらには行選択ウィジェット2030のコレクションを含む。

10

【0207】

特に、列2015は、列ヘッダ2035とともに候補インスタンス識別子のコレクションをも含む。候補インスタンス識別子は、電子文書コレクション102の文書から直接的に、またはデータセンター208を介して間接的に抽出されている可能性がある。列ヘッダ2035で、候補インスタンス識別子が列2015内に見つかることが識別される。

【0208】

列2020は、列ヘッダ2040とともに信頼値のコレクションをも含む。信頼値は、列2015で識別されている候補インスタンスが追加される可能性を示す。信頼値は、数値または単語で表すことができる。例えば、信頼値は、例えば、インスタンスがユーザー指定制約条件に合致する確率をパーセンテージで表すことができる。列ヘッダ2040で、信頼値が列2020内に見つかることが識別される。

20

【0209】

列2025は、列ヘッダ2045とともにソース識別子のコレクションをも含む。ソース識別子は、列2015において識別された候補インスタンスの1つまたは複数のソースを識別する。ソースは、例えば、電子文書の表題、ドメイン名、作成者の名前、または同様のものを使用して識別することができる。いくつかの実装では、ソース識別子は、列2015内の候補インスタンスの識別子を含むテキストスニペットを含むことができる。列ヘッダ2045で、ソース識別子が列2020内に見つかることが識別される。

30

【0210】

選択ウィジェットコレクション2030は、ユーザーから入力を受け取るために1つまたは複数のユーザーインタラクティブ要素を備える。ユーザー入力は、列2015で識別された候補インスタンスが構造化プレゼンテーションに追加されることを識別することができる。

【0211】

いくつかの実装では、ユーザーインターフェイスコンポーネント2000は、信頼値に基づく順序で候補インスタンスを提示することができる。例えば、最高の信頼値を持つ候補インスタンスは、列2015の一番上に提示され、最低の信頼値を持つ候補インスタンスは、列2015の一番下に提示されうる。

【0212】

いくつかの実装では、ユーザーインターフェイスコンポーネント2000は、列2025で識別されている特定のソース内のインスタンス識別子の周りのテキストのスニペットも含みうる。このようなスニペットを使用することで、ユーザーは背景状況におけるインスタンスを見ることができる。

40

【0213】

本明細書で説明されているシステムおよびプロセスを使用して構造化プレゼンテーションに加えられた変更は、それらの変更が追加のインスタンス、属性、および/または値を識別するために使用される反復プロセスの一部とすることができる。例えば、プロセス800(図8)は、数回繰り返すことができる。既存のコンテンツの範囲は増大するので、識別される追加のインスタンス、属性、および/または値は、増大する確かさを持つものとなる

50

可能性がある。

【0214】

図21は、事前に存在している構造化プレゼンテーションを拡張するために新規インスタンスを追加することができるプロセス2100の略図である。プロセス2100は、1つまたは複数の機械可読命令セットによってオペレーションを実行する1つまたは複数のコンピュータのシステム、例えば、システム200(図2)によって実行されうる。

【0215】

プロセス2100は、電子文書コレクション102内の文書から引き出された情報に基づいて新規インスタンスを事前に存在している構造化プレゼンテーションに追加する抽出オペレーション2105とマージオペレーション2110とを含む。特に、プロセス2100は、事前に存在している構造化プレゼンテーション106で提示される情報に基づき1つまたは複数の新規インスタンスを提案する。例えば、構造化プレゼンテーションがいくつかの動画に対応する多数のインスタンスを含む場合、システム200は、電子文書コレクションから引き出された情報に従って動画の追加のインスタンスを提案することができる。つまり、システム200は、属性識別子の類似度、属性値の測定単位、属性値の値、またはこれらの組み合わせに従って追加のインスタンスを識別し、提案することができる。例えば、システム200は、類似のショートタイム、劇場、または放映時間を持つ動画を提案することができる。

【0216】

図21に示されているように、抽出オペレーション2105では、事前に存在している構造化プレゼンテーション106の特性を使用して、電子文書コレクション102から新規インスタンス提案のコレクションを抽出する。特性の例として、事前に存在している構造化プレゼンテーション内のインスタンス、事前に存在している構造化プレゼンテーション内の属性、および事前に存在している構造化プレゼンテーション内の属性の値が挙げられる。事前に存在している構造化プレゼンテーション106の特性は、機械可読情報のコレクションとして表すことができ、これを1つまたは複数の機械可読命令セットによりオペレーションを実行する1つまたは複数のコンピュータのシステムが受け取ることができる。例えば、検索エンジン202(図2)が、事前に存在している構造化プレゼンテーション106の特性を受け取ることができる。

【0217】

抽出オペレーション2105において、電子文書コレクション102内の文書のコンテンツおよび事前に存在している構造化プレゼンテーション106の特性に基づいて、1つまたは複数の新規インスタンス提案を構成することができる。以下でさらに説明されるように、新規インスタンス提案を構成するためのさまざまな異なる技術が使用されうる。

【0218】

マージオペレーション2110で新規インスタンス提案の一部または全部を事前に存在している構造化プレゼンテーション106にマージして、拡張構造化プレゼンテーション106を形成することができる。閲覧者に対して、例えば、ディスプレイ画面104などのディスプレイデバイスに拡張構造化プレゼンテーションを表示することができる。

【0219】

抽出オペレーション2105で構成されたすべての新規インスタンス提案は、事前に存在している構造化プレゼンテーション106とマージされ、閲覧者向けに表示される必要はない。例えば、いくつかの実装では、新規インスタンス提案のコレクションを、閲覧者が追加される1つまたは複数のインスタンスを選択することを可能にするインタラクティブ要素とともに閲覧者に対して提示することができる。しかし、他の実装では、新規インスタンス提案は、ユーザーインタラクションなしで、また表示する前の新規インスタンス提案の選別なしで、自動的に追加することができる。マージャーに関するさらなる詳細は、例えば、図9~20および関連する本文にある。

【0220】

図22は、電子文書コレクション内の文書のコンテンツに基づいてインスタンスを構造化プレゼンテーションに追加するための例示的なプロセス2200の流れ図である。プロセス22

10

20

30

40

50

00は、1つまたは複数の機械可読命令セットによってオペレーションを実行する1つまたは複数のコンピュータのシステムによって実行されうる。例えば、プロセス2200は、システム200内の検索エンジン202によって実行されうる。

【0221】

いくつかの実装では、プロセス2200は、例えば、ユーザーまたは構造化プレゼンテーション106の更新をトリガーする他のシステムもしくはプロセスから入力を受け取ったことに応答して実行されうる。例えば、プロセス2200は、1つまたは複数の新規インスタンスを構造化プレゼンテーション106に追加するユーザー要求への応答として実行されうる。他の例として、プロセス2200は、検索クエリを受け取ったことに対する応答として、検索エンジン、例えば、検索エンジン202(図2)によって実行されうる。

10

【0222】

プロセス2200を実行するシステムは、事前に存在している構造化表示の1つまたは複数の特性を受け取ることができる(ステップ2205)。例えば、システムは、事前に存在している構造化プレゼンテーションの1つまたは複数の属性識別子を受け取ることができる。他の例として、システムは、事前に存在している構造化プレゼンテーション内に出現する1つまたは複数のインスタンス識別子を受け取ることができる。

【0223】

プロセス2200を実行するシステムは、事前に存在している構造化プレゼンテーションの1つまたは複数の特性に基づいて電子文書コレクション内の文書から1つまたは複数のインスタンス提案を構成することができる(ステップ2210)。インスタンス提案は、多数の異なる方法でこれらの特性に基づき構成することができる。例えば、一実装では、システムは、事前に存在している構造化プレゼンテーションから引き出された属性識別子を使用して検索クエリを構築することによって電子文書コレクション102内の文書からインスタンス提案を構成することができる。これらの検索クエリを使用し、文字列比較結果または他のマッチング技術を使用して類似の属性を共有できるインスタンスを識別することができる。他のアプローチの例について以下でさらに説明する。

20

【0224】

プロセス2200を実行するシステムは、1つまたは複数のインスタンス提案をユーザーに提供することができる(ステップ2215)。例えば、インスタンス提案のリストを、事前に存在している構造化プレゼンテーションを表示する同じディスプレイ画面上にユーザー向けに表示することができる。

30

【0225】

プロセス2200を実行するシステムは、1つまたは複数のインスタンス提案のユーザー選択を受け取ることができる(ステップ2220)。例えば、ユーザーインターフェイスコンポーネントがユーザーとやり取りして、1つまたは複数のインスタンス提案を選択する1つまたは複数のユーザー入力(例えば、マウスのクリック、キーの押下、または他のユーザー入力)を受け取ることができる。

【0226】

プロセス2200を実行するシステムは、選択されたインスタンス提案を構造化プレゼンテーションに新規構造化レコードとして追加することができる(ステップ2225)。例えば、構造化プレゼンテーションが、テーブル300(図3)などのテーブルである場合、システムは新規行302を追加することができる。他の例として、構造化プレゼンテーションが、カードのコレクション500(図5)などのカードのコレクションである場合、システムは新規カード502を追加することができる。

40

【0227】

図23は、事前に存在している構造化プレゼンテーションの特性に基づいて電子文書コレクション内の電子文書からインスタンス提案を構成するための例示的なプロセス2300の流れ図である。プロセス2300は、単独で、または他の活動と連携して実行されうる。例えば、プロセス2300は、プロセス2200(図22)におけるステップ2210で実行されうる。

【0228】

50

プロセス2300は、1つまたは複数の機械可読命令セットによってオペレーションを実行する1つまたは複数のコンピュータのシステムによって実行されうる。例えば、プロセス2300は、システム200(図2)内の検索エンジン202によって実行されうる。

【0229】

プロセス2300を実行するシステムは、事前に存在している構造化プレゼンテーションにおいて指定されているインスタンスに関係する構造化コンポーネントを含む1つまたは複数の文書を識別することができる(ステップ2305)。構造化コンポーネントは、構造化されている電子文書のいくつかの部分または領域である。構造化コンポーネントの例としては、テーブル、リスト、レコード、属性-値の対のコレクション、および同様のものが挙げられる、こうして、構造化コンポーネントは、構造化プレゼンテーションとほぼ同様にし、定義済みの構造と一致するように属性値およびインスタンス識別子を編成することができる。

10

【0230】

構造化コンポーネントを含む電子文書の全体が構造化されている必要はない。例えば、電子文書は、非構造化テキストの2つの段落の間にテーブルが挟まれていてもよい。さらに、異なる文書内の構造化コンポーネントが、同じフォーマットを有する、または所定の、もしくは永続的な構造に従っている必要はない。実際、1つの構造化コンポーネント内の情報の編成は、一般的には、他の文書内に出現する構造化コンポーネント内の情報の編成とは無関係に変更されうる。例えば、1人の人の履歴書中の学校の構造化されたリストが卒業年を削除するように変更される場合、他の履歴書中の学校の他の構造化されたリストも同様に変更されることを保証する必要はない。

20

【0231】

プロセス2300を実行するシステムは、さまざまな方法で構造化コンポーネントを含む文書を識別することができる。例えば、テーブルおよび他の構造化コンポーネントは、文書それ自体の中に見つかる、メタデータラベル、例えば、HTMLタグを使用して識別されうる。他の例として、構造化コンポーネントは、文書内の反復要素(例えば、一連のコンマまたはタブ区切り記号)を識別することによって識別されうる。

【0232】

構造化コンポーネントは、それらが指定されたインスタンスに関連する情報を含む場合に事前に存在している構造化プレゼンテーションで指定されたインスタンスに関係する。例えば、1つまたは複数の属性値を持つ指定されたインスタンスのうちの1つまたは複数のインスタンスを特徴付ける構造化コンポーネントは、事前に存在している構造化プレゼンテーション内で指定されたインスタンスに関連するものとして考えることができる。他の例として、事前に存在している構造化プレゼンテーションで指定されたインスタンスとは異なるインスタンスの同じ属性のうちの1つまたは複数の属性を特徴付ける構造化コンポーネントは、指定されたインスタンスに関連するものとして考えることができる。多くの実装において、インスタンスおよび/または属性識別子は同じである必要はない。むしろ、概念的に関係するインスタンスおよび属性は、構造化コンポーネントを含む文書を識別するために使用されうる。

30

【0233】

したがって、いくつかの実装では、プロセス2300を実行するシステムは、事前に存在している構造化プレゼンテーション内に見つかるような同じもしくは関係するインスタンス識別子および/または事前に存在している構造化プレゼンテーション内に見つかるような同じもしくは関係する属性識別子を含む文書を識別することによって事前に存在している構造化プレゼンテーションにおいて指定されているインスタンスに関係する構造化コンポーネントを含む1つまたは複数の文書を識別することができる。

40

【0234】

プロセス2300を実行するシステムは、構造化コンポーネントから1つまたは複数のインスタンス提案を選択することができる(ステップ2310)。この選択プロセスでは、ユーザーに対して提案されるインスタンスを選別して個数を減らすことができる。インスタンス提

50

案の選択は、さまざまな方法で実行できる。例えば、システムは、以下でさらに説明するように、構造化コンポーネント内のインスタンス、構造化コンポーネント内のインスタンスの属性、および/または構造化コンポーネント内のインスタンスの属性の値のカテゴリに基づいてインスタンス提案を選択することができる。

#### 【0235】

図24は、事前に存在している構造化プレゼンテーションの特性に基づく電子文書コレクション内の電子文書からのインスタンス提案の構成を表す表現2400である。特に、表現2400は、プロセス2300(図23)の一実装を使用するインスタンス提案の構成を示している。

#### 【0236】

図示されているように、事前に存在している構造化プレゼンテーションでは、インスタンス2405のコレクションを指定する(つまり、インスタンス「Philadelphia」および「Chicago」)。さらに、電子文書コレクション(例えば、コレクション102)内の異なる文書は、異なる構造化コンポーネント2410、2415、2420を含む。構造化コンポーネント2410、2415、2420は、例えば中に出現する同じインスタンス識別子「Philadelphia」および「Chicago」に基づいて指定されたインスタンス2405に関連するものとして識別される。

#### 【0237】

図示されているように、構造化コンポーネント2410、2415、2420は、異なる背景状況に基づくさまざまな異なる潜在的インスタンス提案を含む。特に、構造化コンポーネント2410の背景状況では、インスタンス「Philadelphia」および「Chicago」は、さまざまな市の特性を表す表形式コンポーネントの一部である。構造化コンポーネント2415の背景状況では、インスタンス「Philadelphia」および「Chicago」は、1970年代のある時期のナショナルリーグ東地区の順位表の一部を表す構造化コンポーネントの一部である。構造化コンポーネント2420の背景状況では、インスタンス「Philadelphia」および「Chicago」は、さまざまな動画の特性を表す表形式コンポーネントの一部である。

#### 【0238】

構造化コンポーネント2410、2415、2420内に見つかるさまざまなインスタンスすべてをユーザーに対して提案する代わりに、それらのインスタンスを特徴付けるために使用される属性に基づいてコンポーネント2410、2415、2420からインスタンス選択を選択することができる。特に、図示されているように、事前に存在している構造化プレゼンテーション106は、属性「year」、「rating」、および「box office receipts」の値を使用してインスタンス「Philadelphia」および「Chicago」を特徴付ける。構造化コンポーネント2410は、属性「population」および「area」の値を使用してインスタンス「Philadelphia」および「Chicago」を特徴付ける。構造化コンポーネント2415は、属性「wins」、「losses」、および「GB」(つまり、ゲーム差)の値を使用してインスタンス「Philadelphia」および「Chicago」を特徴付ける。構造化コンポーネント2420は、属性「year」、「runtime」、および「rating」の値を使用してインスタンス「Philadelphia」および「Chicago」を特徴付ける。

#### 【0239】

システムは、これらの特徴付けられた属性に基づいて構造化コンポーネント2410、2415、2420内のインスタンスから選択することができる。例えば、システムは、事前に存在している構造化プレゼンテーション106における属性識別子「year」および「rating」と構造化コンポーネント2420における属性識別子「year」および「rating」との間の対応関係を識別して、インスタンス「Peter Pan」および「Star Wars」を事前に存在している構造化プレゼンテーション106に追加する提案として選択することができる。

#### 【0240】

図37~51および関連する本文において説明されているように、システムは、追加の属性識別子を提案もしくは追加することもできる。例えば、構造化コンポーネント2420は、属性識別子「runtime」を含む。したがって、このようなシステムでは、対応する属性値を使用して、または対応する属性値なしで、属性識別子「runtime」を提案することができる。

10

20

30

40

50

## 【0241】

いくつかの実装では、構造化コンポーネント2410、2415から引き出されたインスタンスが特定の構成で提案されていない場合であっても、そのようなインスタンスは、将来の情報要求の際に使用できるように格納することができる。例えば、構造化コンポーネント2410で表されている市が、インスタンス提案として選択されていないとしても、それらの市は、各属性識別子(例えば、「population」および「area」)およびデータコレクション(例えば、データセンター208など)内の属性値とともに格納することができる。その後ユーザーが1つまたは複数の市に関する情報を要求した場合、そのようなシステムは、この格納されている情報にアクセスして、追加の情報をユーザーに提供することができる。

## 【0242】

図25は、事前に存在している構造化プレゼンテーションの特性に基づいて電子文書コレクション内の電子文書からインスタンス提案を構成するための例示的なプロセス2500の流れ図である。プロセス2500は、単独で、または他の活動と連携して実行されうる。例えば、プロセス2500は、プロセス2200(図22)におけるステップ2210で実行されうる。

## 【0243】

プロセス2500は、1つまたは複数の機械可読命令セットによってオペレーションを実行する1つまたは複数のコンピュータのシステムによって実行されうる。例えば、プロセス2500は、システム200(図2)内の検索エンジン202によって実行されうる。

## 【0244】

プロセス2500を実行するシステムは、1つまたは複数の指定されたインスタンスに関連する1つまたは複数の文書を識別することができる(ステップ2505)。例えば、プロセス2500を実行するシステムは、文字列比較を使用することで、指定されたインスタンスおよびその属性および/または値のうちの1つまたは複数と電子文書コレクション102などの電子文書コレクション内の文書とのマッチングを行うことができる。他の例では、プロセス2500を実行するシステムは、格納されている情報(データセンター208内の情報など)にアクセスして、指定されたインスタンスに関連する電子文書を識別することができる。

## 【0245】

プロセス2500を実行するシステムは、識別された文書のうちの1つまたは複数の文書のテンプレートを抽出することができる(ステップ2510)。文書テンプレートは、電子文書コレクション内の文書のサブコレクション内の個別の文書のコンテンツの配列に対するパターンとして使用される。サブコレクション内の文書は、一般的に、単一の企業などの単一のソースに由来するものである。例えば、書籍販売店は、単一の文書テンプレートを、異なる書籍を記述したコンテンツの配列に対するパターンとして使用することができる。他の例では、家具小売店は、単一の文書テンプレートを、異なるソファのパフレットのコンテンツの配列に対するパターンとして使用することができる。例えば、ソファ用の電子パフレットのテンプレートにより、ソファのブランド名、ソファの写真、ソファが示されている色をユーザーが選択することを可能にするインタラクティブ要素、テキスト形式のソファの説明、ならびにソファの寸法、入手状況、および価格を特徴付けるテーブルの配列をパフレット上に指定することができる。こうして、文書テンプレートは、構造化プレゼンテーションとほぼ同様にし、定義済みの構造と一致するようにインスタンスに関する情報を編成することができる。

## 【0246】

一般に、文書テンプレートは、電子文書のコンテンツ全体に対するパターンとして使用することができる。上述のように、文書内の構造化コンポーネントの配列を指定することさえできる。しかし、文書テンプレートでは非構造化電子文書コレクション内の文書のサブコレクションのコンテンツの配列を指定するだけなので、電子文書コレクションそれ自体は非構造化のままである。例えば、AMAZON.COMでは1つのテンプレートをAMAZON.COMが販売するすべての書籍の説明の配列に対するパターンとして使用する場合であっても、BARNESANDNOBLE.COMおよび他の書籍販売店は、その同じテンプレートを書籍店が販売する書籍の説明の配列に対するパターンとして必ずしも使用しない。さらに、文書テンプレートは

10

20

30

40

50

、その変更を必ずしも電子文書のコレクション全体、さらにはサブコレクションに伝搬させなくても変更できる。

【0247】

図26は、動画(つまり、動画「Philadelphia」)を記述するためのパターンとして使用されるハイパーテキストマークアップ言語(HTML)のテンプレート2600の一部を表す図である。テンプレート部分2600のハイパーテキストマークアップ言語(HTML)コードは、機械可読かつ人間可読である。例えば、ブラウザでテンプレート部分2600のHTMLコードを使用して、ウェブページを生成することができる。

【0248】

示されている例では、テンプレート部分2600は、2つのサブセクション2605、2610に分割される。サブセクション2605は、動画「Philadelphia」を識別するテキストの配列に対するパターンとして使用される。サブセクション2610は、さまざまな属性識別子およびその値の配列に対するパターンとして使用される。一般に、異なる動画を記述するために電子文書内の文書の特定のサブコレクションにおいて、サブセクション2605、2610内のパターンを何回も繰り返す。

【0249】

HTMLパーサーを使用して、テンプレート部分2600から出力書式を抽出することができ、したがって、出力書式を使用して、同じテンプレートを有する文書を識別することができる。例えば、HTMLタグ<title>、<div>、他のHTMLタグ、および互いに相対的な位置は、HTMLパーサーによって識別されうる。このようなHTMLパーサーは、HTMLタグ<title>がHTMLタグ<div>の前に出現するものとして判定することができる。したがって、HTMLパーサーは、テンプレートに従って配列されているコンテンツからのテンプレート部分2600から出力書式を抽出することができる。

【0250】

図25に戻ると、テンプレートを抽出した後に、プロセス2500を実行するシステムは、同じテンプレートを有する1つまたは複数の文書を識別することができる(ステップ2515)。例えば、システムは、電子文書コレクション内の文書のテンプレートを抽出されたテンプレートと比較することができる。

【0251】

プロセス2500を実行するシステムは、同じテンプレートを有するものとして識別されている文書から1つまたは複数のインスタンス提案を構成することもできる(ステップ2520)。特に、システムは、文書のサブコレクション内でテンプレートの反復を使用して、サブコレクション内の文書がインスタンスの同じカテゴリに関する同じ種類のコンテンツを含むと推論することができる。言い換えると、システムは、同じテンプレートが異なる文書のパターンとして使用されているため、2つの文書の背景状況が同じであると推論することができる。

【0252】

同様の背景状況の文書が識別された後、テンプレートそれ自体を使用して、インスタンス提案を構成することができる。例えば、テンプレート部分2600(図26)内のHTMLタグは、その文書内に記述されている映画の表題が「Philadelphia (1993)」であると識別する。同じテンプレートを共有する文書内の同様にタグ付けされたテキストを検索することによって、システムは、他の映画の表題を識別することができる。

【0253】

さらに、いくつかの実装では、文書テンプレート内の追加のコンテンツは、インスタンス提案を構成する際に使用することができる。例えば、特定の値(例えば、George Lucas)を「映画監督」として識別することを利用して、文書のサブコレクションから特定のインスタンス提案を選択することができる。言い換えると、テンプレート部分2600のサブセクション2610の構文解析を行うか、または他の何らかの形で分析して、属性が類似の値、識別子、または他の特性を有するかどうかを判定することができる。そのような状況では、インスタンス識別子は、サブセクション2605から抽出されうる。

## 【0254】

図27は、事前に存在している構造化プレゼンテーション106内の情報に基づいて新規インスタンス提案2115のコレクションを構成することができるプロセス2700の略図である。プロセス2700は、1つまたは複数の機械可読命令セット(図2)によってオペレーションを実行する1つまたは複数のコンピュータのシステムによって実行されうる。

## 【0255】

プロセス2700は、事前に存在している構造化プレゼンテーション106内の情報に基づきインスタンス/属性コレクション2710に対し抽出オペレーション2705を実行する。インスタンス/属性コレクション2710は、インスタンスをその属性に、またいくつかの実装では、それらの属性の値にも関連付ける情報のコレクションである。コレクション2710内の情報は、トリガーを受け取ったことに対する応答として(例えば、検索クエリ)、またはトリガーを受け取ることを予想して、例えば、電子文書コレクション102のインデックスを作成するプロセスの一部として、電子文書コレクション102内の文書から抽出されうる。いくつかの実装では、コレクション2710内の情報は、現在のユーザーまたは他のユーザーに対して提示された前の構造化プレゼンテーションのコンテンツを含むことができる。一般に、インスタンス提案は、例えばステップ2215、2220、2225(図22)において説明されているように、構造化プレゼンテーションに追加されるインスタンス提案を選択するユーザーに提供される。

## 【0256】

インスタンスとその属性の間の関連付けは、コレクション2710内の情報記憶域を構造化することによってコレクション2710内で確立されうる。図28は、コレクション2710内の属性およびインスタンスを関連付けるテーブル2800の略図である。テーブル2800は、レコード2802、2804、2806、2808、2810、2812、2814のコレクションを含み、それぞれのレコードはインスタンスの識別子を文書のロケーションの記述およびそれらの文書内の識別されたインスタンスを特徴付ける属性に関連付ける。レコード2802、2804、2806、2808、2810、2812、2814内の情報は、列2815、2820、2825、2830、2835、2840のコレクション内に編成されうる。特に、列2815は、インスタンス識別子を含むことができる。列2820は、列2815で識別されているインスタンスを含む電子文書のロケーションの記述を含むことができる。列2825、2830、2835、2840は、列2820内にロケーションが記述されている文書内の列2815で識別されるインスタンスを特徴付ける属性を識別することができる。

## 【0257】

図示されているように、異なる電子文書は、同じインスタンスを特徴付ける異なるカテゴリおよび量の情報を含むことができる。例えば、ロケーションがレコード2804の列2820で識別される文書は、インスタンス「INSTANCE\_2」の2つの属性を含むが、ロケーションがレコード2810の列2820で識別される文書は、インスタンス「INSTANCE\_2」の3つの属性を含む。さらに、レコード2804内の属性(つまり、属性「ATTR\_5」および属性「ATTR\_6」)は、レコード2810内の属性(つまり、属性「ATTR\_5」、属性「ATTR\_8」、および属性「ATTR\_9」)と一部異なる。

## 【0258】

属性およびインスタンスを関連付けるデータコレクション2710(テーブル280など)は、いくつかの異なる方法で形成されうる。例えば、内部的な構造化コンポーネントを含む文書が識別されうる。このような内部的な構造化コンポーネントの例として、HTML文書中に出現するテーブルおよびリストが挙げられる。これらの内部的な構造化コンポーネント内の属性とインスタンスとの間の関係をコピーしてデータコレクション2710を形成することができる。

## 【0259】

他の例として、コレクション2710は、現在のユーザーまたは他のユーザーに対して提示された前の構造化プレゼンテーションのコンテンツから形成することができる。

## 【0260】

さらに他の例として、第1の文書が属性およびインスタンスを含むものとして識別され

10

20

30

40

50

た後、その文書のテンプレートを使用して、同じテンプレートを含む他の文書から属性およびインスタンスを抽出することができる。例えば、ステレオ小売業者が同じ文書テンプレートを使用して販売のため提供される異なるステレオを記述する場合、第1のステレオに関する第1の電子文書内の情報の配列を使用して、他のステレオに関する他の電子文書から情報を抽出することができる。

**【 0 2 6 1 】**

いくつかの実装では、自然言語構文解析などの技術を使用して、インスタンスおよび属性を識別することができる。例えば、電子文書の構文解析を行って、電子文書内の「[Instance] has a/an [attribute]」などのフレーズを識別することができる。

**【 0 2 6 2 】**

いくつかの実装では、データコレクション2710は、インスタンスおよびその属性を分類することができる。例えば、インスタンスおよび属性は、北米の都市、ナショナルリーグ東地区チーム、または人気のある映画としてカテゴリ化することができる。データコレクション2710への情報の格納は、そのような分類に基づくことができる。例えば、異なるカテゴリを異なるファイル、レコード、および同様のものに格納することができる。

**【 0 2 6 3 】**

図27を再び参照すると、プロセス2700は、事前に存在している構造化プレゼンテーション106で提示される情報に基づき1つまたは複数の新規インスタンスを提案することができる。例えば、構造化プレゼンテーションがいくつかの動画に対応する多数のインスタンスを含む場合、システム200は、データコレクション2710から引き出された情報に従って動画の追加のインスタンスを提案することができる。つまり、システム200は、属性識別子の類似度に従って追加のインスタンスを識別し、提案することができる。例えば、システム200は、類似のショータイム、劇場、または放映時間を持つ動画を提案することができる。

**【 0 2 6 4 】**

図29は、事前に存在している構造化プレゼンテーションの特性に基づいてインスタンスおよび属性のコレクションからインスタンス提案を構成するためのプロセス2900の流れ図である。プロセス2900は、1つまたは複数の機械可読命令セットによってオペレーションを実行する1つまたは複数のコンピュータのシステムによって実行されうる。例えば、プロセス2900は、システム200(図2)内の検索エンジン202によって実行されうる。

**【 0 2 6 5 】**

プロセス2900は、単独で、または他の活動と連携して実行されうる。例えば、プロセス2900は、プロセス2700(図27)で実行されうる。他の例として、プロセス2900は、プロセス2200(図22)におけるステップ2210で、単独で、またはプロセス2300、2500(図23、25)の一方または両方と連携して実行されうる。

**【 0 2 6 6 】**

プロセス2900を実行するシステムは、インスタンスおよびその属性のコレクションにアクセスすることができる(ステップ2905)。例えば、プロセス2900を実行するシステムは、データセンター208(図2)に格納されているインスタンスおよび属性コレクション2710(図27)にアクセスすることができる。

**【 0 2 6 7 】**

プロセス2900を実行するシステムは、事前に存在している構造化プレゼンテーションにおいて指定されているインスタンス属性の特性に基づいて1つまたは複数の関連するインスタンスを識別することができる(ステップ2910)。例えば、システムは、インスタンスおよび属性コレクション2710(図27)内に格納されているインスタンスのインスタンス属性を構造化プレゼンテーション106で指定されているインスタンスと比較することができる。システムは、この比較結果を使用して、もしあれば格納されているインスタンスのうちのどれが属性識別子、または関係する属性を事前に存在している構造化プレゼンテーション106で指定されている属性と共有するかを決定することができる。例えば、事前に存在している構造化プレゼンテーション106は、属性「ATTR\_3」および「ATTR\_5」を使用してイ

10

20

30

40

50

インスタンスのコレクションを特徴付けると推定する。テーブル2800(図28)などのインスタンスおよび属性コレクション2710を検討した後、システムは、インスタンス「INSTANCE\_1」および「INSTANCE\_2」を、同じ属性「ATTR\_3」および「ATTR\_5」がレコード2802、2806内のインスタンスを特徴付けるために使用されることに基づいて、提案することができる。

#### 【0268】

図30は、事前に存在している構造化プレゼンテーション106内の情報に基づいて新規インスタンス提案2115のコレクションを構成するためのプロセス3000の流れ図である。プロセス3000は、1つまたは複数の機械可読命令セットによってオペレーションを実行する1つまたは複数のコンピュータのシステム、例えば、システム200(図2)によって実行されうる。プロセス3000は、単独で、または他の活動と連携して実行されうる。例えば、プロセス3000は、プロセス2700(図27)で実行されうる。他の例として、プロセス3000は、プロセス2200(図22)におけるステップ2210で、単独で、またはプロセス2300、2500、2900(図23、25、29)のうちの1つまたは複数と連携して実行されうる。いくつかの実装では、プロセス3000は、入力を(例えば、システム200のユーザーから、または他のシステムもしくはプロセスから)受け取ったことに応答して実行されうる。例えば、プロセス3000は、検索文字列を受け取ったことに応答して検索エンジン202によって実行されうる。

#### 【0269】

プロセス3000を実行するシステムは、1つまたは複数の指定されたインスタンスに関する1つまたは複数の信頼できるソースを識別することができる(ステップ3005)。例えば、システムは、例えばプログラマーによってアSEMBルされた電子文書コレクション102内の文書の信頼できるソースのコレクションにアクセスすることができる。

#### 【0270】

他の例として、システムは、電子文書コレクション102内の文書の1つまたは複数の信頼できるソースをそのユーザーの観点から「信頼できる」ものとして識別するユーザー特有の入力を受け取ることができる。例えば、事前に存在している構造化プレゼンテーション106を表示するディスプレイ画面104は、閲覧者が文書の信頼できるソースを指定することを可能にするGUIコンポーネントを備えることができる。信頼できるソースの識別は、検索クエリを受け取ることと連動して受け取ることができる。例えば、閲覧者は、JD POWER AND ASSOCIATES、AMAZON.COM、およびMAJOR LEAGUE BASEBALLをそれぞれ<http://www.jdpower.com/>、<http://www.amazon.com/>、および<http://www.mlb.com/>にある文書の信頼できるソースとして識別することができる。いくつかの実装では、ユーザー特有の入力は、ソースが信頼できるものである主題を識別することができる。例えば、MAJOR LEAGUE BASEBALLは、ベースボールの統計の信頼できるソースとして識別されうるが、薬物検査に関する情報の信頼できるソースとしてはみなされえない。

#### 【0271】

さらに他の例として、プロセス3000を実行するシステムは、信頼できるソースの「一般」コレクションをアSEMBルするために複数のユーザーからの信頼できるソースを識別するユーザー特有の入力のコレクションを分析することができる。例えば、多数のユーザーが、AMERICAN AUTOMOBILE ASSOCIATION (AAA)を信頼できるものとして識別することができる。次いで、これらの識別結果の統計分析に基づき、AAAを信頼できるソースのコレクションに追加することができる。

#### 【0272】

プロセス3000を実行するシステムは、事前に存在している構造化プレゼンテーションにおいて指定されているインスタンスに対する信頼できるソースから追加の属性を決定することができる(ステップ3010)。例えば、システムは、信頼できるソースによって提供される文書にアクセスし、1つまたは複数の属性を使用して指定されたインスタンスを特徴付ける1つまたは複数の文書を識別することができる。システムは、パーサーまたは他の文字列比較技術を使用してこれらの文書から属性識別子を抽出することができる。

#### 【0273】

他の例として、システムは、テーブル2800(図28)などの属性およびインスタンスを関連付けるデータコレクションにアクセスすることができる。システムは、レコード2802、2804、2806、2808、2810、2812、2814などのレコードを、事前に存在している構造化プレゼンテーションで識別されたインスタンスおよびロケーションがレコード2802、2804、2806、2808、2810、2812、2814で識別される文書が信頼できるソースに由来するものであったかどうかの両方に基づいてフィルター処理することができる。例えば、AMAZON.COMが信頼できるソースである場合、属性およびインスタンスを関連付けるコレクションをスキャンして、<http://www.amazon.com/>ドメインを持つ文書を識別することができる。

【0274】

プロセス3000を実行するシステムは、これらの追加の属性をテーブル2800(図28)などのインスタンスおよび属性コレクション内の属性と比較することができる(ステップ3015)。例えば、システムは、文字列比較、または他の比較技術を使用して、追加の属性をインスタンスおよび属性コレクションに格納されている属性と比較することができる。

【0275】

プロセス3000を実行するシステムは、これらの比較結果に基づいてインスタンスおよび属性コレクション内のインスタンスを識別することができる(ステップ3020)。例えば、システムは、信頼できるソースからの文書内のインスタンスを特徴付けるために使用される属性の数ならびにインスタンスおよび属性コレクション内の他のインスタンスに関連付けられている属性を決定することができる。

【0276】

図31は、事前に存在している構造化プレゼンテーション106内の情報に基づいて新規インスタンス提案2115のコレクションを構成するためのプロセス3100の流れ図である。プロセス3100は、1つまたは複数の機械可読命令セットによってオペレーションを実行する1つまたは複数のコンピュータのシステム、例えば、システム200(図2)によって実行されうる。プロセス3100は、単独で、または他の活動と連携して実行されうる。例えば、プロセス3100は、プロセス2700(図27)で実行されうる。他の例として、プロセス3100は、プロセス2200(図22)におけるステップ2210で、単独で、またはプロセス2300、2500、2900、3000(図23、25、29、30)のうちの1つまたは複数と連携して実行されうる。いくつかの実装では、プロセス3100は、入力を(例えば、システム200のユーザーから、または他のシステムもしくはプロセスから)受け取ったことに応答して実行されうる。例えば、プロセス3100は、検索文字列を受け取ったことに応答して検索エンジン202によって実行されうる。

【0277】

プロセス3100を実行するシステムは、事前に存在している構造化プレゼンテーションにおいて指定されているインスタンスの属性値に基づいて1つまたは複数の関連するインスタンスを識別することができる。例えば、システムは、指定されたインスタンスの属性値を他のインスタンスの属性値と比較することによって関連するインスタンスを識別し、他のインスタンスが適切な提案であるかどうかを判定することができる。このような比較では、例えば、属性値が同一のものである、または属性値が特定の範囲内にあるということが要求される場合がある。このような範囲は、例えば、インタラクティブ要素を介してユーザーによって指定されるか、または事前に存在している構造化プレゼンテーション内のインスタンスをすでに特徴付けている属性値の範囲に基づいて決定することができる。

【0278】

いくつかの実装では、プロセス3100を実行するシステムは、属性値を比較する前に属性値を通常の測定単位に変換することができる。例えば、指定された測定単位がフィートであるが、1つまたは複数の抽出された属性値の測定単位がメートルである場合、システムは、通常の技術を使用してメートル単位の値をフィートに変換することができる。このようなアプローチの概略表現を以下で詳述する。

【0279】

図32は、データコレクション内の属性、インスタンス、およびそれらの値を関連付けるテーブル3200の略図である。テーブル3200は、属性およびインスタンスを関連付けるので

10

20

30

40

50

、テーブル3200は、インスタンス/属性コレクション2710(図27)として使用することもできる。テーブル3200は、電子文書のコレクション、例えば電子文書コレクション102から引き出された情報に基づいて生成されうる。テーブル3200は、例えば、クローリングプロセスの実行中に生成され、例えば、その後使用するためデータセンター208内に格納されうる。

#### 【0280】

テーブル3200は、レコード3202、3204、3206、3208、3210、3212、3214のコレクションを含み、それぞれのレコードはインスタンスの識別子を文書のロケーションの記述、それらの文書内の識別されたインスタンスを特徴付ける属性、およびそれらの文書内のそれらの属性を特徴付ける値に関連付ける。レコード3202、3204、3206、3208、3210、3212、3214内の情報は、列3215、3220、3225、3230、3235、3240のコレクション内に編成されうる。特に、列3215は、インスタンス識別子を含むことができる。列3220は、列3215で識別されているインスタンスを含む電子文書のロケーションの記述を含むことができる。列3225、3235は、列3220内にロケーションが記述されている文書内の列3215で識別されるインスタンスを特徴付ける属性を識別することができる。列3230、3240は、列3225、3235で示されている属性を特徴付ける値を含むことができる。

10

#### 【0281】

示されている例では、それぞれのレコード3202、3204、3206、3208、3210、3212、3214は異なるインスタンスに係する(例えば、INSTANCE\_10からINSTANCE\_N)。これらのインスタンスのそれぞれが、属性識別子ATTR\_3、ATTR\_6によって少なくとも1つの文書内で特徴付けられる。そのようなものとして、インスタンス提案がINSTANCE\_10からINSTANCE\_Nを特徴付けるために使用できる属性のみに基づいて構成された場合、すべてのINSTANCE\_10からINSTANCE\_Nをユーザーに対して提案することができたことになる。

20

#### 【0282】

多くの状況において、これは受け入れられない。例えば、世界中のすべての専門学校または大学の同じ属性の多く(例えば、学生数、学生/教師の比率、所在地など)がインターネット上で利用可能な何らかの電子文書において特徴付けられる可能性がある。しかし、すべての専門学校または大学を含む提案のリストは、通う学校を探している学生にとっては役立たない。

#### 【0283】

したがって、プロセス3100(図31)について上で説明されているように、関連するインスタンスは、事前に存在している構造化プレゼンテーションで指定されているインスタンスの属性値を他のインスタンスの属性値と比較することによって識別されうる。例えば、構造化プレゼンテーション内の指定されたインスタンスが、単位「unit\_a」の値「VAL\_3」でインスタンスの属性「ATTR\_3」を、単位「unit\_c」の値「VAL\_6」でインスタンスの属性「ATTR\_6」を特徴付ける場合、検索エンジン202などのシステムは、共通の値(単位は異なるとしても)に基づいてレコード3202、3206で識別されたインスタンス(つまり、「INSTANCE\_10」および「INSTANCE\_12」)がユーザーに提案されうると識別することができる。そこで、システムは、セル3245、3250、および3255、3260内の値を共通測定単位に変換し、それらの値を比較してそれらの値が類似しているかどうかを判定することができる。こうして、それらの値を表す単位が異なっても、同様のインスタンスを選択することができる。

30

40

#### 【0284】

さらに、レコード3208で識別されたインスタンス(つまり、INSTANCE\_13)は、属性「ATTR\_3」の共通の値を構造化プレゼンテーション内の指定されたインスタンスと共有するけれども、レコード3208で示されたインスタンスは、ユーザーに対して提案されなくてもよい。特に、このインスタンスの属性「ATTR\_6」を特徴付ける値は、値「VAL\_8」であり、これは構造化プレゼンテーション内の指定されたインスタンスのこの属性を特徴付ける値と異なる。この相違に基づいて、レコード3208で識別されたインスタンスを提案されるインスタンスのリストから除外することができる。

50

## 【0285】

インスタンスを提案されるインスタンスのリストに入れ、また除外するために異なる基準を使用することができる。例えば、類似していなければならない属性値の個数は異なってもよい。他の例として、いくつかの実装では、ユーザーは、提案されるインスタンスのリストを構成する際に考慮される属性値の個数および/または性質を指定することができる。さらに他の例では、インスタンスの順位付けを、それらの属性値と事前に存在している構造化プレゼンテーション内の1つまたは複数の指定されたインスタンスの属性値との間の対応関係に基づいて実行できる。さらに他の例として、ある範囲の値を、構造化プレゼンテーションで指定されている1つまたは複数のインスタンスの属性を特徴付ける値に基づいて決定することができ、またこの範囲を使用して、提案されるインスタンスのリストに入れる関連するインスタンスを識別することができる。

10

## 【0286】

図9~20および関連する本文で説明されているように、いくつかの実装では、システムは、値が適切である確率を反映する基準に基づいて異なる値のコレクションのうちから値を選択することができる。このようなものの例として、ユーザー指定範囲、特定の値を持つ属性を特徴付ける文書の数、および/または特定の値を持つ属性を特徴付ける文書の品質が挙げられる。

## 【0287】

図33は、事前に存在している構造化プレゼンテーション106内の情報に基づいて新規インスタンス提案2115のコレクションを構成するためのプロセス3300の流れ図である。プロセス3300は、1つまたは複数の機械可読命令セットによってオペレーションを実行する1つまたは複数のコンピュータのシステム、例えば、システム200(図2)によって実行される。プロセス3300は、単独で、または他の活動と連携して実行される。例えば、プロセス3300は、プロセス2700(図27)で実行される。他の例として、プロセス3300は、プロセス2200(図22)におけるステップ2210で、単独で、またはプロセス2300、2500、2900、3000、3100(図23、25、29、30、31)のうちの1つまたは複数と連携して実行される。いくつかの実装では、プロセス3300は、入力を(例えば、システム200のユーザーから、または他のシステムもしくはプロセスから)受け取ったことに応答して実行される。例えば、プロセス3300は、検索文字列を受け取ったことに応答して検索エンジン202によって実行される。

20

30

## 【0288】

プロセス3300を実行するシステムは、インスタンスおよび属性のカテゴリ化されたコレクションにアクセスすることができる(ステップ3305)。例えば、システムは、前の検索時に生成されたインスタンスおよび属性の1つまたは複数のカテゴリ化されたコレクションにアクセスするために、インスタンスおよび属性コレクション2710にアクセスすることができる。

## 【0289】

プロセス3300を実行するシステムは、指定されたインスタンスを含むカテゴリを識別することができる(ステップ3310)。いくつかの実装では、システムは、類似の属性、類似の属性値、これらの特性の組み合わせ、および/または他の技術に基づくインスタンスを含むカテゴリを識別することができる。

40

## 【0290】

プロセス3300を実行するシステムは、識別されたカテゴリから1つまたは複数のインスタンス提案を選択することができる(ステップ3315)。例えば、いくつかの実装では、インスタンス提案を識別されたカテゴリから、指定されたインスタンスの属性値とそのカテゴリ内のインスタンスの属性値との間の類似度に基づいて選択することができる。

## 【0291】

図34は、事前に存在している構造化プレゼンテーションの特性に基づく電子文書コレクション内の電子文書からのインスタンス提案の構成を表す表現3400である。特に、表現3400は、プロセス3300(図33)の一実装を使用するインスタンス提案の構成を示している。

50

## 【0292】

図示されているように、事前に存在している構造化プレゼンテーションでは、インスタンス2405のコレクションを指定する(つまり、インスタンス「Philadelphia」および「Chicago」)。さらに、電子文書コレクション(例えば、コレクション102)内の異なる文書から引き出されたインスタンスは、異なるコレクション3410、3415、3420にすでにカテゴリ化されている。カテゴリ化されたインスタンスコレクション3410、3415、3420は、例えば中に出現する同じインスタンス識別子「Philadelphia」および「Chicago」に基づいて指定されたインスタンス2405に関連するものとして識別されうる。

## 【0293】

図示されているように、カテゴリ化されたインスタンスコレクション3410、3415、3420は、さまざまな異なる方法でカテゴリ化されている。特に、カテゴリ化されたインスタンスコレクション3410は、「North American Cities」のコレクションとしてカテゴリ化されている。カテゴリ化されたインスタンスコレクション3415は、「National League East Teams」のコレクションとしてカテゴリ化されている。カテゴリ化されたインスタンスコレクション3420は、「Popular Movies」のコレクションとしてカテゴリ化されている。システム200の背景状況において、カテゴリ化されたインスタンスコレクション3410、3415、3420は、データセンター208(図2)に格納することができる。つまり、システム200は、すでに受け取っている検索文字列に基づいてインスタンスの1つまたは複数のカテゴリを生成することができる。したがって、検索エンジン202が、検索文字列に基づいて検索を実行した後、検索エンジン202は、結果をカテゴリ化し、それらをデータセンター208内に格納することができる。その後の検索時に、これらのカテゴリ化された結果にアクセスし、分析して、インスタンス提案を生成することができる。

## 【0294】

事前に存在している構造化プレゼンテーションで指定されているインスタンスを含むカテゴリ化されたインスタンスコレクションは、例えば、指定されたインスタンスを特徴付ける属性と異なるカテゴリに含まれるインスタンスを特徴付ける属性との間の類似度に基づいて識別されうる。例えば、属性「year」および「rating」の一般的な使用では、事前に存在している構造化プレゼンテーションおよびカテゴリ化されたインスタンスコレクション3420を使用して、カテゴリ化されたインスタンスコレクション3420がインスタンス2405を含むものとして識別することができる。

## 【0295】

いくつかの実装では、カテゴリ化されたインスタンスコレクション内のインスタンスのサブセットは、カテゴリに含まれるインスタンスを特徴付ける値に基づいてインスタンス提案として選択されうる。例えば、インスタンス「Star Wars」は、「Philadelphia」および「Chicago」の「rating」属性を特徴付ける値に類似している「Star Wars」の「rating」属性を特徴付ける値に基づいてインスタンス提案のリスト上に含めることができる。他の例として、インスタンス「Peter Pan」は、「Philadelphia」および「Chicago」の「rating」属性を特徴付ける値と異なる「Peter Pan」の「rating」属性を特徴付ける値に基づいてインスタンス提案のリスト上で除外することができる。

## 【0296】

図35は、事前に存在している構造化プレゼンテーション106内の情報に基づいて新規インスタンス提案2115のコレクションを構成するために使用されうるプロセスのコレクション3500の略図である。

## 【0297】

コレクション3500内のプロセスは、潜在的インスタンス3505の大きなコレクションに連続的に適用され、インスタンス提案のより小さなコレクション3510を生成するフィルターとみなすことができる。プロセスコレクション3500は、カテゴリフィルター3515、関係する属性フィルター3520、および関係する値フィルター3525を備える。カテゴリフィルター3515は、例えば、プロセス3300(図33)のいくつかの態様を含むことができる。関係する属性フィルター3520は、例えば、プロセス2300(図23)の態様、プロセス2500(図25)、プロセ

10

20

30

40

50

ス2900(図29)、および/またはプロセス3000(図30)の態様を含むことができる。関係する値フィルター3525は、例えば、プロセス2300(図23)の態様、プロセス2500(図25)、プロセス3100(図31)の態様、および/またはプロセス3300(図33)の態様を含むことができる。それぞれのフィルターは、ユーザーに提示されるか、または構造化プレゼンテーションに直接追加されうるインスタンス提案コレクション3510から潜在的インスタンス3505を除外することができる。

【0298】

フィルター3515、3520、3525は、任意の順序で適用できる。しかし、一般に、フィルター3515、3520、3525は、粒度の順序で適用される。特に、潜在的インスタンスの数を最大量減らすフィルター3515、3520、3525が最初に適用され、潜在的インスタンスの数を最小量減らすフィルター3515、3520、3525が最後に適用される。

10

【0299】

いくつかの実装では、フィルター3515、3520、3525のうちのどれかをコレクション3500から省き、および/または追加のフィルターをコレクション3500に追加することができる。例えば、ユーザーによって供給される入力に従って潜在的インスタンス3505をフィルター処理することができるユーザー指定フィルターをコレクション3500に追加することができる。

【0300】

図36は、事前に存在している構造化プレゼンテーション106内の情報に基づいて新規インスタンス提案2115のコレクションを構成するためのプロセス3600の流れ図である。プロセス3600は、1つまたは複数の機械可読命令セットによってオペレーションを実行する1つまたは複数のコンピュータのシステム、例えば、システム200(図2)によって実行されうる。プロセス3600は、単独で、または他の活動と連携して実行されうる。例えば、プロセス3600は、プロセスコレクション3500におけるプロセスの1つまたは複数と連携して反復実行されうる。いくつかの実装では、プロセス3600は検索文字列を受け取ったことに応答して実行されうる。

20

【0301】

プロセス3600を実行するシステムは、事前に存在している構造化プレゼンテーション106で指定されたインスタンスと文書コレクションから引き出されたインスタンスとの間の初期マッチングを行うことができる(ステップ3605)。初期マッチングは、プロセスコレクション3500内のフィルター処理プロセスのうちの1つまたは複数に基づくことができる。

30

【0302】

プロセス3600を実行するシステムは、マッチングの数が大きすぎるか、小さすぎるか、または適切であるかを判定することができる(ステップ3610)。マッチングの数が少なすぎる場合、システムは、指定されたインスタンスの名前を広げることができる(ステップ3615)。例えば、プロセス3600を実行するシステムは、プロセスコレクション3500内のプロセスのうちの1つまたは複数で指定されているインスタンスについて代替スペリング、略語、同義語、代替名、ニックネーム、および/または他のキーワードを使用することができる。

【0303】

プロセス3600を実行するシステムは、関係する値フィルター処理3525(図35)で使用される属性値の1つまたは複数の範囲を広げることにもできる(ステップ3618)。この範囲は、ユーザーから受け取った入力に基づいて、または自動的に、ユーザー入力なしで広げることができる。例えば、いくつかの実装では、システムは、例えば、特定の割合のインスタンスまたは所定の数のインスタンスを含むように、インスタンスの選択されたグループに対する属性値の分布に基づいて範囲を広げることができる。

40

【0304】

プロセス3600を実行するシステムは、関係する属性フィルター処理3520(図35)で使用される属性および/またはインスタンスの数を減らすこともできる(ステップ3620)。属性および/またはインスタンスの数は、例えば、特定の属性および/またはインスタンスによ

50

て除外される潜在的インスタンスの数に基づいて減らすことができる。例えば、潜在的インスタンスを特徴付けるために特定の属性を使用するという要求条件により、潜在的インスタンスのすべてが除外される場合、この属性は、関係する属性フィルター処理から省くことができる。除去すべき属性および/またはインスタンスは、例えば、自動的に、ユーザー入力なしで、またはユーザーから受け取った入力に基づいて決定されうる。

**【0305】**

ここでもまた、システムは、事前に存在している構造化プレゼンテーション106で指定されたインスタンスと文書コレクションから引き出されたインスタンスとの間のマッチングを行おうとすることができるが、このときには、変更されたパラメータを使用する(ステップ3622)。このマッチングは、プロセスコレクション3500内のフィルター処理プロセスのうちの1つまたは複数を使用して行うこともできる。

10

**【0306】**

マッチングの数が多すぎると判定された場合(ステップ3610)、プロセス3600を実行するシステムは、関係する値フィルター処理3525(図35)で使用される属性値の1つまたは複数の範囲を狭めることができる(ステップ3625)。この範囲は、ユーザーから受け取った入力に基づいて、または自動的に、ユーザー入力なしで狭めることができる。例えば、いくつかの実装では、システムは、例えば、特定の割合のインスタンスまたは所定の数のインスタンスを除外するように、インスタンスの選択されたグループに対する属性値の分布に基づいて範囲を狭めることができる。

**【0307】**

20

プロセス3600を実行するシステムは、関係する属性フィルター3520(図35)で使用される属性および/またはインスタンスの数を増やすこともできる(ステップ3628)。属性および/またはインスタンスの数は、例えば、特定の属性および/またはインスタンスによって除外される潜在的インスタンスの数に基づいて増やすことができる。追加すべき属性および/またはインスタンスは、例えば、自動的に、ユーザー入力なしで、またはユーザーから受け取った入力に基づいて決定されうる。

**【0308】**

プロセス3600を実行するシステムは、変更されたパラメータに基づいてマッチングを選別することができる(ステップ3630)。特に、属性および/またはインスタンスの狭められた範囲および/または増やした数を関係する値フィルター処理3525(図35)で使うことができる。

30

**【0309】**

マッチングの数が受け入れられると判定された場合(ステップ3610)、プロセス3600を実行するシステムは、マッチしたインスタンスをユーザーに提案することができる(ステップ3635)。例えば、プロセス3600を実行するシステムは、ディスプレイ画面、例えば、ディスプレイ画面104上のGUIで1つまたは複数のインスタンス提案を提示することができる。

**【0310】**

図37は、事前に存在している構造化プレゼンテーションを拡張するために新規属性を追加することができるプロセス3700の略図である。プロセス3700は、1つまたは複数の機械可読命令セットを実行することによってオペレーションを実行する1つまたは複数のコンピュータを備えるシステム、例えば、システム200(図2)によって実行されうる。

40

**【0311】**

プロセス3700は、電子文書コレクション102内の文書から引き出された情報に基づいて新規属性を事前に存在している構造化プレゼンテーションに追加する抽出オペレーション3705とマージオペレーション3710とを含む。特に、プロセス3700は、事前に存在している構造化プレゼンテーション106で提示される情報に基づき1つまたは複数の新規属性を提案する。例えば、構造化プレゼンテーションがいくつかの動画に対応する多数のインスタンスを含む場合、システム200は、電子文書コレクションから引き出された情報に従って追加の動画属性を提案することができる。つまり、システム200は、インスタンス識別子の

50

類似度、インスタンスのカテゴリ、属性の値、またはこれらの組み合わせに従って追加の属性を識別し、提案することができる。

【0312】

図37に示されているように、抽出オペレーション3705では、事前に存在している構造化プレゼンテーション106の特性を使用して、電子文書コレクション102から新規属性提案のコレクションを抽出する。特性の例として、事前に存在している構造化プレゼンテーション内のインスタンス、事前に存在している構造化プレゼンテーション内のインスタンスのカテゴリ、および事前に存在している構造化プレゼンテーション内の属性の値が挙げられる。事前に存在している構造化プレゼンテーション106の特性は、機械可読情報のコレクションとして表すことができ、これを1つまたは複数の機械可読命令セットを実行することによってオペレーションを実行する1つまたは複数のコンピュータが受け取ることができる。例えば、検索エンジン202(図2)が、事前に存在している構造化プレゼンテーション106の特性を受け取ることができる。

10

【0313】

抽出オペレーション3705において、電子文書コレクション102内の文書のコンテンツおよび事前に存在している構造化プレゼンテーション106の特性に基づいて、1つまたは複数の新規属性提案を構成することができる。以下でさらに説明されるように、新規属性提案を構成するためのさまざまな異なる技術が使用されうる。

【0314】

マージオペレーション3710で新規属性提案の一部または全部を事前に存在している構造化プレゼンテーション106にマージして、拡張構造化プレゼンテーション106を形成することができる。閲覧者に対して、例えば、ディスプレイ画面104などのディスプレイデバイスに拡張構造化プレゼンテーションを表示することができる。

20

【0315】

抽出オペレーション3705で構成されたすべての新規属性提案は、事前に存在している構造化プレゼンテーション106とマージされ、閲覧者向けに表示される必要はない。例えば、いくつかの実装では、新規属性提案のコレクションを、閲覧者が追加される1つまたは複数の属性を選択することを可能にするインタラクティブ要素とともに閲覧者に対して提示することができる。しかし、他の実装では、新規属性提案は、ユーザーインタラクションなしで、また表示する前の新規属性提案の選別なしで、自動的に追加することができる。マージャーに関するさらなる詳細は、例えば、図9~20および関連する本文にある。

30

【0316】

図38は、電子文書コレクション内の文書のコンテンツに基づいて属性を構造化プレゼンテーションに追加するための例示的なプロセス3800の流れ図である。プロセス3800は、1つまたは複数の機械可読命令セットを実行することによってオペレーションを実行する1つまたは複数のコンピュータのシステムによって実行されうる。例えば、プロセス3800は、システム200内の検索エンジン202によって実行されうる。

【0317】

いくつかの実装では、プロセス3800は、例えば、ユーザーまたは構造化プレゼンテーション106の更新をトリガーする他のシステムもしくはプロセスから入力を受け取ったことに応答して実行されうる。例えば、プロセス3800は、1つまたは複数の新規属性を構造化プレゼンテーション106に追加するユーザー要求への応答として実行されうる。他の例として、プロセス3800は、検索クエリを受け取ったことに対する応答として、検索エンジン、例えば、検索エンジン202(図2)によって実行されうる。

40

【0318】

プロセス3800を実行するシステムは、事前に存在している構造化プレゼンテーションの1つまたは複数の特性を受け取ることができる(ステップ3805)。例えば、システムは、事前に存在している構造化プレゼンテーション内に出現する1つまたは複数のインスタンス識別子を受け取ることができる。他の例として、システムは、事前に存在している構造化プレゼンテーションで識別されたインスタンスを含むカテゴリの記述を受け取ることがで

50

きる。

【0319】

プロセス3800を実行するシステムは、事前に存在している構造化プレゼンテーションの1つまたは複数の特性に基づいて電子文書コレクション内の文書から1つまたは複数の属性提案を構成することができる(ステップ3810)。属性提案は、多数の異なる方法でこれらの特性に基づき構成することができる。例えば、一実装では、システムは、事前に存在している構造化プレゼンテーションから引き出されたインスタンス識別子を使用して検索クエリを構築することによって電子文書コレクション102内の文書から属性提案を構成することができる。これらの検索クエリを使用し、文字列比較結果または他のマッチング技術を使用して同じまたは類似のインスタンスを特徴付けることができる属性を識別することができる。他のアプローチの例について以下でさらに説明する。

10

【0320】

プロセス3800を実行するシステムは、1つまたは複数の属性提案をユーザーに提供することができる(ステップ3815)。例えば、属性提案のリストを、事前に存在している構造化プレゼンテーションを表示する同じディスプレイ画面上にユーザー向けに表示することができる。

【0321】

プロセス3800を実行するシステムは、1つまたは複数の属性提案のユーザー選択を受け取ることができる(ステップ3820)。例えば、ユーザーインターフェイスコンポーネントがユーザーとやり取りして、1つまたは複数の属性提案を選択する1つまたは複数のユーザー入力(例えば、マウスのクリック、キーの押下、または他のユーザー入力)を受け取ることができる。

20

【0322】

プロセス3800を実行するシステムは、選択された属性提案を構造化プレゼンテーションに追加することができる(ステップ3825)。特に、選択された属性提案は、構造化プレゼンテーション内の既存の構造化レコードを拡張するために使用することができる。例えば、構造化プレゼンテーションが、テーブル300(図3)などのテーブルである場合、システムは新規列304を追加することができる。他の例として、構造化プレゼンテーションが、カードのコレクション500(図5)などのカードのコレクションである場合、システムは新規属性識別子308をカード502に追加することができる。

30

【0323】

図39は、事前に存在している構造化プレゼンテーションの特性に基づいて電子文書コレクション内の電子文書から属性提案を構成するための例示的なプロセス3900の流れ図である。プロセス3900は、単独で、または他の活動と連携して実行されうる。例えば、プロセス3900は、プロセス3800(図38)におけるステップ3810で実行されうる。

【0324】

プロセス3900は、1つまたは複数の機械可読命令セットを実行することによってオペレーションを実行する1つまたは複数のコンピュータのシステムによって実行されうる。例えば、プロセス3900は、システム200(図2)内の検索エンジン202によって実行されうる。

【0325】

プロセス3900を実行するシステムは、事前に存在している構造化プレゼンテーションにおいて指定されているインスタンスに関係する構造化コンポーネントを含む1つまたは複数の文書を識別することができる(ステップ3905)。構造化コンポーネントは、構造化されている電子文書のいくつかの部分または領域である。構造化コンポーネントの例としては、テーブル、リスト、レコード、属性-値の対のコレクション、および同様のものが挙げられる、こうして、構造化コンポーネントは、構造化プレゼンテーションとほぼ同様にし、定義済みの構造と一致するように属性値およびインスタンス識別子を編成することができる。

40

【0326】

構造化コンポーネントを含む電子文書の全体が構造化されている必要はない。例えば、

50

電子文書は、非構造化テキストの2つの段落の間にテーブルが挟まれていてもよい。さらに、異なる文書内の構造化コンポーネントが、同じフォーマットを有する、または所定の、もしくは永続的な構造に従っている必要はない。実際、1つの構造化コンポーネント内の情報の編成は、一般的には、他の文書内に出現する構造化コンポーネント内の情報の編成とは無関係に変更されうる。例えば、1人の人の履歴書中の学校の構造化されたリストが卒業年を削除するように変更される場合、他の履歴書中の学校の他の構造化されたリストも同様に変更されることを保証する必要はない。

**【 0 3 2 7 】**

プロセス3900を実行するシステムは、さまざまな方法で構造化コンポーネントを含む文書を識別することができる。例えば、テーブルおよび他の構造化コンポーネントは、文書  
10  
それ自体の中に見つかる、HTMLタグなどのメタデータラベルを使用して識別されうる。他の例として、構造化コンポーネントは、文書内の反復要素(例えば、一連のコンマまたはタブ区切り記号)を識別することによって識別されうる。

**【 0 3 2 8 】**

構造化コンポーネントは、それらが指定されたインスタンスに関連する情報を含む場合に事前に存在している構造化プレゼンテーションで指定されたインスタンスに関連する。例えば、1つまたは複数の属性値を持つ指定されたインスタンスのうちの1つまたは複数の  
20  
インスタンスを特徴付ける構造化コンポーネントは、事前に存在している構造化プレゼンテーション内で指定されたインスタンスに関連するものとして考えることができる。他の例として、事前に存在している構造化プレゼンテーションで指定されたインスタンスとは異なるインスタンスの同じ属性のうちの1つまたは複数の属性を特徴付ける構造化コンポーネントは、指定されたインスタンスに関連するものとして考えることができる。多くの実装において、インスタンスおよび/または属性識別子は同じである必要はない。むしろ、概念的に関連するインスタンスおよび属性は、構造化コンポーネントを含む文書を識別するために使用されうる。

**【 0 3 2 9 】**

したがって、いくつかの実装では、プロセス3900を実行するシステムは、事前に存在している構造化プレゼンテーション内に見つかるような同じもしくは関係するインスタンス  
30  
識別子および/または事前に存在している構造化プレゼンテーション内に見つかるような同じもしくは関係する属性識別子を含む文書を識別することによって事前に存在している構造化プレゼンテーションにおいて指定されているインスタンスに関連する構造化コンポーネントを含む1つまたは複数の文書を識別することができる。

**【 0 3 3 0 】**

プロセス3900を実行するシステムは、構造化コンポーネントから1つまたは複数の属性提案を選択することができる(ステップ3910)。この選択プロセスでは、ユーザーに対して提案される属性を選別して個数を減らすことができる。属性提案の選択は、さまざまな方法で実行できる。例えば、システムは、以下でさらに説明するように、構造化コンポーネント内のインスタンスおよび/または構造化コンポーネント内のインスタンスの属性の値の  
40  
カテゴリに基づいて属性提案を選択することができる。

**【 0 3 3 1 】**

図40は、事前に存在している構造化プレゼンテーションの特性に基づく電子文書コレクション内の電子文書からの属性提案の構成を表す表現4000である。特に、表現4000は、プロセス3900(図39)の一実装を使用する属性提案の構成を示している。

**【 0 3 3 2 】**

図示されているように、事前に存在している構造化プレゼンテーションでは、インスタンス4005のコレクションを指定する(つまり、インスタンス「Philadelphia」および「Chicago」)。さらに、電子文書コレクション(例えば、コレクション102)内の異なる文書は、異なる構造化コンポーネント4010、4015、4020を含む。構造化コンポーネント4010、4015、4020は、例えば中に出現する同じインスタンス識別子「Philadelphia」および「Chicago」に基づいて指定されたインスタンス4005に関連するものとして識別されうる。  
50

## 【0333】

図示されているように、構造化コンポーネント4010、4015、4020は、異なる背景状況に基づくさまざまな異なる潜在的属性提案を含む。特に、構造化コンポーネント4010の背景状況では、インスタンス「Philadelphia」および「Chicago」は、さまざまな市の特徴を表す表形式コンポーネントの一部である。構造化コンポーネント4015の背景状況では、インスタンス「Philadelphia」および「Chicago」は、1970年代のある時期のナショナルリーグ東地区の順位表の一部を表す構造化コンポーネントの一部である。構造化コンポーネント4020の背景状況では、インスタンス「Philadelphia」および「Chicago」は、さまざまな動画の特徴を表す表形式コンポーネントの一部である。

## 【0334】

構造化コンポーネント4010、4015、4020内に見つかるさまざまな属性すべてをユーザーに対して提案する代わりに、それらのインスタンスを特徴付けるために使用される属性に基づいてコンポーネント4010、4015、4020から属性選択を選択することができる。特に、図示されているように、事前に存在している構造化プレゼンテーション106は、属性「year」、「rating」、および「box office receipts」の値を使用してインスタンス「Philadelphia」および「Chicago」を特徴付ける。構造化コンポーネント4010は、属性「population」および「area」の値を使用してインスタンス「Philadelphia」および「Chicago」を特徴付ける。構造化コンポーネント4015は、属性「wins」、「losses」、および「GB」(つまり、ゲーム差)の値を使用してインスタンス「Philadelphia」および「Chicago」を特徴付ける。構造化コンポーネント4020は、属性「year」、「runtime」、および「rating」の値を使用してインスタンス「Philadelphia」および「Chicago」を特徴付ける。

## 【0335】

システムは、これらの特徴付けられた属性に基づいて構造化コンポーネント4010、4015、4020内の属性から選択することができる。例えば、システムは、事前に存在している構造化プレゼンテーション106における属性識別子「year」および「rating」と構造化コンポーネント4020における属性識別子「year」および「rating」との間の対応関係を識別して、属性「director」および「runtime」を事前に存在している構造化プレゼンテーション106に追加する提案として選択することができる。

## 【0336】

図21~36および関連する本文において説明されているように、いくつかの実装では、システムは、追加のインスタンス識別子を提案もしくは追加することもできる。例えば、構造化コンポーネント4020は、インスタンス識別子「Peter Pan」および「Star Wars」を含む。したがって、このようなシステムは、これらのインスタンス識別子を構造化プレゼンテーションに含むように提案することができる。

## 【0337】

いくつかの実装では、構造化コンポーネント4010、4015から引き出された属性が特定の構成で提案されていない場合であっても、そのような属性は、将来の情報要求の際に使用できるように格納することができる。例えば、構造化コンポーネント4010で表されている市が、属性提案として選択されていないとしても、それらの市は、各属性識別子(例えば、「population」および「area」)およびデータコレクション(例えば、データセンター208など)内の属性値とともに格納することができる。その後ユーザーが1つまたは複数の市に関する情報を要求した場合、そのようなシステムは、この格納されている情報にアクセスして、追加の情報をユーザーに提供することができる。

## 【0338】

図41は、事前に存在している構造化プレゼンテーションの特性に基づいて電子文書コレクション内の電子文書から属性提案を構成するための例示的なプロセス4100の流れ図である。プロセス4100は、単独で、または他の活動と連携して実行されうる。例えば、プロセス4100は、プロセス3800(図38)におけるステップ3810で実行されうる。

## 【0339】

プロセス4100は、1つまたは複数の機械可読命令セットを実行することによってオペレ

10

20

30

40

50

ーションを実行する1つまたは複数のコンピュータのシステムによって実行されうる。例えば、プロセス4100は、システム200(図2)内の検索エンジン202によって実行されうる。

【0340】

プロセス4100を実行するシステムは、1つまたは複数の指定されたインスタンスに関連する1つまたは複数の文書を識別することができる(ステップ4105)。例えば、プロセス4100を実行するシステムは、文字列比較を使用することで、指定されたインスタンスおよびその属性および/または値のうちの1つまたは複数と電子文書コレクション102などの電子文書コレクション内の文書とのマッチングを行うことができる。他の例では、プロセス4100を実行するシステムは、格納されている情報(例えば、データセンター208内の情報)にアクセスして、指定されたインスタンスに関連する電子文書を識別することができる。

10

【0341】

プロセス4100を実行するシステムは、識別された文書のうちの1つまたは複数の文書のテンプレートを抽出することができる(ステップ4110)。文書テンプレートは、電子文書コレクション内の文書のサブコレクション内の個別の文書のコンテンツの配列に対するパターンとして使用される。サブコレクション内の文書は、一般的に、単一のソース、例えば、単一の企業に由来するものである。例えば、書籍販売店は、単一の文書テンプレートを、異なる書籍を記述したコンテンツの配列に対するパターンとして使用することができる。他の例では、家具小売店は、単一の文書テンプレートを、異なるソファのパフレットのコンテンツの配列に対するパターンとして使用することができる。例えば、ソファ用の電子パフレットのテンプレートにより、ソファのブランド名、ソファの写真、ソファが示されている色をユーザーが選択することを可能にするインタラクティブ要素、テキスト形式のソファの説明、ならびにソファの寸法、入手状況、および価格を特徴付けるテーブルの配列をパフレット上に指定することができる。こうして、文書テンプレートは、構造化プレゼンテーションとほぼ同様にし、定義済みの構造と一致するようにインスタンスに関する情報を編成することができる。

20

【0342】

一般に、文書テンプレートは、電子文書のコンテンツ全体に対するパターンとして使用することができる。上述のように、文書内の構造化コンポーネントの配列を指定することさえできる。しかし、文書テンプレートでは非構造化電子文書コレクション内の文書のサブコレクションのコンテンツの配列を指定するだけなので、電子文書コレクションそれ自体は非構造化のままである。例えば、AMAZON.COMでは1つのテンプレートをAMAZON.COMが販売するすべての書籍の説明の配列に対するパターンとして使用する場合であっても、BARNESANDNOBLE.COMおよび他の書籍販売店は、その同じテンプレートを書籍店が販売する書籍の説明の配列に対するパターンとして必ずしも使用しない。

30

【0343】

さらに、文書テンプレートは、その変更を必ずしも電子文書のコレクション全体、さらにはサブコレクションに伝搬させなくても変更できる。

【0344】

図42は、動画(つまり、動画「Philadelphia」)を記述するためのパターンとして使用されるハイパーテキストマークアップ言語(HTML)のテンプレート4200の一部を表す図である。テンプレート部分4200のハイパーテキストマークアップ言語(HTML)コードは、機械可読かつ人間可読である。例えば、ブラウザでテンプレート部分4200のHTMLコードを使用して、ウェブページを生成することができる。

40

【0345】

示されている例では、テンプレート部分4200は、2つのサブセクション4205、4210に分割される。サブセクション4205は、動画「Philadelphia」を識別するテキストの配列に対するパターンとして使用される。サブセクション4210は、さまざまな属性識別子およびその値の配列に対するパターンとして使用される。一般に、異なる動画を記述するために電子文書内の文書の特定のサブコレクションにおいて、サブセクション4205、4210内のパターンを何回も繰り返す。

50

## 【 0 3 4 6 】

HTMLパーサーを使用して、テンプレート部分4200から出力書式を抽出することができ、したがって、出力書式を使用して、同じテンプレートを有する文書を識別することができる。例えば、HTMLタグ<title>、<div>、他のHTMLタグ、および互いに相対的な位置は、HTMLパーサーによって識別されうる。このようなHTMLパーサーは、HTMLタグ<title>がHTMLタグ<div>の前に出現するものとして判定することができる。したがって、HTMLパーサーは、テンプレートに従って配列されているコンテンツからのテンプレート部分4200から出力書式を抽出することができる。

## 【 0 3 4 7 】

再び図41を参照すると、テンプレートを抽出した後に、プロセス4100を実行するシステムは、テンプレートを使用して1つまたは複数の新規属性を識別することができる(ステップ4115)。例えば、システムは、そのテンプレート内の事前に存在している構造化表示から引き出された属性の配列を識別することができる。この配列は、他の属性を推論するために使用できる。システム4100を実行するシステムは、そのテンプレートを使用して識別される属性から1つまたは複数の属性提案を構成することもできる(ステップ4120)。したがって、これらのテンプレートそれ自体を使用して、属性提案を構成することができる。例えば、テンプレート部分4200(図42)内のHTMLタグは、映画に題名が付けられていることを識別する。

10

## 【 0 3 4 8 】

「Philadelphia (1993)」は、属性「Director」、「Writer」、および「Release Date」によって特徴付けられる。これらの属性のどれかを使用して、1つまたは複数の属性提案を構成することができる。

20

## 【 0 3 4 9 】

さらに、いくつかの実装では、文書テンプレート内の追加のコンテンツは、属性提案を構成する際に使用することができる。例えば、属性の値は、属性提案を構成する際に使用することができる。例えば、属性「year」の値が、例えば、1976である場合、属性「start time」は、映画を特徴付けるために属性提案のコレクションから除外されうる。

## 【 0 3 5 0 】

図43は、事前に存在している構造化プレゼンテーション106内の情報に基づいて新規属性提案3715のコレクションを構成することができるプロセス4300の略図である。プロセス4300は、1つまたは複数の機械可読命令セットを実行することによってオペレーションを実行する1つまたは複数のコンピュータのシステム、例えば、システム200(図2)によって実行されうる。

30

## 【 0 3 5 1 】

プロセス4300は、事前に存在している構造化プレゼンテーション106内の情報に基づきインスタンス/属性コレクション4310に対し抽出オペレーション4305を実行する。インスタンス/属性コレクション4310は、インスタンスをその属性に、またいくつかの実装では、それらの属性の値にも関連付ける情報のコレクションである。コレクション4310内の情報は、トリガーを受け取ったことに対する応答として(例えば、検索クエリ)、またはトリガーを受け取ることを予想して、例えば、電子文書コレクション102のインデックスを作成するプロセスの一部として、電子文書コレクション102内の文書から抽出されうる。いくつかの実装では、コレクション4310内の情報は、現在のユーザーまたは他のユーザーに対して提示された前の構造化プレゼンテーションのコンテンツを含むことができる。一般に、属性提案は、ステップ3815、3820、3825(図38)において説明されているような、構造化プレゼンテーションに追加される属性提案を選択するユーザーに提供される。

40

## 【 0 3 5 2 】

インスタンスとその属性の間の関連付けは、コレクション4310内の情報記憶域を構造化することによってコレクション4310内で確立されうる。図44は、コレクション4310内の属性およびインスタンスを関連付けるテーブル4400の略図である。テーブル4400は、レコード4402、4404、4406、4408、4410、4412、4414のコレクションを含み、それぞれのレコー

50

ドはインスタンスの識別子を文書のロケーションの記述およびそれらの文書内の識別されたインスタンスを特徴付ける属性に関連付ける。レコード4402、4404、4406、4408、4410、4412、4414内の情報は、列4415、4420、4425、4430、4435、4440のコレクション内に編成されうる。特に、列4415は、インスタンス識別子を含むことができる。列4420は、列4415で識別されているインスタンスを含む電子文書のロケーションの記述を含むことができる。列4425、4430、4435、4440は、列4420内にロケーションが記述されている文書内の列4415で識別されるインスタンスを特徴付ける属性を識別することができる。

**【 0 3 5 3 】**

図示されているように、異なる電子文書は、同じインスタンスを特徴付ける異なるカテゴリおよび量の情報を含むことができる。例えば、ロケーションがレコード4404の列4420で識別される文書は、インスタンス「INSTANCE\_2」の2つの属性を含むが、ロケーションがレコード4410の列4420で識別される文書は、インスタンス「INSTANCE\_2」の3つの属性を含む。さらに、レコード4404内の属性(つまり、属性「ATTR\_5」および属性「ATTR\_6」)は、レコード4410内の属性(つまり、属性「ATTR\_5」、属性「ATTR\_8」、および属性「ATTR\_9」)と一部異なる。

10

**【 0 3 5 4 】**

属性およびインスタンスを関連付けるデータコレクション4310(テーブル4400など)は、いくつもの異なる方法で形成されうる。例えば、内部的な構造化コンポーネントを含む文書が識別されうる。このような内部的な構造化コンポーネントの例として、HTML文書中に出現するテーブルおよびリストが挙げられる。これらの内部的な構造化コンポーネント内の属性とインスタンスとの間の関係をコピーしてデータコレクション4310を形成することができる。

20

**【 0 3 5 5 】**

他の例として、コレクション4310は、現在のユーザーまたは他のユーザーに対して提示された前の構造化プレゼンテーションのコンテンツから形成することができる。

**【 0 3 5 6 】**

さらに他の例として、第1の文書が属性およびインスタンスを含むものとして識別された後、その文書のテンプレートを使用して、同じテンプレートを含む他の文書から属性およびインスタンスを抽出することができる。例えば、ステレオ小売業者が同じ文書テンプレートを使用して販売のため提供される異なるステレオを記述する場合、第1のステレオに関する第1の電子文書内の情報の配列を使用して、他のステレオに関する他の電子文書から情報を抽出することができる。

30

**【 0 3 5 7 】**

いくつかの実装では、自然言語構文解析などの技術を使用して、インスタンスおよび属性を識別することができる。例えば、電子文書の構文解析を行って、電子文書内の「[Instance] has a/an [attribute]」などのフレーズを識別することができる。

**【 0 3 5 8 】**

いくつかの実装では、データコレクション4310は、インスタンスおよびその属性を分類することができる。例えば、インスタンスおよび属性は、北米の都市、ナショナルリーグ東地区チーム、または人気のある映画としてカテゴリ化することができる。データコレクション4310への情報の格納は、そのような分類に基づくことができる。例えば、異なるカテゴリを異なるファイル、レコード、および同様のものに格納することができる。

40

**【 0 3 5 9 】**

図43を再び参照すると、プロセス4300は、事前に存在している構造化プレゼンテーション106で提示される情報に基づき1つまたは複数の新規属性を提案している。例えば、構造化プレゼンテーションがいくつかの動画に対応する多数のインスタンスを含む場合、システム200は、データコレクション4310から引き出された情報に従って動画の追加の属性を提案することができる。つまり、システム200は、同じインスタンスを特徴付けるために使用される属性に基づいて追加の属性を識別し、提案することができる。例えば、システム200は、ショータイム、劇場、または放映時間などの、動画を特徴付けるために通常使

50

用される他の属性を提案することができる。

【0360】

図45は、事前に存在している構造化プレゼンテーションの特性に基づいてインスタンスおよび属性のコレクションから属性提案を構成するためのプロセス4500の流れ図である。プロセス4500は、1つまたは複数の機械可読命令セットを実行することによってオペレーションを実行する1つまたは複数のコンピュータのシステムによって実行されうる。例えば、プロセス4500は、システム200(図2)内の検索エンジン202によって実行されうる。

【0361】

プロセス4500は、単独で、または他の活動と連携して実行されうる。例えば、プロセス4500は、プロセス4300(図43)で実行されうる。他の例として、プロセス4500は、プロセス3800(図38)におけるステップ3810で、単独で、またはプロセス3900、4100(図39、41)の一方または両方と連携して実行されうる。

【0362】

プロセス4500を実行するシステムは、インスタンスおよびその属性のコレクションにアクセスすることができる(ステップ4505)。例えば、プロセス4500を実行するシステムは、データセンター208(図2)に格納されているインスタンスおよび属性コレクション4310(図43)にアクセスすることができる。

【0363】

プロセス4500を実行するシステムは、事前に存在している構造化プレゼンテーションにおいて指定されているインスタンス属性の特性に基づいて1つまたは複数の関連する属性を識別することができる(ステップ4510)。例えば、システムは、インスタンスおよび属性コレクション4310(図43)内に格納されているインスタンスのインスタンス属性を構造化プレゼンテーション106で指定されているインスタンスと比較することができる。システムは、この比較結果を使用して、もしあれば格納されているインスタンスのうちのどれが属性識別子、または関係する属性を事前に存在している構造化プレゼンテーション106で指定されている属性と共有するかを決定することができる。例えば、事前に存在している構造化プレゼンテーション106は、属性「ATTR\_3」および「ATTR\_5」を使用してインスタンスのコレクションを特徴付けると推定する。テーブル4400(図44)などのインスタンスおよび属性コレクション4310を検討した後、システムは、属性「ATTR\_7」および「ATTR\_7」を、レコード4402、4406内のインスタンス「INSTANCE\_1」および「INSTANCE\_2」を特徴付ける際に「ATTR\_3」および「ATTR\_5」と連携して使用することに基づいて、提案することができる。

【0364】

図46は、事前に存在している構造化プレゼンテーション106内の情報に基づいて新規属性提案3715のコレクションを構成するためのプロセス4600の流れ図である。プロセス4600は、1つまたは複数の機械可読命令セットを実行することによってオペレーションを実行する1つまたは複数のコンピュータのシステム、例えば、システム200(図2)によって実行されうる。プロセス4600は、単独で、または他の活動と連携して実行されうる。例えば、プロセス4600は、プロセス4300(図43)で実行されうる。他の例として、プロセス4600は、プロセス3800(図38)におけるステップ3810で、単独で、またはプロセス3900、4100、4500(図39、41、45)のうちの1つまたは複数と連携して実行されうる。いくつかの実装では、プロセス4600は、入力を(例えば、システム200のユーザーから、または他のシステムもしくはプロセスから)受け取ったことに応答して実行されうる。例えば、プロセス4600は、検索文字列を受け取ったことに応答して検索エンジン202によって実行されうる。

【0365】

プロセス4600を実行するシステムは、1つまたは複数の指定されたインスタンスに関する1つまたは複数の信頼できるソースを識別することができる(ステップ4605)。例えば、システムは、例えばプログラマーによってアSEMBルされた電子文書コレクション102内の文書の信頼できるソースのコレクションにアクセスすることができる。

【0366】

他の例として、システムは、電子文書コレクション102内の文書の1つまたは複数の信頼できるソースをそのユーザーの観点から「信頼できる」ものとして識別するユーザー特有の入力を受け取ることができる。例えば、事前に存在している構造化プレゼンテーション106を表示するディスプレイ画面104は、閲覧者が文書の信頼できるソースを指定することを可能にするGUIコンポーネントを備えることができる。信頼できるソースの識別は、検索クエリと連動して受け取ることができる。例えば、閲覧者は、JD POWER AND ASSOCIATE S、AMAZON.COM、およびMAJOR LEAGUE BASEBALLをそれぞれ<http://www.jdpower.com/>、<http://www.amazon.com/>、および<http://www.mlb.com/>にある文書の信頼できるソースとして識別することができる。いくつかの実装では、ユーザー特有の入力は、ソースが信頼できるものである主題を識別することができる。例えば、MAJOR LEAGUE BASEBALLは、ベースボールの統計の信頼できるソースとして識別されうるが、薬物検査に関する情報の信頼できるソースとしてはみなされえない。

10

## 【0367】

さらに他の例として、プロセス4600を実行するシステムは、信頼できるソースの「一般」コレクションをアSEMBルするために複数のユーザーからの信頼できるソースを識別するユーザー特有の入力のコレクションを分析することができる。例えば、多数のユーザーが、AMERICAN AUTOMOBILE ASSOCIATION (AAA)を信頼できるものとして識別することができる。次いで、これらの識別結果の統計分析に基づき、AAAを信頼できるソースのコレクションに追加することができる。

## 【0368】

20

プロセス4600を実行するシステムは、事前に存在している構造化プレゼンテーションにおいて指定されているインスタンスに対する信頼できるソースから追加の属性を決定することができる(ステップ4610)。例えば、システムは、信頼できるソースによって提供される文書にアクセスし、1つまたは複数の属性を使用して指定されたインスタンスを特徴付ける1つまたは複数の文書を識別することができる。システムは、パーサーまたは他の文字列比較技術を使用してこれらの文書から属性識別子を抽出することができる。

## 【0369】

他の例として、システムは、テーブル4400(図44)などの属性およびインスタンスを関連付けるデータコレクションにアクセスすることができる。システムは、レコード4402、4404、4406、4408、4410、4412、4414などのレコードを、事前に存在している構造化プレゼンテーションで識別されたインスタンスおよびロケーションがレコード4402、4404、4406、4408、4410、4412、4414で識別される文書が信頼できるソースに由来するものであったかどうかの両方に基づいてフィルター処理することができる。例えば、AMAZON.COMが信頼できるソースである場合、属性およびインスタンスを関連付けるコレクションをスキャンして、<http://www.amazon.com/>ドメインを持つ文書を識別することができる。

30

## 【0370】

プロセス4600を実行するシステムは、これらの追加のインスタンスをテーブル4400(図44)などのインスタンスおよび属性コレクション内の属性と比較することができる(ステップ4615)。例えば、システムは、文字列比較、または他の比較技術を使用して、追加のインスタンスをインスタンスおよび属性コレクションに格納されているインスタンスと比較することができる。

40

## 【0371】

プロセス4600を実行するシステムは、これらの比較結果に基づいてインスタンスおよび属性コレクション内の属性を識別することができる(ステップ4620)。

## 【0372】

図47は、事前に存在している構造化プレゼンテーション106内の情報に基づいて属性提案を構成する際に使用する関係するインスタンスを識別するためのプロセス4700の流れ図である。プロセス4700は、1つまたは複数の機械可読命令セットを実行することによってオペレーションを実行する1つまたは複数のコンピュータのシステム、例えば、システム200(図2)によって実行されうる。プロセス4700は、単独で、または他の活動と連携して実

50

行されうる。例えば、プロセス4700は、プロセス1500(図15)で実行されうる。他の例として、プロセス4700は、プロセス3800(図38)におけるステップ3810で、単独で、またはプロセス3900、4100、4500、4600(図39、41、45、46)のうちの1つまたは複数と連携して実行されうる。いくつかの実装では、プロセス4700は、入力を(例えば、システム200のユーザーから、または他のシステムもしくはプロセスから)受け取ったことに応答して実行されうる。例えば、プロセス4700は、検索文字列を受け取ったことに応答して検索エンジン202によって実行されうる。

【0373】

プロセス4700を実行するシステムは、事前に存在している構造化プレゼンテーションにおいて指定されているインスタンスを特徴付ける属性および/または属性値に基づいて1つまたは複数の関係するインスタンスを識別することができる。例えば、システムは、指定されたインスタンスの属性値を他のインスタンスの属性値と比較することによって関係するインスタンスを識別し、他のインスタンスが関係しているかどうかを判定することができる。このような比較では、例えば、属性値が同一のものである、または属性値が特定の範囲内にあるということが要求される場合がある。このような範囲は、例えば、インタラクティブ要素を介してユーザーによって指定されるか、または事前に存在している構造化プレゼンテーション内のインスタンスをすでに特徴付けている属性値の範囲に基づいて決定することができる。

10

【0374】

いくつかの実装では、プロセス4700を実行するシステムは、属性値を比較する前に属性値を通常の測定単位に変換することができる。例えば、指定された測定単位がフィートであるが、1つまたは複数の抽出された属性値の測定単位がメートルである場合、システムは、通常の技術を使用してメートル単位の値をフィートに変換することができる。

20

【0375】

図48は、事前に存在している構造化プレゼンテーション106内の情報に基づいて新規属性提案3715のコレクションを構成するためのプロセス4800の流れ図である。プロセス4800は、1つまたは複数の機械可読命令セットを実行することによってオペレーションを実行する1つまたは複数のコンピュータのシステム、例えば、システム200(図2)によって実行されうる。プロセス4800は、単独で、または他の活動と連携して実行されうる。例えば、プロセス4800は、プロセス1500(図15)で実行されうる。他の例として、プロセス4800は、プロセス3800(図38)におけるステップ3810で、単独で、またはプロセス3900、4100、4500、4600、4700(図39、41、45、46、47)のうちの1つまたは複数と連携して実行されうる。いくつかの実装では、プロセス4800は、入力を(例えば、システム200のユーザーから、または他のシステムもしくはプロセスから)受け取ったことに応答して実行されうる。例えば、プロセス4800は、検索文字列を受け取ったことに応答して検索エンジン202によって実行されうる。

30

【0376】

プロセス4800を実行するシステムは、インスタンスおよび属性のカテゴリ化されたコレクションにアクセスすることができる(ステップ4805)。例えば、システムは、前の検索時に生成されたインスタンスおよび属性の1つまたは複数のカテゴリ化されたコレクションにアクセスするために、インスタンスおよび属性コレクション1510にアクセスすることができる。

40

【0377】

プロセス4800を実行するシステムは、指定されたインスタンスを含むカテゴリを識別することができる(ステップ4810)。いくつかの実装では、システムは、類似の属性、類似の属性値、これらの特性の組み合わせ、および/または他の技術に基づくインスタンスを含むカテゴリを識別することができる。

【0378】

プロセス4800を実行するシステムは、識別されたカテゴリから1つまたは複数の属性提案を選択することができる(ステップ4815)。例えば、いくつかの実装では、属性提案を識

50

別されたカテゴリから、そのカテゴリ内のインスタンスを特徴付けるためにそれらの属性が使用される回数に基づいて選択することができる。

【0379】

図49は、事前に存在している構造化プレゼンテーションの特性に基づく電子文書コレクション内の電子文書からの属性提案の構成を表す表現4900である。特に、表現4900は、プロセス4800(図48)の一実装を使用する属性提案の構成を示している。

【0380】

図示されているように、事前に存在している構造化プレゼンテーションでは、インスタンス4005のコレクションを指定する(つまり、インスタンス「Philadelphia」および「Chicago」)。さらに、電子文書コレクション(例えば、コレクション102)内の異なる文書から引き出されたインスタンスは、異なるコレクション4910、4915、4920にすでにカテゴリ化されている。カテゴリ化されたインスタンスコレクション4910、4915、4920は、例えば中に出現する同じインスタンス識別子「Philadelphia」および「Chicago」に基づいて指定されたインスタンス4005に関連するものとして識別されうる。

【0381】

図示されているように、カテゴリ化されたインスタンスコレクション4910、4915、4920は、さまざまな異なる方法でカテゴリ化されている。特に、カテゴリ化されたインスタンスコレクション4910は、「North American Cities」のコレクションとしてカテゴリ化されている。カテゴリ化されたインスタンスコレクション4915は、「National League East Teams」のコレクションとしてカテゴリ化されている。カテゴリ化されたインスタンスコレクション4920は、「Popular Movies」のコレクションとしてカテゴリ化されている。システム200の背景状況において、カテゴリ化されたインスタンスコレクション4910、4915、4920は、データセンター208(図2)に格納することができる。つまり、システム200は、すでに受け取っている検索文字列に基づいてインスタンスの1つまたは複数のカテゴリを生成することができる。したがって、検索エンジン202が、検索文字列に基づいて検索を実行した後、検索エンジン202は、結果をカテゴリ化し、それらをデータセンター208内に格納することができる。その後の検索時に、これらのカテゴリ化された結果にアクセスし、分析して、属性提案を生成することができる。

【0382】

事前に存在している構造化プレゼンテーションで指定されているインスタンスを含むカテゴリ化されたインスタンスコレクションは、例えば、指定されたインスタンスを特徴付ける属性と異なるカテゴリに含まれるインスタンスを特徴付ける属性との間の類似度に基づいて識別されうる。例えば、属性「year」および「rating」の一般的な使用では、事前に存在している構造化プレゼンテーションおよびカテゴリ化されたインスタンスコレクション4920を使用して、カテゴリ化されたインスタンスコレクション4920がインスタンス4005を含むものとして識別することができる。

【0383】

いくつかの実装では、カテゴリ化されたインスタンスコレクション内の属性のサブセットは、カテゴリに含まれるインスタンスを特徴付ける属性に基づいて属性提案として選択されうる。例えば、属性「Start time」を使用して動画インスタンスを特徴付けることは、現在再生中の動画に関する情報のみが構造化表示に含まれるという指示としてみなすことができる。したがって、「playing at」および「coupons available」などの属性は、属性提案のリストに含まれうる。他の例として、属性「year」は、属性「Start time」を使用して事前に存在している構造化表示における動画インスタンスを特徴付けることに基づき属性提案のリストから除外することができる。

【0384】

図50は、事前に存在している構造化プレゼンテーション106内の情報に基づいて新規属性提案3715のコレクションを構成するために使用されうるプロセスのコレクション5000の略図である。

【0385】

10

20

30

40

50

コレクション5000内のプロセスは、潜在的属性5005の大きなコレクションに連続的に適用され、属性提案のより小さなコレクション5010を生成するフィルターとみなすことができる。それぞれのフィルターは、ユーザーに提示されるか、または構造化プレゼンテーションに直接追加されうる属性提案コレクション5010から潜在的属性5005を除外することができる。

**【0386】**

フィルター5015、5020、5025は、任意の順序で適用できる。しかし、一般に、フィルター5015、5020、5025は、粒度の順序で適用される。特に、潜在的属性の数を最大量減らすフィルター5015、5020、5025が最初に適用され、潜在的属性の数を最小量減らすフィルター5015、5020、5025が最後に適用される。

10

**【0387】**

いくつかの実装では、フィルター5015、5020、5025のうちのどれかをコレクション5000から省き、および/または追加のフィルターをコレクション5000に追加することができる。例えば、ユーザーによって供給される入力に従って潜在的属性5005をフィルター処理することができるユーザー指定フィルターをコレクション5000に追加することができる。

**【0388】**

図51は、事前に存在している構造化プレゼンテーション106内の情報に基づいて新規属性提案3715のコレクションを構成するためのプロセス5100の流れ図である。プロセス5100は、1つまたは複数の機械可読命令セットを実行することによってオペレーションを実行する1つまたは複数のコンピュータのシステム、例えば、システム200(図2)によって実行されうる。プロセス5100は、単独で、または他の活動と連携して実行されうる。例えば、プロセス5100は、プロセスコレクション5000におけるプロセスの1つまたは複数と連携して反復実行されうる。いくつかの実装では、プロセス5100は検索文字列を受け取ったことに応答して実行されうる。

20

**【0389】**

プロセス5100を実行するシステムは、事前に存在している構造化プレゼンテーション106で指定されたインスタンスと文書コレクションから引き出された属性との間の初期マッチングを行うことができる(ステップ5105)。初期マッチングは、プロセスコレクション5000内のフィルター処理プロセスのうちの1つまたは複数に基づくことができる。

30

**【0390】**

プロセス5100を実行するシステムは、マッチングの数が大きすぎるか、小さすぎるか、または適切であるかを判定することができる(ステップ5110)。マッチングの数が少なすぎる場合、システムは、指定されたインスタンスの名前を広げることができる(ステップ5115)。例えば、プロセス5100を実行するシステムは、プロセスコレクション5000内のプロセスのうちの1つまたは複数で指定されているインスタンスについて代替スペリング、略語、同義語、代替名、ニックネーム、および/または他のキーワードを使用することができる。

**【0391】**

プロセス5100を実行するシステムは、関係するインスタンスおよび/または属性フィルター処理5020(図50)で使用される属性および/またはインスタンスの数を減らすこともできる(ステップ5120)。属性および/またはインスタンスの数は、例えば、特定の属性および/またはインスタンスによって除外される潜在的インスタンスの数に基づいて減らすことができる。例えば、特定のインスタンスが属性によって特徴付けられるという要求条件により潜在的属性のすべてが除外される場合、このインスタンスは、関係するインスタンスおよび/または属性フィルター処理から省くことができる。除去すべき属性および/またはインスタンスは、例えば、自動的に、ユーザー入力なしで、またはユーザーから受け取った入力に基づいて決定されうる。

40

**【0392】**

ここでもまた、システムは、事前に存在している構造化プレゼンテーション106で指定されたインスタンスと文書コレクションから引き出されたインスタンスとの間のマッチン

50

グを行おうとすることができるが、このときには変更されたパラメータを使用する(ステップ5122)。このマッチングは、プロセスコレクション5000内のフィルター処理プロセスのうちの1つまたは複数を使用して行うこともできる。

【0393】

マッチングの数が多すぎると判定された場合(ステップ5110)、プロセス5100を実行するシステムは、関係する属性および/またはインスタンスフィルター処理5020(図50)で使用される属性および/またはインスタンスの数を増やすことができる(ステップ5128)。属性および/またはインスタンスの数は、例えば、特定の属性および/またはインスタンスによって除外される潜在的属性の数に基づいて増やすことができる。追加すべき属性および/またはインスタンスは、例えば、自動的に、ユーザー入力なしで、またはユーザーから受け取った入力に基づいて決定されうる。例えば、追加するインスタンスは、プロセス4800(図48)を使用して決定することができる。

10

【0394】

プロセス5100を実行するシステムは、変更されたパラメータに基づいてマッチングを選別することができる(ステップ5130)。特に、インスタンスの狭められた範囲および/または増やした数を関係する値フィルター処理5025(図50)で使用することができる。

【0395】

マッチングの数が受け入れられると判定された場合(ステップ5110)、プロセス5100を実行するシステムは、マッチした属性をユーザーに提案することができる(ステップ5135)。例えば、プロセス5100を実行するシステムは、ディスプレイ画面、例えば、ディスプレイ画面104上のGUIで1つまたは複数の属性提案を提示することができる。

20

【0396】

図52は、電子文書コレクション内の2つまたはそれ以上の電子文書から引き出される属性値307が構造化プレゼンテーションでユーザーに提示されるシステム5200を示す略図である。電子文書コレクション102、ディスプレイ画面104、およびデータ通信経路108に加えて、システム5200は、構造化データ5205およびマージモジュール5210を備える。動作中、システム5200は、電子文書コレクション102における電子文書の非構造化コレクションから属性値を抽出し、その情報を構造化データ5205から引き出された情報とマージして構造化プレゼンテーション106に書き込む。

【0397】

30

システム5200は、構造化プレゼンテーション106のすべてまたは一部にのみ、属性値を書き込むことができる。構造化プレゼンテーションの一部にのみ属性値が書き込まれる状況は多数ある。例えば、このような書き込みは、新規インスタンス(したがって新規構造化レコード)を構造化プレゼンテーション106に追加する一環として行われうる。他の例として、このような書き込みは、新規属性を構造化プレゼンテーション106に追加する一環として行われうる。さらに他の例として、このような書き込みは、構造化プレゼンテーション106内で既存の属性値の一部の改善の一環とすることができる。例えば、オリジナルの属性値のある部分は、正確かどうかをチェックしたり、または特徴付けられているインスタンスが変更されていないことを確認するためにチェックすることができる。

【0398】

40

構造化データ5205は、情報の構造化コレクションである。構造化データ5205内の情報は、定義済みデータモデルに従って編成されうる。例えば、構造化データ5205は、階層または関係データモデルに従って編成され、データ記憶デバイスに格納されうる。いくつかの場合において、構造化データ5205の全部または一部を構造化プレゼンテーションでユーザーに提示することができる。例えば、いくつかの実装では、構造化データ5205は、コレクション102から引き出された新規属性値が書き込まれた構造化プレゼンテーション106を提示する同じディスプレイ画面104上でユーザーに対し提示される事前に存在している構造化プレゼンテーション106とすることができる。

【0399】

マージモジュール5210は、1つまたは複数のデータ処理デバイス上に配備される1つまた

50

は複数の機械可読命令セットのコレクションである。マージモジュール5210は、新規属性値を識別するとともに、書き込まれる構造化プレゼンテーション106内に値を配置するための機能を備えることができる。マージモジュール5210によって実行されるオペレーションについては、以下でさらに詳しく説明する。

**【0400】**

図53は、電子文書コレクション102内の2つまたはそれ以上の電子文書から引き出される属性値が構造化プレゼンテーションでユーザーに提示されるシステム5300の一実装を示す略図である。電子文書コレクション102、ディスプレイ画面104、データ通信経路108、検索エンジン202、クローラー204、ならびに圧縮、インデックス付け、および順位付けモジュール210に加えて、システム5300は、データセンター208に属性/値/インスタンスコレクション5305も備える。

10

**【0401】**

属性/値/インスタンスコレクション5305は、インスタンスをその属性に、さらにそれらの属性の値にも関連付ける情報のコレクションである。コレクション5305内の情報は、トリガーを受け取ったことに対する応答として(例えば、検索クエリ)、またはトリガーを受け取ることを予想して、例えば、電子文書コレクション102のインデックスを作成するプロセスの一部として、コレクション5305内の電子文書から抽出されうる。

**【0402】**

インスタンス、属性、およびその値の間の関連付けは、コレクション5305内の情報記憶域を構造化することによってコレクション5305内で確立されうる。例えば、図54は、コレクション5305(図53)内の属性、値、およびインスタンスを関連付けることができるテーブル5400の略図である。テーブル5400は、レコード5402、5404、5406、5408、5410のコレクションを含み、それぞれのレコードはインスタンスの識別子を文書のロケーションの記述および識別されたインスタンスを特徴付ける属性および値に関連付ける。レコード5402、5404、5406、5408、5410内の情報は、列5415、5420、5425、5430、5435、5440、5445、5450のコレクション内に編成されうる。特に、列5415は、インスタンス識別子を含むことができる。列5420は、列5415で識別されているインスタンスを含む電子文書のロケーションの記述を含むことができる。列5425、5435、5445は、列5415で識別されるインスタンスを特徴付けるとともに、列5420内にロケーションが記述されている文書内の値でそれ自体特徴付けられる属性の記述を含むことができる。列5430、5440、5450は、列5420内にロケーションが記述されている文書内の列5415で識別されるインスタンスの列5425、5435、5445に記述されている属性を特徴付ける値の記述を含むことができる。

20

30

**【0403】**

図示されているように、異なる電子文書は、同じインスタンスを特徴付ける異なるカテゴリおよび量の情報を含むことができる。例えば、ロケーションがレコード5404の列5420で識別される文書は、インスタンス「INSTANCE\_1」の3つの属性を特徴付ける情報を含むが、ロケーションがレコード5406の列5420で識別される文書は、インスタンス「INSTANCE\_1」の2つの属性を特徴付ける情報を含む。さらに、レコード5404で特徴付けられる属性(つまり、属性「ATTR\_5」、属性「ATTR\_6」、属性「ATTR\_7」)は、レコード5406で特徴付けられる属性(つまり、属性「ATTR\_3」、属性「ATTR\_4」)と異なる。

40

**【0404】**

さらに、たとえ同じエンティティの同じ属性を特徴付けるために使用される値でも、異なる電子文書内では異なることがある。例えば、ロケーションがレコード5402の列5420で識別される文書は、インスタンス「INSTANCE\_1」の属性「ATTR\_3」を特徴付ける値「VALUE\_3A」を含むが、ロケーションがレコード5406の列5420で識別される文書は、同じインスタンス「INSTANCE\_1」の同じ属性「ATTR\_3」を特徴付ける値「VALUE\_3B」を含む。

**【0405】**

異なる文書において同じエンティティの同じ属性を特徴付ける値の間のそのような不一致の異なる潜在的発生源は多数ある。例えば、文書は、エンティティの属性を間違っ

50

更される可能性もある。これの例としては、例えば、ハイスクールのバスケットボール選手のインスタンスの「height」属性の値、家のインスタンスの「list price」属性の値、市のインスタンスの「mayor」属性の値が挙げられる。正しい値で更新することができる文書があるが、オリジナルであるが現在は不正な値を保持する文書もある。

**【0406】**

さらに、完全に正確な文書であっても、異なる方法で同じエンティティの同じ属性を特徴付ける場合がある。例えば、異なる文書は、異なる単位を使用して、同じ値を表現することができる。他の例として、異なる文書は、異なる精度で同じ値を表すことができる(例えば、「about a two hour drive to Phoenix」(フェニックスまで車で約2時間)と「a 130 minute drive to Phoenix at the posted speed limits」(標識に示されている制限速度でフェニックスまで車で130分))。

10

**【0407】**

そのような不一致は、非構造化電子文書コレクション、例えば、文書コレクション102の背景状況に本質的に固有なものである。この点に関して、上で説明したように、情報を異なる仕方で提示する異なるユーザーが異なる文書をコレクション102に追加することができる。文書を追加するユーザーは、情報が矛盾なく提示されるように共同作業をすることはせず、また異なる文書内の情報の提示が変更を受けないままであることを保証するための正式なメカニズムもない。

**【0408】**

属性、値、およびインスタンスを関連付けるデータコレクション5305(例えば、テーブル5400)は、いくつもの異なる方法で形成されうる。例えば、内部的な構造化コンポーネントを含む文書が識別されうる。このような内部的な構造化コンポーネントの例として、HTML文書中に出現するテーブルおよびリスト、および同様のものが挙げられる。これらの内部的な構造化コンポーネント内の属性、値、およびインスタンスの間の関係をコピーしてデータコレクション5305を形成することができる。

20

**【0409】**

他の例として、第1の文書が属性、値、および/またはインスタンスを含むものとして識別された後、その文書のテンプレートを使用して、同じテンプレートを含む他の文書から属性、値、およびインスタンスを抽出することができる。例えば、ステレオ小売業者が同じ文書テンプレートを使用して販売のため提供される異なるステレオを記述する場合、第1のステレオに関する第1の電子文書内の情報の配列を使用して、他のステレオに関する他の電子文書から情報を抽出することができる。

30

**【0410】**

さらに他の例では、単一の文書のテンプレートを使用して、その文書から属性、値、および/またはインスタンスを抽出することができる。例えば、テンプレートにより、複数の属性およびそれらの属性を特徴付ける値の配列をインスタンスの識別子に関して指定することができる。これらの属性および/または値のいくつかは知られている場合、それらの知られている属性および/または値の配列を識別し、使用して、他の属性および/または値を識別することができる。

**【0411】**

40

例えば、単一のウェブページのテンプレートにより、動画インスタンス「Citizen Kane」の識別子に関して属性/値の対の配列「Director: Orson Welles」、「Writer: Orson Welles, Herman J. Mankiewicz」、および「Release Date: May 1, 1941」を指定することができる。属性と値「Director: Orson Welles」と「Release Date: May 1, 1941」がすでに知られていた場合、動画インスタンス識別子「Citizen Kane」に関するこれらの属性および値の配列を使用して、属性/値の対「Writer: Orson Welles, Herman J. Mankiewicz」を外挿することができる。

**【0412】**

いくつかの実装では、自然言語構文解析などの技術を使用して、インスタンス、属性、およびそれらの値を識別することができる。例えば、電子文書の構文解析を行って、電子

50

文書内の「[Instance] has a/an [attribute] of [value]」などのフレーズを識別することができる。

【0413】

図55は、電子文書コレクション内の2つまたはそれ以上の電子文書から引き出される属性値を構造化プレゼンテーションでユーザーに提示するための例示的なプロセス5500の流れ図である。プロセス5500は、1つまたは複数の機械可読命令セットを実行することによってオペレーションを実行する1つまたは複数のコンピュータによって実行されうる。プロセス5500は、他のデータ処理活動から独立して、または連携して実行されうる。例えば、プロセス5500は、プロセス600(図6)の一部として実行されうる。

【0414】

プロセス5500を実行するシステムは、インスタンス識別子および属性識別子を受け取ることができる(ステップ5505)。プロセス5500を実行するシステムは、ユーザーから直接的に(例えば、検索クエリの形で)、または間接的に(例えば、構造化データコレクション905(図9)の一部として)インスタンス識別子および属性識別子を受け取ることができる。

【0415】

プロセス5500を実行するシステムは、属性の値を含む受け取ったインスタンスに関連する電子文書を識別することができる(ステップ5510)。例えば、システムは、データセンター208(図53)内の属性/値/インスタンスコレクション5305にアクセスして、関連する電子文書を識別することができる。他の例として、検索エンジン(例えば、検索エンジン202)は、インスタンスおよび属性識別子を使用してキーワード検索を実行することで関連する文書を識別することができる。いくつかの場合において、そのようなキーワード検索は、言語構文解析、または値の識別を容易にする他の技術で補うことができる。

【0416】

プロセス5500を実行するシステムは、構造化プレゼンテーションで提示するために識別されたインスタンスの識別された属性に対する値のサブセットを画定することができる。これらの値のサブセットは、構造化プレゼンテーションに書き込むために適切であると考えられる、または適切である可能性がある、1つまたは複数の値を含むことができる。特に、これらの値のサブセットは、識別されたインスタンスの識別された属性を、正確に、かつ構造化プレゼンテーションの閲覧者の望みに一致する形で、特徴付けるものとして考えることができる。さらに以下で説明されているように、構造化プレゼンテーションの閲覧者の望みは、例えば、閲覧者から受け取った値の選択結果に基づいて、または例えば、構造化データ905(図9)などの情報の事前に存在している構造化コレクション内の同じ、または他のインスタンスの同じ、または他の属性の特徴付けに基づいて確認することができる。

【0417】

プロセス5500を実行するシステムは、値のサブセットが書き込まれた構造化プレゼンテーションを表示するための命令を備えることができる(ステップ5520)。そこで、構造化プレゼンテーションは、電子文書のコレクションから収集された情報(つまり、電子文書コレクションから収集された値のサブセット)に基づいて提示することができる(ステップ615、図6)。

【0418】

いくつかの実装では、プロセス5500を数回、例えば、インスタンス識別子および/または属性識別子の個数の分だけ実行することができる。

【0419】

図56は、構造化プレゼンテーションで提示する1つまたは複数の値を確定するためのプロセス5600の流れ図である。プロセス5600は、単独で、または他の活動と連携して実行されうる。例えば、プロセス5600は、プロセス5500におけるステップ5515で実行されうる。

【0420】

プロセス5600を実行するシステムは、電子文書コレクション内の2つまたはそれ以上の文書からの属性の値を2つまたはそれ以上のグループに分けることができる(ステップ5605

10

20

30

40

50

)。グループ分けされた値は、電子文書から直接引き出されるか、またはテーブル5400(図54)のような属性、値、およびインスタンスの関連付けなどの電子文書のコンテンツの記述から引き出されうる。

#### 【0421】

システムは、値をいつグループ分けすべきかを決定するための1つまたは複数の異なる標準を使用して値をグループ分けすることができる。例えば、1つの標準では、グループ分けされた値が同一であることを要求することができる。例えば、2つの値「4」をグループにまとめることができる。他の標準では、数値が同一である特定の範囲内にあることを要求することができる。例えば、値「3.14」および「3.14159」をグループにまとめることができる。他の標準では、語の値が同一であるか、または同一の意味を持つことを要求することができる。例えば、語「czar」、「czar」、および「tsar」はすべて、グループにまとめることができる。他の標準では、語の値が概念のオントロジーにおいて同じ概念を表すことを要求することができる。例えば、語「pink」および「mauve」をグループにまとめることができる。他の標準では、異なるフォーマットで書き込まれる値をグループにまとめることができる。例えば、日付「July 25, 1982」および「7/25/1982」をグループにまとめることができる。他の標準では、異なる単位で書き込まれる値をグループにまとめることができる。例えば、測定単位「1m」および「100cm」をグループにまとめることができる。他の標準では、異なるが、ほぼ等しい単位で書き込まれる値をグループにまとめることができる。例えば、測定単位「1m」および「39 inches」をグループにまとめることができる。

10

20

#### 【0422】

これらのグループ分けを使用することで、プロセス5600を実行するシステムは、以下のサブプロセスのうちの1つまたは複数を任意の順序で実行し、1つのグループを選択し、したがって、属性値のコレクションから値のサブセットを選択することができる。

#### 【0423】

第1のサブプロセスでは、プロセス5600を実行するシステムは、構造化プレゼンテーションで提示するため最高の「value」を持つグループを選択することができる(ステップ5610)。いくつかの実装では、グループの「value」は、そのグループ内の値のカウントを反映する。統計用語によれば、プロセス5500を実行するシステムは、電子文書コレクション内で高い頻度を有する値を選択することができる。実際、このアプローチにより、電子文書コレクション内の文書はインスタンスの属性の値について「vote」することができる。

30

#### 【0424】

他の実装では、グループの「value」は、値のカウントを反映するだけでなく、値が引き出される文書の特性を反映するパラメータによる個別のカウントの重みまたはスコアとなる。例えば、カウントは、例えば、値が引き出される文書のページランク、ユーザーによって与えられたその文書に対する重み係数、その文書から引き出された他の値の個数、または文書の「age」に基づいて重み付けすることができる。例えば、最近作成された文書ほど、特定のインスタンスの属性を特徴付ける際の正確さが増すものと考えられる。

#### 【0425】

このサブプロセスのアプローチは、例えば、1つの電子文書上の値が誤字であるときの、不注意による属性の特徴付けの誤りをなくすうえで有効である。しかし、単独では、このアプローチは、特定の状況の下で、不適切な値を選択する可能性がある。例えば、多数の文書が体積属性をリットル単位で特徴付けているとしても、閲覧者は、構造化プレゼンテーションにおいてガロン単位で特徴付けられたその属性を有することに関心がある場合がある。

40

#### 【0426】

第2のサブプロセスでは、プロセス5600を実行するシステムは、例えば、許容可能な値の範囲、または許容可能な値の単位に対する制約条件のユーザー指定を受け取ることができる(ステップ5615)。例えば、システムは、ディスプレイ画面、例えば、ディスプレイ10

50

4(図1)に、ユーザーが値の範囲もしくは測定制約条件の単位を選択することを可能にするGUIコンポーネントを備えることができる。制約条件は、上限なし(例えば、「値>1」)または上下限が決まっている(例えば、1から10までの範囲の値)とすることができる。

【0427】

この制約条件を受け取った後、プロセス5600を実行するシステムは、構造化プレゼンテーションで提示するため受け取った制約条件と合致するグループを選択することができる(ステップ5620)。例えば、ユーザーが「meters」を適切な測定単位として選択した場合、システムの実行により、メートル単位で表される値の1つまたは複数のグループを選択することができる。

【0428】

このサブプロセスのアプローチは、構造化プレゼンテーションで提示される値が編成された、システム配置で提示されることを保証するうえで有効である。例えば、Michael Jordanの身長を例えば特徴付けるために使用される値の測定単位は、Magic Johnsonの身長を特徴付けるために使用される値の測定単位と同一であるように制約することができる。このような編成された、システム配置により、ユーザーは、値を提示する単位を気にせずに、異なるインスタンスの同じ属性の値を容易に比較することができる。

【0429】

第3のサブプロセスでは、プロセス5600を実行するシステムは、それぞれのグループ内の属性値を引き出した文書の「quality」を決定することができる(ステップ5625)。文書の「quality」は、文書内の情報が正確であり、属性の値を誤って特徴付けることをしない確率を反映するものとしてよい。例えば、商品の商業的サプライヤーは、一般的に、それらの商品に関する正確な情報を提供する。したがって、商業的サプライヤーによって提供される情報の「quality」は、個人によって提供される情報の「quality」より高いとみなすことができる。他の例として、情報が引き出される文書の品質を決定する際にバイアスを考慮することができる。例えば、独立情報源であると言われているソース(例えば、連邦議会予算事務局など)から引き出された情報は、政党から引き出された情報に比べて品質が高いと考えてよい。

【0430】

他の例として、文書の品質は、ユーザーによる文書のソース、または文書それ自体の品質の指定に基づくものとしてすることができる。例えば、ユーザーは、コンシューマーユニオン(コンシューマーレポートの発行元)から引き出された自動車信頼性情報は高品質であるが、ロード&トラック誌から引き出された自動車信頼性情報はそうでないと指示することができる。プロセス5600を実行するシステムは、最高品質の文書から引き出された値を含むグループを選択することもできる(ステップ5630)。

【0431】

このサブプロセスのアプローチは、属性の値が議論のテーマとなっている状況では特に有効である。例えば、大学のポイントガードの真の身長、フィラデルフィア最高のチーズステーキ、および特定のホテルに与えられた星の数など、さまざまな値に関する意見の相違がある。閲覧者がそのような属性を特徴付ける値の好ましい「high quality」提供者を指定できるようにすることで、構造化プレゼンテーションを閲覧者の期待にそって修正することができる。

【0432】

上述のように、これらの活動および他の活動の組み合わせを実行して、構造化プレゼンテーションで提示するための1つまたは複数の値を選択することができる。例えば、値の総数のうちのあるパーセンテージ(例えば、>10%)を代表する値のすべてのグループを、第1のスクリーニングで選択し(ステップ5610)、その後、単位指定に基づいてこれらのグループをさらに選別することができる(ステップ5620)。残りのグループは、それぞれのグループ内の属性値を引き出した文書の品質に基づいてさらに選別することができる(ステップ5630)。したがって、いくつかの実装では、プロセス5600は、首尾一貫した測定単位とともに、特徴付けの誤りのない、また閲覧者が好むソースから引き出された、残りのグル

10

20

30

40

50

ープからの1つまたは複数の値を提供することができる。

【0433】

図57は、構造化プレゼンテーションで提示する1つまたは複数の値を選択するためのプロセス5700の流れ図である。プロセス5700は、単独で、または他の活動と連携して実行されうる。例えば、プロセス5700は、プロセス5500(図55)におけるステップ5515で、単独で、またはプロセス5600(図56)のサブプロセスのうちの1つまたは複数と連携して実行されうる。

【0434】

プロセス5700を実行するシステムは、電子文書コレクション内の2つまたはそれ以上の文書から抽出された属性の値を2つまたはそれ以上のグループに分けることができる(ステップ5605)。プロセス5700を実行するシステムは、値のグループの記述をユーザーに提示することができる(ステップ5705)。例えば、システムは、それぞれのグループ内の最も一般的な値、またはそれぞれのグループ内の値のうちのいくつかの値のリストを、ディスプレイに、例えば、ディスプレイ画面104(図1)に、ユーザーに向けて表示することができる。いくつかの実装では、値のグループの記述は、それらのグループを特徴付ける追加の情報を含んでいてもよい。例えば、それぞれのグループ内の値の個数の計数を表示すること、それぞれのグループ内に見つかる抽出された値のパーセントを反映するパーセンテージを表示すること、および/またはグループ内の測定単位の記述を表示することが可能である。他の例として、それぞれのグループ内の値の抽出元であった電子文書の品質の推定を表示することができる。さらに他の例として、それぞれのグループ内の値の抽出元であった文書の識別、ロケーション、および/またはスニペットもしくは他の抜粋を表示することができる。

【0435】

いくつかの実装では、値のグループの記述は、確かさベースの順序付けでソートされる。つまり、値のグループの記述は、プロセス5700を実行するシステムに対するそれぞれのグループ内の値の正確さに関する確かさに応じて順序付けられる。それぞれのグループ内の値の正確さの確かさは、例えば、それぞれのグループ内の値の個数、それらの値の抽出元となった文書の品質、および同様のものに基づいて決定できる。

【0436】

プロセス5700を実行するシステムは、値の所望のグループのユーザー選択を受け取ることができる(ステップ5710)。例えば、システムは、所望の値のグループの選択を識別するユーザーインタラクションを受け取ることができる。

【0437】

いくつかの実装では、プロセス5700を実行するシステムは、ユーザー選択に基づいて構造化プレゼンテーションの他の態様を変更することもできる(ステップ5715)。例えば、ユーザーがメートルの測定単位で値のグループを選択し、他のインスタンスの同じ属性を特徴付けるが、フィートの測定単位で提示される他の値がある場合、そのような値は、構造化プレゼンテーション106で変換され、メートル単位で掲示されうる。

【0438】

図58は、構造化プレゼンテーションで提示する1つまたは複数の値を選択するための例示的なプロセス5800の流れ図である。プロセス5800は、単独で、または他の活動と連携して実行されうる。例えば、プロセス5800は、プロセス5500(図55)におけるステップ5515で、単独で、またはプロセス5600(図56)のサブプロセスのうちの1つまたは複数および/またはプロセス5700(図57)と連携して実行されうる。

【0439】

プロセス5800を実行するシステムは、構造化データコレクション、例えば、構造化データコレクション905においてインスタンスおよび他の属性に関連する電子文書コレクション内の電子文書を識別することができる(ステップ5805)。上述のように、構造化データコレクション905は、構造化プレゼンテーション106のバージョンとすることができる。

【0440】

10

20

30

40

50

構造化データコレクション内のインスタンスおよび他の属性に関連する文書は、さまざまな方法で識別されうる。例えば、プロセス5800を実行するシステムは、インスタンス、それらの属性、およびこれらの属性を特徴付ける値を関連付けるデータコレクション、例えば、属性/値/インスタンスコレクション5305(図53)にアクセスすることができる。構造化データコレクション内のインスタンスおよび他の属性に関連する情報を含む文書は、例えば、両方のデータコレクション内のインスタンスの識別子および属性を比較することによって、その中で識別されうる。他の例として、プロセス5800を実行するシステムは、インスタンスの識別子および属性を1つまたは複数の検索クエリにおける検索語として使用することができる。このような検索クエリは、単独で、または言語構文解析および文字列比較などの他の抽出技術と連携して、電子文書コレクション内の関連する文書を識別するために使用することができる。

10

**【0441】**

プロセス5800を実行するシステムは、識別された文書から構造化プレゼンテーションで提示する1つまたは複数の値を選択することもできる(ステップ5810)。図59は、電子文書コレクション102内の電子文書から引き出される属性値が構造化プレゼンテーション106でユーザーに提示される状況を示す略図である。特に、システム900(図9)などのシステムは、電子文書コレクション102内の電子文書から引き出された属性、値、およびインスタンスを関連付けるテーブル5400から属性値を引き出す。このシステムは、これらの属性値を初期構造化プレゼンテーション106とマージして、最終構造化プレゼンテーション106を形成することも行う。そのため、初期構造化プレゼンテーション106は、構造化データ905(

20

**【0442】**

示されている例では、初期構造化プレゼンテーションは、新規属性(つまり、属性「AIRPORT」)の値を構造化プレゼンテーションで識別されたインスタンスに関連付けるように修正されている。特に、新規列5905が初期構造化プレゼンテーションに追加されている。列5905の先頭に、語「AIRPORT」を使用して新規属性を識別する属性識別子5910が付く。構造化プレゼンテーションへの新規属性の値の追加は、例えば、さらに図9~20および関連する本文で説明されているように、ユーザーによるインタラクティブな操作に基づいて、または自動的にトリガーされうる。

**【0443】**

検索エンジン202などのシステムは、電子文書コレクション102(テーブル5400など)内の電子文書から引き出された属性、値、およびインスタンスを関連付けるデータコレクションにアクセスすることができる。このようなデータコレクションを使用することで、システムは、初期構造化プレゼンテーションにおけるインスタンスの1つまたは複数の新規属性を特徴付ける1つまたは複数の値を選択することができる。例えば、例示されている状況では、値5915(つまり、値「value\_ai」)は、文書「DOC\_3」内にインスタンス「NEW YORK」の属性「AIRPORT」を特徴付ける。必要ならば、システムは、例えばプロセス5600、5700、5800(図56、57、58)のうちの1つまたは複数を使用して、表示する新規属性の1つまたは複数の値を選択することができる。

30

**【0444】**

初期構造化プレゼンテーションにおけるインスタンスのうちの1つまたは複数のインスタンスの新規属性を特徴付ける値が選択された後、最終構造化プレゼンテーション106を閲覧者に提示することができる。最終構造化プレゼンテーション106は、構造化プレゼンテーションにおけるインスタンスのうちの1つまたは複数のインスタンスの新規属性を特徴付ける選択された値を含むことができる。例えば、図示されているように、値5915を閲覧者が見られるように最終構造化プレゼンテーション106で提示することができる。

40

**【0445】**

図60は、属性と属性値の両方が電子文書コレクション内の電子文書から引き出され、構造化プレゼンテーションでユーザーに提示されるプロセス6000を示す略図である。

**【0446】**

50

プロセス6000では、初期構造化データコレクション905は、事前に存在している構造化プレゼンテーション6005を含むことができる。事前に存在している構造化プレゼンテーション6005は、1つまたは複数の属性値、例えば列6010内の属性値を使用してインスタンスを特徴付けることができる。事前に存在している構造化プレゼンテーション6005におけるインスタンスを特徴付ける新規属性は、図37～51および関連する本文において説明されているように、電子文書コレクション102内の電子文書のコンテンツに基づいて構成することができる。新規属性は、ステップ6015で事前に存在している構造化プレゼンテーション6005に追加することができ、構造化プレゼンテーション6020の一部として表示することができる。このような属性の新規値は、本明細書で説明されているように、電子文書コレクション102内の電子文書のコンテンツに基づいて構成することができる。新規値は、ステップ6025で事前に存在している構造化プレゼンテーション6005に追加することができ、構造化プレゼンテーション6020の一部として表示することができる。特に、示されている例では、新規列6030は、新規属性を識別する新規属性識別子308(つまり、属性識別子6035)および新規属性を特徴付ける属性値307(つまり、属性値6040、6045)の新規コレクションを含みうる。実際、事前に存在している構造化プレゼンテーション6005のコンテンツは、電子文書コレクション102から引き出されたコンテンツとマージされている。

10

## 【0447】

図61は、電子文書コレクション内の文書のコンテンツに基づいて値を構造化プレゼンテーションに追加するためのプロセス6100の流れ図である。プロセス6100は、1つまたは複数の機械可読命令セットを実行することによってオペレーションを実行する1つまたは複数のコンピュータのシステムによって実行されうる。例えば、プロセス6100は、システム200内の検索エンジン202によって実行されうる。プロセス6100は、単独で、または他の活動と連携して実行されうる。例えば、プロセス6100は、プロセス700、600、800(図6、7、8)のうちの1つまたは複数のプロセスの一部として実行されうる。

20

## 【0448】

いくつかの実装では、プロセス6100は、例えば、ユーザーまたは新規構造化プレゼンテーションの作成もしくは構造化プレゼンテーションの更新をトリガーする他のシステムもしくはプロセスから入力を受け取ったことに応答して実行されうる。例えば、プロセス6100は、1つまたは複数の新規属性を構造化プレゼンテーション106に追加するユーザー要求への応答として実行されうる。他の例として、プロセス6100は、検索クエリを受け取ったことに対する応答として、検索エンジン、例えば、検索エンジン202(図2)によって実行されうる。

30

## 【0449】

プロセス6100を実行するシステムは、構造化プレゼンテーションにおけるインスタンスおよび属性の指定を受け取ることができる(ステップ6105)。構造化プレゼンテーションは、新規もしくは事前に存在している構造化プレゼンテーションであるものとしてよい。例えば、システムは、構造化プレゼンテーションにおいて特徴付けられるインスタンス、もしくはインスタンスのカテゴリを指定する検索クエリを受け取ることができる。

## 【0450】

他の例として、ユーザーが事前に存在している構造化プレゼンテーションをインタラクティブに操作して、インスタンス、および属性、またはその両方を指定することができる。事前に存在している構造化プレゼンテーションに対するユーザーインタラクションにより、インスタンスおよび/または属性を内在的に、または手動で指定することができる。内在的指定では、構造化表示におけるインスタンスおよび属性識別子のシステム配置を利用して、セルに対するユーザーインタラクションによりそのセルに関連付けられているインスタンスおよび属性を指定する。

40

## 【0451】

それとは対照的に、手動により指定では、ユーザーは、どのセルがセルに関連付けられているインスタンスおよび属性の識別子を含むかを手動で識別する。例えば、ユーザーは、構造化プレゼンテーション内のインスタンス識別子、属性識別子、またはその両方の配

50

列を指定する検索クエリをセルに入力することができる。例えば、式「(CELL\_1, CELL\_2)」を含む検索クエリは、このセルがセル「CELL\_1」で識別されるインスタンスのセル「CELL\_2」で指定される属性に関連付けられること、およびこのインスタンスのこの属性に対する検索を実行することを指定することができる。インスタンスおよび属性識別子のこのような手動による指定は、インスタンスおよび属性識別子の位置がユーザーには明らかであるが、構造化プレゼンテーションを提示するデータ処理デバイスには未知のものである場合に、スプレッドシートテーブルなどの構造化プレゼンテーションにおいて特に有用である。1つまたは複数のインスタンスおよび属性を指定するための事前に存在している構造化プレゼンテーションに対するユーザーインタラクションの他の例について、以下で詳述する。

10

**【0452】**

プロセス6100を実行するシステムは、インスタンスの指定された属性に対し電子文書コレクション内の文書から1つまたは複数の値提案を構成することができる(ステップ6110)。値提案は、多数の異なる方法で指定された属性について構成することができる。例えば、一実装では、システムは、指定されたインスタンスおよび属性を使用して構築される検索クエリを使用して検索を実行することによって電子文書コレクション102内の文書から値提案を構成することができる。例えば、値提案は、例えば、図52~60および関連する本文で説明されているように、指定されたインスタンスおよび属性に関係する構造化コンポーネントを含む文書を見つけることによって構成される。

**【0453】**

20

他の例としては、検索クエリにおいて、インスタンスの属性を特徴付ける値が出現する可能性が高いことを示す言語パターン中に指定されたインスタンスおよび属性の識別子が見つかることが要求される場合がある。このようなパターンの例として、「the <attribute> of <entity> is」、「<entity> with an <attribute> of」、「<entity> has an <attribute> of」、「<entity>'s <attribute> is」、および同様のものが挙げられる。このようなパターンを使用して、電子文書内のテキストコンテンツから値提案を抽出することができる。

**【0454】**

プロセス6100を実行するシステムは、1つまたは複数の値提案をユーザーに提供することができる(ステップ6115)。例えば、値提案のリストを、事前に存在している構造化プレゼンテーションを表示する同じディスプレイ画面上にユーザー向けに表示することができる。値提案のリストの表示は、事前に存在している構造化プレゼンテーションに追加するために値を選択する前に、実行することができる。

30

**【0455】**

他の例として、いくつかの実装では、構造化プレゼンテーションにおいて、検索情報およびインタラクティブ要素とともに、値提案を隠すこともできる。そのような実装の例について以下でさらに説明する。

**【0456】**

プロセス6100を実行するシステムは、構造化表示で提示される値提案のユーザー選択を受け取ることができる(ステップ6120)。例えば、インタラクティブ要素がユーザーとやり取りして、値提案を選択する1つまたは複数のユーザー入力(例えば、マウスのクリック、キーの押下、または他のユーザー入力)を受け取ることができる。いくつかの実装では、インタラクティブ要素は、以下でさらに説明するように、構造化プレゼンテーションにおいて隠すことができる。プロセス6100を実行するシステムは、選択された値を構造化プレゼンテーションに追加し(ステップ6125)、構造化プレゼンテーションにおいて選択された値を表示することもできる。

40

**【0457】**

図62は、検索インターフェイスが隠されている構造化プレゼンテーション、つまり、構造化プレゼンテーション6200の略図である。検索インターフェイスは、検索情報、1つまたは複数の検索インタラクティブ要素、またはそれらの組み合わせを含むことができる。

50

インタラクティブ要素は、ユーザーとやり取りすることができる、例えば、入力された命令を受け取ることができるグラフィカルユーザーインターフェイスのコンポーネントである。検索インタラクティブ要素および検索情報は、検索に関連するものである。検索は、電子文書コレクション内の情報を見つけるプロセスである。検索インターフェイスは、例えば、構造化プレゼンテーションに値を書き込むために検索が利用可能かどうかを示す情報、そのような検索を実行すべきであることをユーザーが示すことを可能にするインタラクティブ要素、検索時に特定された電子文書を識別する表示、構造化プレゼンテーションに値を書き込むために複数の電子文書のうちから電子文書をユーザーが選択することを可能にするインタラクティブ要素、またはこれらの機能および他の機能の組み合わせを含むことができる。

10

**【0458】**

構造化プレゼンテーション6200は、上述の構造化プレゼンテーションのうちのどれかを含む、任意の形態の構造化プレゼンテーションであってよい。例えば、構造化プレゼンテーション6200は、図示されているように、スプレッドシートフレームワーク内に表示されるデータテーブルであってもよい。構造化プレゼンテーション6200のデータテーブルは、行302および列304のコレクションを含む。それぞれの行302は、各インスタンス識別子306を含み、それぞれの列304は、各属性識別子308を含む。行302および列304内のインスタンス識別子306および属性識別子308の配列および位置決めは、構造化プレゼンテーション6200が表示されるスプレッドシートフレームワークのそれぞれのセルをインスタンスおよび属性に関連付ける。例えば、構造化プレゼンテーション6200内のセル6205は、「Tesla Roadster」として識別されるインスタンスおよび「mpg」として識別される属性に関連付けられる。構造化プレゼンテーション6200内のセル6210は、「Chevy Volt」として識別されるインスタンスおよび「range」として識別される属性に関連付けられる。構造化プレゼンテーション6200内のセル6215は、「Myers NmG」として識別されるインスタンスおよび「top speed」として識別される属性に関連付けられる。構造化プレゼンテーション6200内のセル6220は、「Myers NmG」として識別されるインスタンスおよび「mpg」として識別される属性に関連付けられる。

20

**【0459】**

インスタンス、属性、およびセル6205、6210、6215、6220などのセルの間の関連付けを使用して、ユーザーからインスタンスおよび属性の指定を受け取ることができる。例えば、セル6220を選択するユーザーインタラクションを受け取るとは、「Myers NmG」として識別されるインスタンスおよび「mpg」として識別される属性を指定する入力とみなすことができる。セルを選択するユーザーインタラクションは、例えば、セル上のカーソル6225の位置を決める入力を受け取ること、ユーザーがセルをクリックすること、または同様のものを含むものとしてよい。いくつかの実装では、セルの選択は、セルの中、またはセルの周りで周囲ハイライト6230などの視覚的な印を位置決めすることによって表すことができる。

30

**【0460】**

例示されている実装では、選択時には、選択されたセル6220に値307が入っていない。これに対する理由はいくつかある。例えば、構造化プレゼンテーション6200は、値をまだ書き込まれていない新規構造化プレゼンテーションとすることができる。他の例として、構造化プレゼンテーション6200は、値が削除された事前に存在している構造化プレゼンテーションであってもよい。さらに他の例として、構造化プレゼンテーション6200は、何らかの理由により、もはや値の供給源として機能しえないソース文書から以前の値を引き出した事前に存在している構造化プレゼンテーションであってもよい。

40

**【0461】**

図63は、検索インターフェイスが隠されている他の構造化プレゼンテーション6300の略図である。構造化プレゼンテーション6200との関連で、構造化プレゼンテーション6300は、選択されたセル6220内の値307を含む。これに対する理由はいくつかある。例えば、セル6220は、自動的に、例えば、検索クエリを受け取ったことに対する応答として、値307

50

を書き込まれていることがある。他の例として、セル6220への書き込みは、ユーザーが手動でセル6220をインタラクティブに操作して値を入力することによって行われている場合がある。さらに他の例として、セル6220は、ユーザーが（内在的にまたは手動で）セル6220に関連付けられているインスタンス、属性、またはその両方を指定したことに応答して値307が書き込まれている可能性がある。いずれにせよ、セル6220の選択では、「Myers N mG」として識別されるインスタンスおよびそれに関連付けられている「mpg」として識別される属性を指定する。

#### 【0462】

図64は、検索インターフェイスが隠されている他の構造化プレゼンテーション6400の略図である。構造化プレゼンテーション6400は、視覚的な印6405を含む。視覚的な印6405は、隠されている検索情報またはインタラクティブ要素が構造化プレゼンテーション6400からアクセス可能であることを視覚的に示す。

10

#### 【0463】

例示されている実装では、それぞれの視覚的な印6405は、セル6205、6210、6215、6220などの個別のセルに見られる。セル内の視覚的な印6405の位置決めおよび配列（それに付随して、行302および列304内のインスタンス識別子306および属性識別子308に相対的な視覚的な印6405の位置決めおよび配列）は、以下でさらに説明するように、隠されている検索情報およびインタラクティブ要素の関連性および機能を視覚的に示すことができる。

#### 【0464】

動作中、構造化プレゼンテーション6200、6300、6400に対するユーザーインタラクションは、隠されている検索インターフェイスを提示するトリガーとなりうる。上述のように、検索インターフェイスは、検索情報、検索インタラクティブ要素、またはその両方を含むことができる。検索インターフェイスは、検索情報およびインタラクティブ要素が構造化プレゼンテーション内で常に認識できる必要はないという点で構造化プレゼンテーション内に隠することができる。むしろ、隠されている検索インターフェイスは、構造化プレゼンテーションがいくつかの状態にある間、視界から完全にまたは部分的に隠することができる。例えば、閲覧者が構造化プレゼンテーションの他の情報コンテンツをレビューしている可能性が高い状態において、隠し検索インターフェイスを隠することができる。このように隠すことで、他の情報の提示に利用可能な構造化プレゼンテーションの部分を拡大し、構造化プレゼンテーションの可読性を改善するために見えている散乱しているものを減らすことができる。

20

30

#### 【0465】

図65は、以前に隠されていた検索インターフェイスが提示される表示要素6500の図である。いくつかの実装では、構造化プレゼンテーションそれ自体に対するユーザーインタラクションに応答して表示要素6500を提示することができる。表示要素6500は、セル6220に対するユーザーインタラクションに応答して、構造化プレゼンテーション(構造化プレゼンテーション6200、6300、6400など)の前に「ポップアップ」し、ウィンドウ6510内に検索インタラクティブ要素6505を提示することができる。例えば、検索インタラクティブ要素6505およびウィンドウ6510は、ユーザーがマウスを使用してセル6220をクリックしたことに応答して提示されうる。検索インタラクティブ要素6505は、セル6220に書き込むために「more options...」が利用可能であることを示すテキストを含むハイパーリンクである。

40

#### 【0466】

図66は、以前に隠されていた検索インターフェイスが提示される表示要素6600の図である。いくつかの実装では、構造化プレゼンテーションそれ自体に対するユーザーインタラクションに応答して表示要素6600を提示することができる。検索インタラクティブ要素6505に加えて、表示要素6600は、セル6220に対するユーザーインタラクションに応答してウィンドウ6510内にソース識別子6605を提示する。ソース識別子6605は、セル6220に書き込む値307のソースである電子文書を識別するテキストまたは他の情報を含む。ソース識別子6605によって識別されたソース文書は、事前検索の結果として特定された文書であって

50

もよい。いくつかの実装では、ソース識別子6605は、ソース文書へのハイパーリンクを含むこともできる。

【0467】

図67は、以前に隠されていた検索インターフェイスが提示される表示要素6700の図である。いくつかの実装では、構造化プレゼンテーションそれ自体に対するユーザーインタラクションにตอบสนองして表示要素6700を提示することができる。検索インタラクティブ要素6505およびソース識別子6605に加えて、表示要素6700は、セル6220に対するユーザーインタラクションにตอบสนองしてウィンドウ6510内にスニペット6705を提示する。スニペット6705は、セル6220に書き込む値307のソースである電子文書内の値の背景状況を記述するテキストまたは他の情報である。

10

【0468】

図68は、以前に隠されていた検索インターフェイスが提示される表示要素6800の図である。いくつかの実装では、構造化プレゼンテーションそれ自体に対するユーザーインタラクションにตอบสนองして、または以前に隠されていた検索インタラクティブ要素6505に対するユーザーインタラクションにตอบสนองして表示要素6800を提示することができる。表示要素6800は、ヘッダ6802、関連するソース選択領域6805、および一致するソース選択領域6810を含む。ヘッダ6802は、値の追加先となるセルを識別するテキストまたは他の情報を含むことができる。例示されている実装では、セル6220は、セル6220内の値307によって特徴付けられている属性と値(つまり、Myers NmG: mpg)によって識別される。

【0469】

20

関連するソース選択領域6805は、指定されたインスタンスおよび属性に対するソース電子文書の関連性を構造化プレゼンテーションに書き込む値を選択する際に使用するとユーザーが指定することを可能にする情報およびインタラクティブ要素を備えることができる。例示されている実装では、ユーザーは、単一の「最も関連性の高い」文書が構造化プレゼンテーションに書き込む値の単独のソースであると指定することができる。文書の関連性により、文書と例えば、検索が定義する属性およびインスタンスとマッチングする程度を特徴付けることができる。

【0470】

例示されている実装では、関連するソース選択領域6805は、ヘッダ6815、選択ウィジェット6820、値識別子6825、およびソース識別子6830を含む。ヘッダ6815は、関連するソース選択領域6805を使用することで最も関連性の高い電子文書をヘッダ6802において識別されているセルに書き込む値のソースとして使用することをユーザーが指定することが可能であることを識別するテキストまたは他の情報を含む。選択ウィジェット6820を使用することで、ユーザーは、最も関連性の高い文書を構造化プレゼンテーションに書き込む値のソースとして使用することを選択することができる。値識別子6825は、現在最も関連性の高い文書から引き出された値を識別するテキストまたは他の情報を含む。ソース識別子6830は、現在最も関連性の高い電子文書を識別するテキストまたは他の情報を含む。いくつかの実装では、ソース識別子6830は、現在最も関連性の高い文書へのハイパーリンクを含むこともできる。最も関連性の高い文書は、時間が経過するにつれ変わることがあるため、値識別子6825によって識別された値およびソース識別子6830によって識別された文書も、時間が経過するにつれ変わらう。

30

40

【0471】

一致するソース選択領域6810は、ソース電子文書を構造化プレゼンテーションに書き込む値を選択する際に矛盾なく使用するとユーザーが指定することを可能にする情報およびインタラクティブ要素を備えることができる。例示されている実装では、ユーザーは、構造化プレゼンテーションに書き込む値のソースとして矛盾なく使用されるその文書を指定するために3つの候補文書のうちから文書を選択することができる。

【0472】

例示されている実装では、首尾一貫したソース選択領域6810は、ヘッダ6835、選択ウィジェット6840のコレクション、値識別子6845のコレクション、およびソース識別子6850の

50

コレクションを含む。

【0473】

ヘッダ6835は、関連するソース選択領域6805を使用することでソース電子文書が値を選択する際に矛盾なく使用されることをユーザーが指定することが可能であることを識別するテキストまたは他の情報を含む。選択ウィジェット6840では、ユーザーは、矛盾なく使用される文書を選択することができる。例示されている実装では、ユーザーは、3つの異なる文書のうちから文書を選択することができる。値識別子6845は、構造化プレゼンテーションに書き込むために特定の文書から引き出されうる現在値を識別するテキストまたは他の情報を含む。ソース識別子6850は、値識別子6825によって識別される値が引き出される電子文書を識別するテキストまたは他の情報を含む。いくつかの実装では、ソース識別子6850は、値識別子6825によって識別される値が引き出される電子文書へのハイパーリンクを含むこともできる。

10

【0474】

電子文書の関連性と電子文書内の値は、両方とも、時間の経過とともに変化しうる。例えば、電子文書を電子文書コレクションに追加する人は、インスタンスおよび属性に対するその文書の関連性が変化するように電子文書のコンテンツを変えることができる。他の例として、電子文書を電子文書コレクションに追加する人は、インスタンスの属性を特徴付けるために使用される値を変えることができる。ヘッダ6815、6835は、発生する可能性のある変化の性質を識別するテキストまたは他の情報を含むことができる。例えば、例示されている実装では、ヘッダ6815は、指定されたインスタンスおよび属性に対するソース電子文書の関連性が構造化プレゼンテーションに書き込む値を選択する際に使用されるとユーザーが指定したときに最も関連性の高い文書と属性の値の両方が変化しうることを識別するテキストを含む。他の例として、例示されている実装では、ヘッダ6835は、ソース電子文書が構造化プレゼンテーションに書き込む値を選択する際に矛盾なく使用されるとユーザーが指定したときに属性の値が変化しうることを述べたテキストを含む。

20

【0475】

図69は、以前に隠されていた検索インターフェイスが提示される表示要素6900の図である。いくつかの実装では、構造化プレゼンテーションそれ自体に対するユーザーインタラクションにตอบสนองして、または以前に隠されていた検索インタラクティブ要素6505に対するユーザーインタラクションにตอบสนองして、表示要素6900を提示することができる。ヘッダ6802、6815、6835、選択ウィジェット6820、6840、値識別子6825、6845、およびソース識別子6830、6850に加えて、表示要素6900は、スニペット6905のコレクションおよび検索インタラクティブ要素6910のコレクションを含む。それぞれのスニペット6905は、識別された値のソースである電子文書内の値識別子6825、6845によって識別される各値の背景状況を記述するテキストまたは他の情報である。検索インタラクティブ要素6910は、各値識別子6845によって識別される値のソースである各電子文書へユーザーが移動することを可能にするハイパーリンクである。

30

【0476】

図70は、以前に隠されていた検索インターフェイスが提示される表示要素7000の図である。いくつかの実装では、構造化プレゼンテーションそれ自体に対するユーザーインタラクションにตอบสนองして、または以前に隠されていた検索インタラクティブ要素6505に対するユーザーインタラクションにตอบสนองして、表示要素7000を提示することができる。ヘッダ6802、6815、6835、選択ウィジェット6820、6840、値識別子6825、6845、およびソース識別子6830、6850、スニペット6905、および検索インタラクティブ要素6910に加えて、表示要素6900は検索トリガー7005を含む。検索トリガー7005は、電子文書コレクションの検索をトリガーするインタラクティブ要素である。前の検索の結果の記述と一緒に表示する場合(値識別子6825、6845、ソース識別子6830、6850、スニペット6905、および検索インタラクティブ要素6910など)、検索トリガー7005を使用することで、ユーザーは前の検索の結果に満足していないことを示すことができる。いくつかの実装では、検索トリガー7005によってトリガーされる検索は、Google(登録商標)検索エンジンなどの汎用検索エンジンを

40

50

使用して実行される「全文検索」とすることができる。いくつかの実装では、検索エンジンは、前のユーザーインタラクションによって指定されたインスタンスおよび属性を使用して自動生成されるクエリとともに提示されうる。

【0477】

図65～70に図示されているように、以前に隠されていた検索情報およびインタラクティブ要素の表示をトリガーするユーザーインタラクションの性質により、表示される検索情報およびインタラクティブ要素のカテゴリが決定されうる。例えば、構造化プレゼンテーション内で単一のセルを指定するユーザーインタラクションにより、同じセルに値を書き込むことに関連する検索情報およびインタラクティブ要素の提示がトリガーされうる。他の実装では、列、行、またはセルの他のコレクションに対するユーザーインタラクションにより、セルのコレクションに値を書き込むことに関連する検索情報およびインタラクティブ要素の提示がトリガーされうる。例えば、列に対するユーザーインタラクションを使用することで、ユーザーは、その列に書き込む値が単一のソース文書またはソース文書群から矛盾なく引き出されることを指定することができる。他の例として、行に対するユーザーインタラクションを使用することで、ユーザーは、その行に書き込む値がインスタンスおよびその行の属性に最も関連するソース文書から引き出されることを指定することができる。

10

【0478】

図71は、電子文書コレクション内の文書のコンテンツから値を引き出すことによって値を構造化プレゼンテーションに追加するためのプロセス7100の流れ図である。プロセス7100は、1つまたは複数の機械可読命令セットを実行することによってオペレーションを実行する1つまたは複数のコンピュータのシステムによって実行されうる。例えば、プロセス7100は、システム200内の検索エンジン202によって実行されうる。プロセス7100は、単独で、または他の活動と連携して実行されうる。例えば、プロセス7100は、プロセス6100(図61)のステップ6105、6115、および6120の一部として実行されうる。

20

【0479】

プロセス7100を実行するシステムは、構造化プレゼンテーションの1つまたは複数のセルを指定するユーザーインタラクションを特徴付けるデータを受け取ることができる(ステップ7105)。構造化プレゼンテーションは、新規または事前に存在している構造化プレゼンテーションであってよい。それに付随して、上で説明されているように、1つまたは複数のセルに対するインタラクティブな操作により、1つまたは複数の属性およびインスタンスを指定することができる。

30

【0480】

プロセス7100を実行するシステムは、1つまたは複数のセルに書き込む1つまたは複数の値が電子文書コレクションの事前検索の結果得られたものかどうかを判定することができる(ステップ7110)。このような判定は、構造化プレゼンテーションに目に見える形で表示される情報を特徴付ける情報だけでなく、構造化プレゼンテーションに書き込むために実行された事前検索を特徴付ける情報をも格納するデータ記憶デバイスにアクセスすることによって行うことができる。事前検索を特徴付ける格納されている情報は、例えば、検索が実際に実行されたことを示す情報、前の検索の結果セット内のソース文書のURL、およびソース文書内の値の背景状況の特徴付けるスニペットを含むことができる。

40

【0481】

プロセス7100を実行するシステムが、値が前の検索の結果として得られたと判定した場合、システムは、事前検索を特徴付ける検索情報を提示することができる(ステップ7115)。そのような情報は、例えば、値が引き出された結果セット内のソース文書を識別する情報、ソース文書内の値の背景状況の特徴付けるスニペット、およびソース文書へのハイパーリンクを含むことができる。例えば、システムは、表示要素6600、6700(図66、67)などのプレゼンテーションにおける単一のソース文書を特徴付ける検索情報を提示することができる。他の例として、システムは、表示要素6800、6900、7000(図68、69、70)などのプレゼンテーションにおける複数のソース文書(構造化プレゼンテーションに目に見える形

50

で書き込む値とは異なる値を有するソース文書を含む)に関する検索情報を提示することができる。

【0482】

いくつかの実装では、プロセス7100を実行するシステムは、ユーザーとやり取りしたその応答として単一のソース文書に関する検索情報の提示と複数のソース文書に関する検索情報との間で遷移することができる。例えば、システムは、検索インタラクティブ要素6505などの検索インタラクティブ要素に対するユーザーインタラクションを受け取り、表示要素6600、6700と表示要素6800、6900、7000(図65~70)との間で遷移することができる。

【0483】

いくつかの実装では、プロセス7100を実行するシステムは、新規検索を実行し、その新規検索によって得られた結果セット内の1つまたは複数の電子文書の特徴付ける情報を供給することができる(ステップ7120)。特徴付ける情報としては、例えば、電子文書の名前およびURL、電子文書のスニペット、電子文書の要約、または同様のものが挙げられる。結果セットは、表示要素6600、6700(図66、67)などのプレゼンテーションにおける単一のソース文書または表示要素6800、6900、7000(図68、69、70)などのプレゼンテーションにおける複数のソース文書の特徴付けることができる。いくつかの実装では、システムは、ユーザーとやり取りしたその応答として単一のソース文書に関する検索情報の提示と複数のソース文書に関する検索情報との間で遷移することができる。例えば、システムは、検索インタラクティブ要素6505などの検索インタラクティブ要素に対するユーザーインタラクションを受け取り、表示要素6600、6700と表示要素6800、6900、7000(図65~70)との間で遷移することができる。

【0484】

図72は、以前に隠されていた検索インターフェイスが提示される表示要素7200の図である。いくつかの実装では、構造化プレゼンテーションそれ自体に対するユーザーインタラクションに回答して、または以前に隠されていた検索インタラクティブ要素7005に対するユーザーインタラクションに回答して表示要素7200を提示することができる。表示要素7200は、検索、例えば、汎用検索エンジンを使用して実行される検索の結果得られる値を受け取ることができる。表示要素7200が受け取る値を構造化プレゼンテーションに追加して、インスタンスの属性を特徴付けることができる。表示要素7200は、ヘッダ7205、インスタンス識別子7210、属性識別子7215、値入力要素7220、値使用トリガー要素7225、およびプレゼンテーション閉じる要素7230を備える。

【0485】

ヘッダ7205は、表示要素7200がインスタンスの属性の値を受け取ることができると記述するテキストまたは他の情報である。ヘッダ7205は、検索の結果得られる値を入力することをユーザーに促すものであってもよい。例えば、ヘッダ7205は、検索が正常に実行されたかどうかを尋ねるテキストであってもよい。

【0486】

インスタンス識別子7210は、表示要素7200を使用して入力された値によって特徴付けられる、インスタンス、もしくはインスタンスのカテゴリを識別するテキストまたは他の情報である。例示されている実装では、インスタンス識別子7210は、インスタンス「China」を識別するテキストである。属性識別子7215は、インスタンス識別子7210によって識別されるインスタンスの属性を識別するテキストまたは他の情報である。属性識別子7215によって識別される属性は、表示要素7200が受け取る値によって特徴付けることができる。例示されている実装では、属性識別子7215は、属性「Population」を識別するテキストである。それとともに、インスタンス識別子7210および属性識別子7215は、中国の人口が表示要素7200を使用して受け取った値によって特徴付けられると識別する。

【0487】

値入力要素7220は、インスタンス識別子7210によって識別されるインスタンスの属性識別子7215によって識別された属性を特徴付ける値をユーザーが指定することを可能にするインタラクティブ要素である。値入力要素7220は、例えば、テキスト入力フィールドであ

10

20

30

40

50

ってよい。

【0488】

値使用トリガー要素7225は、構造化プレゼンテーションにおいてインスタンス識別子7210によって識別されるインスタンスの属性識別子7215によって識別された属性を特徴付けるためにユーザーが値入力要素7220に入力された値の使用をトリガーすることを可能にするインタラクティブ要素である。値使用トリガー要素7225は、例えば、値使用トリガー要素7225に対するユーザーインタラクションの結果、値入力要素7220に入力された値が構造化プレゼンテーションにおいて使用されることを識別するテキストを含むボタンであるものとしてよい。

【0489】

プレゼンテーション閉じる要素7230は、表示要素7200をユーザーが閉じることを可能にするインタラクティブ要素である。表示要素7200は、プレゼンテーション閉じる要素7230に対するユーザーインタラクションにตอบสนองして、値入力要素7220に入力された値が、構造化プレゼンテーションにおいて、インスタンス識別子7210によって識別されるインスタンスの属性識別子7215によって識別された属性を特徴付けるために使用されるかどうかに関係なく閉じられる。プレゼンテーション閉じる要素7230は、例えば、プレゼンテーション閉じる要素7230に対するユーザーインタラクションにより表示要素7200が閉じられることを識別するテキストを含むボタンとすることができる。

【0490】

図73は、以前に隠されていた検索インターフェイスが提示される表示要素7300の図である。いくつかの実装では、構造化プレゼンテーションそれ自体に対するユーザーインタラクションにตอบสนองして、または以前に隠されていた検索インタラクティブ要素7005に対するユーザーインタラクションにตอบสนองして表示要素7300を提示することができる。表示要素7300は、構造化プレゼンテーションに追加するインスタンスの属性の値を受け取ることができる。ヘッダ7205、インスタンス識別子7210、属性識別子7215、値入力要素7220、値使用トリガー要素7225、およびプレゼンテーション閉じる要素7230に加えて、表示要素7300は、ソース入力要素7305およびソース入力要素識別子7310を含む。

【0491】

ソース入力要素7305は、インスタンス識別子7210によって識別されるインスタンスの属性識別子7215によって識別された属性を特徴付ける値のソースをユーザーが指定することを可能にするインタラクティブ要素である。ソース入力要素7305は、例えば、テキスト入力フィールドであってよい。ソース入力要素識別子7310は、ソース入力要素7305を使用して値のソースを指定することができることを記述するテキストまたは他の情報である。

【0492】

いくつかの実装では、表示要素7200、7300は、検索に失敗した後、ユーザーのために、ディスプレイ画面上に表示されうる。例えば、表示要素7200、7300は、前の検索の結果にユーザーが満足していないことを示す情報をユーザーから受け取ったことにตอบสนองして表示することができる。例えば、表示要素7200、7300の表示は、検索トリガー7005(図70)に対するユーザーインタラクションによってトリガーされうる。他の例として、表示要素7200、7300は、インスタンスの属性の値に対する自動検索が実行されたが満足できる結果が得られなかった場合に表示されうる。

【0493】

値の検索から満足できる結果が得られない理由は多数ある。例えば、属性、およびインスタンス、またはその両方の指定が、例えば、綴り違いまたは他の誤りにより正しくないことが考えられる。他の例として、属性またはインスタンスが、誤りはないが、未知の、または不定の値に関して指定される場合もある。例えば、インスタンス「suitable for Jim and Diane」は、不定の値、つまり、JimおよびDianeの識別、さらに彼らにとって「suitable」であるものの性質に関して指定される。他の例として、インスタンス「my car」は、不定の値、つまり、特徴付けられる自動車の持ち主の識別に関して指定される。

【0494】

10

20

30

40

50

値の検索により、事前検索の結果得られた電子文書が構造化プレゼンテーションに対する値を与えるようには機能しないために、満足できる結果が得られない場合もある。例えば、値が引き出されるソース文書が利用不可能になる可能性がある。ソース文書は、例えば、ソース文書を追加した当事者が電子文書コレクションからその文書を引き出すときに利用不可能になることがある。さらに他の例として、そのようなソース文書はまだ利用可能ではあるが、値それ自体はソース文書内で利用不可能になっている場合もある。値は、例えば、ソース文書を電子文書コレクションに追加した当事者がソース文書のコンテンツを変更したときに利用不可能になることがある。

【 0 4 9 5 】

図74は、以前に隠されていた検索インターフェイスが提示される表示要素7400の図である。表示要素7400は、ユーザーインタラクションに応答して、またはトリガーイベントに  
10 応答して自動的に提示されうる。例えば、表示要素7400は、事前検索が機能しなくなったことに応答して自動的に掲示されうる。

【 0 4 9 6 】

表示要素7400は、検索インタラクティブ要素6505、ソース識別子6605、およびウィンドウ6510内のエラーメッセージ7405を含む。検索インタラクティブ要素6505は、セル6220に書き込むために「more options...」が値の検索に利用可能であることを示すアンカーテキストを含むハイパーリンクである。ソース識別子6605は、セル6220に書き込む値307の  
20 ソースである電子文書を識別するテキストのコレクションである。

【 0 4 9 7 】

エラーメッセージ7405は、事前検索の結果が機能しなくなったことを示すテキストまたは他の情報を含むことができる。例えば、エラーメッセージ7405は、値307がソース識別子6605によって識別されるソース文書内で利用不可能になったことを示すことができる。エラーメッセージ7405は、機能不全の性質を記述するか、またはエラーが発生したことを単  
20 純に示す情報を含むことができる。例えば、例示されている実装では、エラーメッセージ7405は、電子文書それ自体はまだ利用可能なのに値がもはや電子文書内で利用できなくなっていることを示す。

【 0 4 9 8 】

図75は、電子文書コレクション内の文書のコンテンツに基づいて値を構造化プレゼンテーションに追加するためのプロセス7500の流れ図である。プロセス7500は、1つまたは複  
30 数の機械可読命令セットを実行することによってオペレーションを実行する1つまたは複数のコンピュータのシステムによって実行されうる。例えば、プロセス7500は、システム200内の検索エンジン202によって実行されうる。プロセス7500は、単独で、または他の活動と連携して実行されうる。例えば、プロセス7100は、プロセス6100(図61)の一部として実行されうる。

【 0 4 9 9 】

プロセス7500を実行するシステムは、事前に存在している構造化プレゼンテーションの1つまたは複数のセルの1つまたは複数の値の更新をトリガーする更新トリガーを受け取ることができる(ステップ7505)。更新トリガーは、例えば、前の更新以降一定期間経過した  
40 ことに応答して自動的に、ユーザーインタラクションに  
40 応答して手動で、または同様の形で生成されうる。例えば、上述のように、セル6220(図62、63、64)に対するユーザーインタラクションは、セルの更新をトリガーしうる。更新トリガーは、構造化プレゼンテーション内の値または単一のセル、セルのコレクションの値、またはすべてのセルの値の更新をトリガーしうる。それに付随して、上で説明されているように、更新トリガーにより、1つまたは複数のインスタンスの1つまたは複数の属性を指定することができる。

【 0 5 0 0 】

プロセス7500を実行するシステムは、構造化プレゼンテーションに値を書き込むために1つまたは複数の事前検索が機能しなくなっているかどうかを判定することができる(ステップ7510)。このような判定は、構造化プレゼンテーションに書き込む値が引き出される  
50 文書にアクセスしようとすることによって行われうる。

## 【0501】

プロセス7100を実行するシステムが、事前検索が機能しなくなっていると判定した場合、システムは、事前検索で識別されている1つまたは複数のソース文書のコンテンツで構造化プレゼンテーションを更新することができる(ステップ7515)。構造化プレゼンテーションを更新するために使用される新規値は、構造化プレゼンテーションに書き込むためにすでに使用されている値と同一である必要はない。むしろ、更新された構造化プレゼンテーションは、その現在のコンテンツとともにソース電子文書によって提供される値を含むことができる。

## 【0502】

プロセス7100を実行するシステムが、事前検索が機能しなくなったと判定した場合、システムは、事前検索が操作不能であることをユーザーに通知することができる(ステップ7520)。例えば、表示要素7400など表示要素を使用して、操作可能であることをユーザーに通知ことができ、またこれにより、ユーザーは新規検索を実行して構造化プレゼンテーションに値を書き込む機会を得ることができる。いくつかの実装では、システムは、プロセス7100(図71)のステップ7120を参照しつつ説明されているように、新規検索を実行し、その新規検索によって得られた結果セット内の1つまたは複数の電子文書の特徴付ける情報を供給することができる。

10

## 【0503】

そこで、本明細書の表示および主題は、例えば、以下に番号付きで示されている実施形態を含む、多数の本発明の実施形態により実現することができる。

20

## 【0504】

実施形態1:機械実装方法であって、この方法は、事前に存在している構造化プレゼンテーションを記述する記述データを受け取る段階であって、事前に存在している構造化プレゼンテーションの視覚的提示は、構造化設計に適合するシステム配置で情報を視覚的に提示し、構造化プレゼンテーションは、構造化プレゼンテーションの視覚的提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す、段階と、事前に存在している構造化プレゼンテーションの特性を電子文書の非構造化コレクション内の電子文書のコンテンツと比較して事前に存在している構造化プレゼンテーションに関連する新規インスタンスを識別する電子文書を特定する段階と、新規インスタンスの識別子を事前に存在している構造化プレゼンテーションに追加して拡張構造化プレゼンテーションを形成する段階と、拡張構造化プレゼンテーションをディスプレイデバイス上に提示するための命令を出力する段階とを含む、機械実装方法。

30

## 【0505】

実施形態2:新規インスタンスの識別子を追加する段階は、インスタンス提案のコレクションを構成する段階と、インスタンス提案コレクションをユーザーに提供する段階と、新規インスタンスのユーザー選択を受け取る段階とを含み、新規インスタンスはインスタンス提案のコレクションに含まれる、実施形態1の方法。

## 【0506】

実施形態3:事前に存在している構造化プレゼンテーションの特性を電子文書のコンテンツと比較する段階は、事前に存在している構造化プレゼンテーションで識別されたインスタンスに関係する構造化コンポーネントを含む電子文書コレクション内の文書を識別する段階を含む、実施形態2の方法。

40

## 【0507】

実施形態4:インスタンス提案のコレクションを構成する段階は、事前に存在している構造化プレゼンテーションで識別されたインスタンスの識別子を含み、テンプレートに従って配列されている電子文書コレクション内の第1の文書を識別する段階と、テンプレートに従って配列されているが、第2のインスタンスに関連している第2の文書を識別する段階と、第2のインスタンスをインスタンス提案コレクションに含める段階とを含む、実施形態2の方法。

## 【0508】

50

実施形態5: 事前に存在している構造化プレゼンテーションの特性を電子文書のコンテンツと比較する段階は、事前に存在している構造化プレゼンテーション内の1つまたは複数のインスタンスに関する情報を含む電子文書コレクション内の文書を識別する段階を含む、実施形態1の方法。

【0509】

実施形態6: 事前に存在している構造化プレゼンテーションの特性を電子文書のコンテンツと比較する段階は、インスタンスと属性との関連付けの格納されているコレクション内の新規インスタンスを特定する段階を含む、実施形態1の方法。

【0510】

実施形態7: 事前に存在している構造化プレゼンテーションの特性を電子文書のコンテンツと比較する段階は、事前に存在している構造化プレゼンテーションの特性を事前に存在している構造化プレゼンテーションにおいて特徴付けられている属性と比較する段階を含む、実施形態1の方法。

10

【0511】

実施形態8: 事前に存在している構造化プレゼンテーションの特性を電子文書のコンテンツと比較する段階は、事前に存在している構造化プレゼンテーション内のインスタンスを特徴付けるために使用される属性を電子文書のコンテンツと比較する段階を含む、実施形態1の方法。

【0512】

実施形態9: 事前に存在している構造化プレゼンテーションの特性を電子文書のコンテンツと比較する段階は、事前に存在している構造化プレゼンテーション内のインスタンスを特徴付けるために使用される属性の値を電子文書のコンテンツと比較する段階を含む、実施形態1の方法。

20

【0513】

実施形態10: 事前に存在している構造化プレゼンテーションの特性を電子文書のコンテンツと比較する段階は、事前に存在している構造化プレゼンテーション内のインスタンスを含むインスタンスのカテゴリを電子文書のコンテンツと比較する段階を含む、実施形態1の方法。

【0514】

実施形態11: 電子文書のコレクションは、インターネット上で利用可能な電子文書を含み、電子文書は、ウェブページを含む、実施形態1の方法。

30

【0515】

実施形態12: 拡張構造化プレゼンテーションは、テーブルを含む、実施形態1の方法。

【0516】

実施形態13: 拡張構造化プレゼンテーションは、カードのコレクションを含む、実施形態1の方法。

【0517】

実施形態14: ディスプレイ画面の1つまたは複数の要素を物理的に変換する段階を含めて、拡張構造化プレゼンテーションをディスプレイ画面上に視覚的に提示する段階をさらに含む、実施形態1の方法。

40

【0518】

実施形態15: 1つまたは複数のデータ処理機にオペレーションを実行させるように動作可能な命令を格納する1つまたは複数の機械可読データ記憶媒体を備える装置であって、これらのオペレーションは、非構造化電子文書コレクション内の2つまたはそれ以上の文書のコンテンツに基づいてインスタンス提案のコレクションを構成する段階であって、電子文書コレクションは、電子文書コレクション内の電子文書のフォーマットが制限的でも永続的でもないという点で非構造化コレクションである、段階と、インスタンス提案コレクションをユーザーに提供する段階と、インスタンス提案のコレクション内の第1のインスタンスのユーザー選択を受け取る段階と、第1のインスタンス提案の識別子をディスプレイデバイス上に提示される構造化プレゼンテーションに追加する段階であって、構造化プ

50

プレゼンテーションの視覚的提示により、情報が整理された形で視覚的に提示され、構造化プレゼンテーションは、構造化プレゼンテーションの視覚的提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す、段階とを含む。

【0519】

実施形態16:インスタンス提案のコレクションを構成する段階は、事前に存在している構造化プレゼンテーションの特性を電子文書コレクション内の電子文書のコンテンツと比較する段階を含む、実施形態15の装置。

【0520】

実施形態17:インスタンス提案のコレクションを構成する段階は、事前に存在している構造化プレゼンテーション内で識別されているインスタンスに関する構造化コンポーネントを含む電子文書コレクション内の文書を識別する段階を含む、実施形態16の装置。

10

【0521】

実施形態18:インスタンス提案コレクションを構成する段階は、事前に存在している構造化プレゼンテーションで識別されたインスタンスに関連し、テンプレートに従って配列されている、電子文書コレクション内の第1の文書を識別する段階と、テンプレートに従って配列されているが、第2のインスタンスに関連している第2の文書を識別する段階と、第2のインスタンスをインスタンス提案コレクションに含める段階とを含む、実施形態16の装置。

【0522】

20

実施形態19:インスタンス提案のコレクションを構成する段階は、事前に存在している構造化プレゼンテーション内の1つまたは複数のインスタンスの識別子を含む電子文書コレクション内の文書を識別する段階を含む、実施形態16の装置。

【0523】

実施形態20:インスタンス提案のコレクションを構成する段階は、事前に存在している構造化プレゼンテーション内のインスタンスを特徴付けるために使用される追加の属性を識別する段階を含む、実施形態16の装置。

【0524】

実施形態21:インスタンス提案のコレクションを構成する段階は、事前に存在している構造化プレゼンテーション内のインスタンスを特徴付けるために使用される属性の値をインスタンス提案の値と比較する段階を含む、実施形態16の装置。

30

【0525】

実施形態22:インスタンス提案のコレクションを構成する段階は、事前に存在している構造化プレゼンテーション内のインスタンスを含むインスタンスのカテゴリを識別する段階と、インスタンスのそのカテゴリ内のインスタンスを使用してインスタンス提案のコレクションを構成する段階とを含む、実施形態16の装置。

【0526】

実施形態23:インスタンス提案のコレクションを構成する段階は、インスタンスと属性との関連付けの格納されているコレクション内のインスタンス提案を識別する段階を含む、実施形態15の装置。

40

【0527】

実施形態24:インスタンス提案のコレクションを構成する段階は、事前に存在している構造化プレゼンテーションにおいて特徴付けられている属性を非構造化電子文書コレクション内の文書のコンテンツと比較する段階を含む、実施形態15の装置。

【0528】

実施形態25:電子文書のコレクションは、インターネット上で利用可能な文書を含み、電子文書は、ウェブページを含む、実施形態15の装置。

【0529】

実施形態26:構造化プレゼンテーションは、テーブルを含む、実施形態15の装置。

【0530】

50

実施形態27:構造化プレゼンテーションは、カードのコレクションを含む、実施形態15の装置。

【0531】

実施形態28:クライアントデバイスと、クライアントデバイスをインタラクティブに操作し、オペレーションを実行するようにプログラムされた1つまたは複数のコンピュータであって、これらのオペレーションは、事前に存在している構造化プレゼンテーションを記述する記述データを受け取る段階であって、事前に存在している構造化プレゼンテーションの視覚的提示は、構造化設計に適合するシステム配置で情報を視覚的に提示し、構造化プレゼンテーションは、構造化プレゼンテーションの視覚的提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す、段階と、事前に存在している構造化プレゼンテーションの特性を電子文書の非構造化コレクション内の電子文書のコンテンツと比較して事前に存在している構造化プレゼンテーションに関連する新規インスタンスを識別する電子文書を特定する段階と、新規インスタンスの識別子を事前に存在している構造化プレゼンテーションに追加して拡張構造化プレゼンテーションを形成する段階と、クライアントデバイスとデータ通信を行うように結合されているディスプレイデバイス上に拡張構造化プレゼンテーションを提示するための命令を出力する段階とを含む、1つまたは複数のコンピュータとを備えるシステム

10

【0532】

実施形態29:クライアントデバイスと、クライアントデバイスをインタラクティブに操作し、オペレーションを実行するようにプログラムされた1つまたは複数のコンピュータであって、これらのオペレーションは、非構造化電子文書コレクション内の2つまたはそれ以上の文書のコンテンツに基づいてインスタンス提案のコレクションを構成する段階であって、電子文書コレクションは、電子文書コレクション内の電子文書のフォーマットが制限的でも永続的でもないという点で非構造化コレクションである、段階と、クライアントデバイスを使用してインスタンス提案コレクションをユーザーに提供する段階と、インスタンス提案のコレクション内の第1のインスタンスのユーザー選択を受け取る段階と、クライアントデバイスとデータ通信を行うように結合されているディスプレイデバイス上に提示される構造化プレゼンテーションに第1のインスタンス提案の識別子を追加する段階とを含み、構造化プレゼンテーションの視覚的提示により、情報が整理された形で視覚的に提示され、構造化プレゼンテーションは、構造化プレゼンテーションの視覚的提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す、段階とを含む、1つまたは複数のコンピュータとを備えるシステム。

20

30

【0533】

実施形態30:1つまたは複数のコンピュータは、データ通信ネットワークを通じてクライアントデバイスをインタラクティブに操作するように動作可能なサーバーを含み、クライアントデバイスは、クライアントとしてサーバーをインタラクティブに操作するように動作可能であり、クライアントデバイスは、ウェブブラウザを実行するパーソナルコンピュータを含み、パーソナルコンピュータは、ディスプレイデバイスを備える、実施形態29のシステム。

40

【0534】

実施形態31:機械実装方法であって、この方法は、事前に存在している構造化プレゼンテーションを記述する記述データを受け取る段階であって、事前に存在している構造化プレゼンテーションの視覚的提示は、構造化設計に適合するシステム配置で情報を視覚的に提示し、構造化プレゼンテーションは、構造化プレゼンテーションの視覚的提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す、段階と、事前に存在している構造化プレゼンテーションの特性を電子文書の非構造化コレクション内の電子文書のコンテンツと比較して事前に存在している構造化プレゼンテーションに関連する新規属性を識別する電子文書を特定する段階と、

50

新規属性の識別子を事前に存在している構造化プレゼンテーションに追加して拡張構造化プレゼンテーションを形成する段階と、拡張構造化プレゼンテーションをディスプレイ画面上に提示するための命令を出力する段階とを含む機械実装方法。

【0535】

実施形態32:新規属性の識別子を追加する段階は、属性提案のコレクションを構成する段階と、属性提案コレクションをユーザーに提供する段階と、新規属性のユーザー選択を受け取る段階とを含み、新規属性はインスタンス提案のコレクションに含まれる、実施形態31の方法。

【0536】

実施形態33:属性提案コレクションを構成する段階は、事前に存在している構造化プレゼンテーションで識別されたインスタンスに関連し、テンプレートに従って配列されている、電子文書コレクション内の第1の文書を識別する段階と、属性提案コレクション内のインスタンスを特徴付けるために第1の文書内で使用される属性を追加する段階とを含む、実施形態32の方法。

10

【0537】

実施形態34:事前に存在している構造化プレゼンテーションの特性を電子文書のコンテンツと比較する段階は、事前に存在している構造化プレゼンテーションで識別されたインスタンスに係る構造化コンポーネントを含む電子文書コレクション内の文書を識別する段階を含む、実施形態31の方法。

【0538】

実施形態35:事前に存在している構造化プレゼンテーションの特性を電子文書のコンテンツと比較する段階は、事前に存在している構造化プレゼンテーション内の1つまたは複数のインスタンスに関する情報を含む電子文書コレクション内の文書を識別する段階を含む、実施形態31の方法。

20

【0539】

実施形態36:事前に存在している構造化プレゼンテーションの特性を電子文書のコンテンツと比較する段階は、インスタンスと属性との関連付けの格納されているコレクション内の新規属性を識別する段階を含む、実施形態31の方法。

【0540】

実施形態37:事前に存在している構造化プレゼンテーションの特性を電子文書のコンテンツと比較する段階は、事前に存在している構造化プレゼンテーションにおいて特徴付けられているインスタンスを電子文書のコンテンツと比較する段階を含む、実施形態31の方法。

30

【0541】

実施形態38:事前に存在している構造化プレゼンテーションの特性を電子文書のコンテンツと比較する段階は、事前に存在している構造化プレゼンテーションにおいて識別されているインスタンスに係る追加のインスタンスを識別する段階を含む、実施形態31の方法。

【0542】

実施形態39:事前に存在している構造化プレゼンテーションの特性を電子文書のコンテンツと比較する段階は、事前に存在している構造化プレゼンテーション内のインスタンスを特徴付けるために使用される属性または属性の値を電子文書のコンテンツと比較する段階を含む、実施形態31の方法。

40

【0543】

実施形態40:事前に存在している構造化プレゼンテーションの特性を電子文書のコンテンツと比較する段階は、事前に存在している構造化プレゼンテーション内のインスタンスを含むインスタンスのカテゴリを電子文書のコンテンツと比較する段階を含む、実施形態31の方法。

【0544】

実施形態41:電子文書のコレクションは、インターネット上で利用可能な電子文書を含

50

み、電子文書は、ウェブページを含む、実施形態31の方法。

【0545】

実施形態42:拡張構造化プレゼンテーションは、テーブルを含む、実施形態31の方法。

【0546】

実施形態43:拡張構造化プレゼンテーションは、カードのコレクションを含む、実施形態31の方法。

【0547】

実施形態44:ディスプレイ画面の1つまたは複数の要素を物理的に変換する段階を含めて、拡張構造化プレゼンテーションをディスプレイ画面上に視覚的に提示する段階をさらに含む、実施形態31の方法。

【0548】

実施形態45:1つまたは複数のデータ処理機にオペレーションを実行させるように動作可能な命令を格納する1つまたは複数の機械可読データ記憶媒体を備える装置であって、これらのオペレーションは、非構造化電子文書コレクション内の2つまたはそれ以上の文書のコンテンツに基づいて属性提案のコレクションを構成する段階であって、電子文書コレクションは、電子文書コレクション内の電子文書のフォーマットが制限的でも永続的でもないという点で非構造化コレクションである、段階と、属性提案コレクションをユーザーに提供する段階と、属性提案のコレクション内の第1の属性のユーザー選択を受け取る段階と、第1の属性提案の識別子をディスプレイ画面上に提示される構造化プレゼンテーションに追加する段階であって、構造化プレゼンテーションの視覚的提示により、情報が整理された形で視覚的に提示され、構造化プレゼンテーションは、構造化プレゼンテーションの提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す、段階とを含む、装置。

【0549】

実施形態46:属性提案のコレクションを構成する段階は、事前に存在している構造化プレゼンテーションの特性を電子文書コレクション内の電子文書のコンテンツと比較する段階を含む、実施形態45の装置。

【0550】

実施形態47:属性提案のコレクションを構成する段階は、事前に存在している構造化プレゼンテーション内で識別されているインスタンスを特徴付ける構造化コンポーネントを含む電子文書コレクション内の文書を識別する段階を含む、実施形態46の装置。

【0551】

実施形態48:属性提案コレクションを構成する段階は、事前に存在している構造化プレゼンテーションで識別されたインスタンスに関連し、テンプレートに従って配列されている、電子文書コレクション内の第1の文書を識別する段階と、属性提案コレクション内のインスタンスを特徴付けるために使用される属性を含める段階を含む、実施形態46の装置。

【0552】

実施形態49:属性提案のコレクションを構成する段階は、事前に存在している構造化プレゼンテーション内の1つまたは複数のインスタンスに関する情報を含む電子文書コレクション内の文書を識別する段階を含む、実施形態46の装置。

【0553】

実施形態50:事前に存在している構造化プレゼンテーションの特性を電子文書のコンテンツと比較する段階は、事前に存在している構造化プレゼンテーションにおいて識別されているインスタンスを電子文書のコンテンツと比較する段階を含む、実施形態46の装置。

【0554】

実施形態51:事前に存在している構造化プレゼンテーションの特性を電子文書のコンテンツと比較する段階は、事前に存在している構造化プレゼンテーション内のインスタンスを特徴付けるために使用される属性または属性の値を電子文書のコンテンツと比較する段階を含む、実施形態46の装置。

10

20

30

40

50

## 【 0 5 5 5 】

実施形態52:属性提案のコレクションを構成する段階は、事前に存在している構造化プレゼンテーション内のインスタンスを含むインスタンスのカテゴリを識別する段階と、インスタンスのそのカテゴリ内のインスタンスを特徴付けるために使用される属性から属性提案のコレクションを構成する段階とを含む、実施形態46の装置。

## 【 0 5 5 6 】

実施形態53:属性提案のコレクションを構成する段階は、インスタンスと属性との関連付けの格納されているコレクション内の属性提案を識別する段階を含む、実施形態45の装置。

## 【 0 5 5 7 】

実施形態54:電子文書のコレクションは、インターネット上で利用可能な電子文書を含み、電子文書は、ウェブページを含む、実施形態45の装置。

## 【 0 5 5 8 】

実施形態55:構造化プレゼンテーションは、テーブルを含む、実施形態45の装置。

## 【 0 5 5 9 】

実施形態56:構造化プレゼンテーションは、カードのコレクションを含む、実施形態45の装置。

## 【 0 5 6 0 】

実施形態57:ディスプレイ画面を備えるクライアントデバイスと、クライアントデバイスをインタラクティブに操作し、オペレーションを実行するようにプログラムされた1つまたは複数のコンピュータであって、これらのオペレーションは、事前に存在している構造化プレゼンテーションを記述する記述データを受け取る段階であって、事前に存在している構造化プレゼンテーションの視覚的提示は、構造化設計に適合するシステム配置で情報を視覚的に提示し、構造化プレゼンテーションは、構造化プレゼンテーションの視覚的提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す、段階と、事前に存在している構造化プレゼンテーションの特性を電子文書の非構造化コレクション内の電子文書のコンテンツと比較して事前に存在している構造化プレゼンテーションに関連する新規属性を識別する電子文書を特定する段階と、新規属性の識別子を事前に存在している構造化プレゼンテーションに追加して拡張構造化プレゼンテーションを形成する段階と、拡張構造化プレゼンテーションをディスプレイ画面上に提示するための命令を出力する段階とを含む、1つまたは複数のコンピュータとを備えるシステム。

## 【 0 5 6 1 】

実施形態58:ディスプレイ画面を備えるクライアントデバイスと、クライアントデバイスをインタラクティブに操作し、オペレーションを実行するようにプログラムされた1つまたは複数のコンピュータであって、これらのオペレーションは、非構造化電子文書コレクション内の2つまたはそれ以上の文書のコンテンツに基づいて属性提案のコレクションを構成する段階であって、電子文書コレクションは、電子文書コレクション内の電子文書のフォーマットが制限的でも永続的でもないという点で非構造化コレクションである、段階と、属性提案コレクションをクライアントデバイスに提供する段階と、クライアントデバイスから属性提案のコレクション内の第1の属性の選択を受け取る段階と、第1の属性提案の識別子をディスプレイ画面上に提示される構造化プレゼンテーションに追加する段階であって、構造化プレゼンテーションの視覚的提示により、情報が整理された形で視覚的に提示され、構造化プレゼンテーションは、構造化プレゼンテーションの視覚的提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す、段階とを含む、1つまたは複数のコンピュータとを備えるシステム。

## 【 0 5 6 2 】

実施形態59:機械実装方法であって、この方法は、事前に存在している構造化プレゼンテーションを記述する記述データを受け取る段階であって、事前に存在している構造化プ

10

20

30

40

50

プレゼンテーションの視覚的提示は、構造化設計に適合するシステム配置で情報を視覚的に提示し、構造化プレゼンテーションは、構造化プレゼンテーションの視覚的提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す、段階と、事前に存在している構造化プレゼンテーションの特性を電子文書の非構造化コレクション内の電子文書のコンテンツと比較して事前に存在している構造化プレゼンテーションに関連する新規値を識別する電子文書を特定する段階と、新規値を事前に存在している構造化プレゼンテーションに追加して新規構造化プレゼンテーションを形成する段階と、新規構造化プレゼンテーションを視覚的に提示するための命令を出力するアクションとを含む、機械実装方法。

【0563】

10

実施形態60:事前に存在している構造化プレゼンテーションの特性を電子文書のコンテンツと比較する段階は、第1の電子文書において構造化プレゼンテーション内に出現する第1のインスタンスの識別子を特定する段階を含み、この方法は、新規値を第1の電子文書から抽出する段階をさらに含む、実施形態59の方法。

【0564】

実施形態61:新規値を追加する段階は、第1のインスタンスの第1の属性の値のコレクションを識別する段階と、識別された値のうちの一つまたは複数の値のサブセットを、第1のインスタンスの第1の属性を適切に特徴付けるものとして確定する段階とを含む、実施形態59の方法。

【0565】

20

実施形態62:値のサブセットを適切なものとして確定する段階は、コレクション内の値をいくつかのグループに分ける段階を含む、実施形態61の方法。

【0566】

実施形態63:値のサブセットを適切なものとして確定する段階は、サブセット内の値の個数に少なくとも一部は基づいてサブセットを選択する段階を含む、実施形態61の方法。

【0567】

実施形態64:値のサブセットを適切なものとして確定する段階は、ユーザー指定制約条件を満たすサブセット内の値に少なくとも一部は基づいてサブセットを選択する段階を含む、実施形態61の方法。

【0568】

30

実施形態65:値のサブセットを適切なものとして確定する段階は、サブセット内の値が高品質文書から引き出される段階に少なくとも一部は基づいてサブセットを選択する段階を含む、実施形態61の方法。

【0569】

実施形態66:値のサブセットを適切なものとして確定する段階は、サブセット内の値が事前に存在している構造化プレゼンテーション内の他のインスタンスに関連する文書から引き出される段階に少なくとも一部は基づいてサブセットを選択する段階を含む、実施形態61の方法。

【0570】

40

実施形態67:値のサブセットを適切なものとして確定する段階は、サブセット内の値が事前に存在している構造化プレゼンテーション内の他の属性に関連する文書から引き出される段階に少なくとも一部は基づいてサブセットを選択する段階を含む、実施形態61の方法。

【0571】

実施形態68:電子文書のコレクションは、インターネット上で利用可能な電子文書を含み、電子文書は、ウェブページを含む、実施形態59の方法。

【0572】

実施形態69:事前に存在している構造化プレゼンテーションは、テーブルを含む、実施形態59の方法。

【0573】

50

実施形態70: 事前に存在している構造化プレゼンテーションは、カードのコレクションを含む、実施形態59の方法。

【0574】

実施形態71: ディスプレイ画面の1つまたは複数の要素を物理的に変換する段階を含めて、新規構造化プレゼンテーションをディスプレイ画面上に視覚的に提示する段階をさらに含む、実施形態59の方法。

【0575】

実施形態72: 1つまたは複数のデータ処理機にオペレーションを実行させるように動作可能な命令を格納する1つまたは複数の機械可読データ記憶媒体を備える装置であって、これらのオペレーションは、第1のインスタンス、第2のインスタンス、および第1の属性を記述する記述データを受け取る段階と、非構造化電子文書コレクションの2つまたはそれ以上の文書から第1のインスタンスの第1の属性の値の第1のコレクションを抽出する段階と、非構造化電子文書コレクションの2つまたはそれ以上の文書から第2のインスタンスの第1の属性の値の第2のコレクションを抽出する段階と、値の第1のコレクションの第1のサブセットを、第1のインスタンスの第1の属性を適切に特徴付けるものとして確定する段階と、値の第2のコレクションの第2のサブセットを、第2のインスタンスの第1の属性を適切に特徴付けるものとして確定する段階と、第1のサブセットの第1の値および第2のサブセットの第2の値を含む構造化プレゼンテーションを表示するための機械可読命令を生成する段階であって、構造化プレゼンテーションは、インスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す、段階とを含む、装置。

10

20

【0576】

実施形態73: 値の第1のサブセットを適切なものとして確定する段階は、第1のコレクション内の値をいくつかのグループに分ける段階を含み、それぞれのグループは値の第1のコレクションのサブセットを含む、実施形態72の装置。

【0577】

実施形態74: 値の第1のサブセットを適切なものとして確定する段階は、第1のサブセット内の値の個数に少なくとも一部は基づき第1のサブセットを選択する段階を含む、実施形態72の装置。

【0578】

実施形態75: 値の第1のサブセットを適切なものとして確定する段階は、第1のサブセット内の値をそれらの値に対するユーザー指定制約条件と比較する段階を含む、実施形態72の装置。

30

【0579】

実施形態76: 値の第1のサブセットを適切なものとして確定する段階は、第1のサブセット内の値が高品質文書から引き出されたと判定する段階を含む、実施形態72の装置。

【0580】

実施形態77: 値の第1のサブセットを適切なものとして確定する段階は、第1のサブセット内の値が第2のインスタンスに関連する文書から引き出されたと判定する段階を含む、実施形態72の装置。

40

【0581】

実施形態78: 値の第1のサブセットを適切なものとして確定する段階は、第1のサブセット内の値が第1のインスタンスおよび第2のインスタンスの両方を特徴付ける他の属性に関連する文書から引き出されたと判定する段階を含む、実施形態72の装置。

【0582】

実施形態79: 第1のインスタンスの記述は、事前に存在している構造化プレゼンテーション内に出現する第1のインスタンスの識別子を含み、第2のインスタンスの記述は、事前に存在している構造化プレゼンテーション内に出現する第2のインスタンスの識別子を含む、実施形態72の装置。

【0583】

50

実施形態80:第1の属性の記述は、事前に存在している構造化プレゼンテーションに追加される新規属性の記述を含む、実施形態72の装置。

【0584】

実施形態81:非構造化電子文書コレクションは、インターネット上で利用可能な電子文書を含む、実施形態72の装置。

【0585】

実施形態82:構造化プレゼンテーションは、テーブルを含む、実施形態72の装置。

【0586】

実施形態83:構造化プレゼンテーションは、カードのコレクションを含む、実施形態72の装置。

10

【0587】

実施形態84:ディスプレイ画面の1つまたは複数の要素を物理的に変換する段階を含めて、構造化プレゼンテーションをディスプレイ画面上に視覚的に提示する段階をさらに含む、実施形態72の装置。

【0588】

実施形態85:デバイスと、デバイスをインタラクティブに操作し、オペレーションを実行するようにプログラムされた1つまたは複数のコンピュータであって、これらのオペレーションは、事前に存在している構造化プレゼンテーションを記述する記述データを受け取る段階であって、事前に存在している構造化プレゼンテーションの視覚的提示は、構造化設計に適合するシステム配置で情報を視覚的に提示し、構造化プレゼンテーションは、構造化プレゼンテーションの視覚的提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す、段階と、事前に存在している構造化プレゼンテーションの特性を電子文書の非構造化コレクション内の電子文書のコンテンツと比較して事前に存在している構造化プレゼンテーションに関連する新規値を識別する電子文書を特定する段階と、新規値を事前に存在している構造化プレゼンテーションに追加して新規構造化プレゼンテーションを形成する段階と、新規構造化プレゼンテーションをデバイス上に視覚的に提示するための命令を出力する段階とを含む、1つまたは複数のコンピュータとを備える、システム。

20

【0589】

実施形態86:デバイスと、デバイスをインタラクティブに操作し、オペレーションを実行するようにプログラムされた1つまたは複数のコンピュータであって、これらのオペレーションは、第1のインスタンス、第2のインスタンス、および第1の属性を記述する記述データを受け取る段階と、非構造化電子文書コレクションの2つまたはそれ以上の文書から第1のインスタンスの第1の属性の値の第1のコレクションを抽出する段階と、非構造化電子文書コレクションの2つまたはそれ以上の文書から第2のインスタンスの第1の属性の値の第2のコレクションを抽出する段階と、値の第1のコレクションの第1のサブセットを、第1のインスタンスの第1の属性を適切に特徴付けるものとして確定する段階と、値の第2のコレクションの第2のサブセットを、第2のインスタンスの第1の属性を適切に特徴付けるものとして確定する段階と、第1のサブセットの第1の値および第2のサブセットの第2の値を含む構造化プレゼンテーションを表示するための機械可読命令を生成する段階であって、構造化プレゼンテーションは、インスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す、段階と、機械可読命令をデバイスに送信する段階とを含む、1つまたは複数のコンピュータとを備える、システム。

30

40

【0590】

実施形態87:機械実装方法であって、この方法は、構造化プレゼンテーションをディスプレイデバイス上に表示する段階であって、構造化プレゼンテーションは、構造化設計に適合する組織的な構造化された配列で情報を視覚的に提示し、構造化プレゼンテーションは、構造化プレゼンテーションの視覚的提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す、段階と、表

50

示される構造化プレゼンテーションに対するユーザーインタラクションを特徴付ける、構造化プレゼンテーションの第1のインスタンスおよび第1の属性の指定を含むデータを受け取る段階と、データを受け取ったことに応答してディスプレイデバイス上に以前には隠されていた検索インターフェイスを表示する段階であって、検索インターフェイスは電子文書コレクション内の第1のインスタンスの第1の属性を特徴付ける第1の値の位置を識別する情報もしくはインタラクティブ要素を含む、段階とを含む、機械実装方法。

【0591】

実施形態88:表示される構造化プレゼンテーションに対するユーザーインタラクションを特徴付けるデータを受け取る段階は、構造化プレゼンテーション内のセルに関連付けられている第1のインスタンスおよび第1の属性の手動によるユーザー指定を受け取る段階を含む、実施形態87の方法。

10

【0592】

実施形態89:ユーザーインタラクションを特徴付けるデータを受け取る段階は、構造化プレゼンテーション内のセルに対するユーザーインタラクションを特徴付けるデータを受け取る段階を含み、セルは構造化プレゼンテーション内の第1のインスタンスおよび第1の属性の識別子に関するセルの配列を用いて第1のインスタンスおよび第1の属性に関連付けられる、実施形態87の方法。

【0593】

実施形態90:セルに対するユーザーインタラクションを特徴付けるデータを受け取る段階は、空のセルに対するユーザーインタラクションを特徴付けるデータを受け取る段階を含む、実施形態89の方法。

20

【0594】

実施形態91:以前には隠されていた検索インターフェイスを表示する段階は、電子文書コレクションの検索をトリガーし、第1の値を特定するためにユーザー側で選択できるインタラクティブ要素を表示する段階を含む、実施形態87の方法。

【0595】

実施形態92:以前には隠されていた検索インターフェイスを表示する段階は、第1のインスタンスの第1の属性を特徴付ける値を指定するためにユーザー側で選択できるインタラクティブ入力要素を表示する段階を含む、実施形態87の方法。

【0596】

実施形態93:以前には隠されていた検索インターフェイスを表示する段階は、電子文書コレクションの第1の文書内の第1の値のコンテキストを特徴付けるスニペットを表示する段階を含む、実施形態87の方法。

30

【0597】

実施形態94:以前には隠されていた検索インターフェイスを表示する段階は、第1の値を特定するために電子文書コレクションの事前検索の結果を表示する段階を含む、実施形態87の方法。

【0598】

実施形態95:第1の値は、第1のインスタンスの第1の属性を特徴付ける値として構造化プレゼンテーション内に出現し、以前には隠されていた検索インターフェイスを表示する段階は、電子文書コレクション内の第1の電子文書の識別子を表示する段階を含み、第1の値は、第1の電子文書から引き出される、実施形態87の方法。

40

【0599】

実施形態96:第1の電子文書が機能せず第1の値を提供することができないと判定する段階と、第1の文書が機能しないことを示す視覚的指示を表示する段階とをさらに含む、実施形態87の方法。

【0600】

実施形態97:以前には隠されていた検索インターフェイスを表示する段階は、第1の文書と第1のインスタンスおよび第1の属性との関連性の変化に関係なく第1の文書から矛盾なく第1の値を選択するオプションをユーザーに提示する段階を含む、実施形態87の方法。

50

## 【0601】

実施形態98:以前には隠されていた検索インターフェイスを表示する段階は、第1のインスタンスおよび第1の属性に最も関連している第1の文書から第1の値を選択するオプションをユーザーに提示する段階を含む、実施形態87の方法。

## 【0602】

実施形態99:検索インターフェイスに対するユーザーインタラクションにตอบสนองして第1の値を特定するために電子文書の非構造化コレクションを検索する段階と、第1の値を構造化プレゼンテーションに追加する段階とをさらに含む、実施形態87の方法。

## 【0603】

実施形態100:第1のインスタンスおよび第1の属性の指定を受け取る段階は、属性のコレクションまたはインスタンスのコレクションの指定を受け取る段階を含む、実施形態87の方法。

10

## 【0604】

実施形態101:時間の経過に応じて構造化プレゼンテーションの表示を更新する段階をさらに含む、実施形態87の方法。

## 【0605】

実施形態102:クライアントデバイスをインタラクティブに操作し、オペレーションを実行するようにプログラムされた1つまたは複数のコンピュータであって、これらのオペレーションは、ディスプレイデバイス上に表示される構造化プレゼンテーションの第1のセルを指定するユーザーインタラクションを特徴付けるデータを受け取る段階であって、構造化プレゼンテーションは、構造化設計に適合する組織的な構造化された配列で情報を視覚的に提示し、構造化プレゼンテーションは、インスタンスとインスタンスの属性を特徴付ける値との間の関連付けをセル内の値の配列によって表す、段階と、第1のセルに第1の値を書き込むために事前検索が実施されたと判定する段階と、事前検索が実施されたと判定したことに応じて、ディスプレイデバイス上に事前検索を特徴付ける情報を表示する段階とを含む、1つまたは複数のコンピュータを備える、システム。

20

## 【0606】

実施形態103:第1のセルを指定するユーザーインタラクションを特徴付けるデータを受け取る段階は、第1のセルに関連付けられている第1のインスタンスおよび第1の属性の手动によるユーザー指定を特徴付けるデータを受け取る段階を含む、実施形態102のシステム。

30

## 【0607】

実施形態104:事前検索を特徴付ける情報を表示する段階は、第1の値の引き出し元となる電子文書を識別する情報を表示する段階を含む、実施形態102のシステム。

## 【0608】

実施形態105:事前検索を特徴付ける情報を表示する段階は、第1の値の引き出し元になった可能性のある電子文書のコレクションを識別する情報を表示する段階を含む、実施形態102のシステム。

## 【0609】

実施形態106:事前検索を特徴付ける情報を表示する段階は、以前には隠されていた検索インターフェイスの表示要素内に情報を表示する段階を含む、実施形態102のシステム。

40

## 【0610】

実施形態107:事前検索を特徴付ける情報を表示する段階は、第1の値の引き出し元となる電子文書コレクション内の第1の電子文書を識別する情報を表示する段階を含む、実施形態102のシステム。

## 【0611】

実施形態108:これらのオペレーションは、第1の電子文書が操作不能であり第1の値を提供することができないと判定する段階と、第1の文書が操作不能であることを示す視覚的指示を表示する段階とをさらに含む、実施形態107のシステム。

## 【0612】

50

実施形態109:これらのオペレーションは、ユーザーインタラクションにตอบสนองして構造化プレゼンテーションの第1のセル内の値の表示を更新する段階をさらに含む、実施形態102のシステム。

【0613】

実施形態110:事前検索を特徴付ける情報を表示する段階は、電子文書コレクションの第1の文書内の第1の値のコンテキストを特徴付けるスニペットを表示する段階を含む、実施形態102のシステム。

【0614】

実施形態111:電子文書のコレクションは、インターネット上で利用可能な電子文書を含み、電子文書は、ウェブページを含む、実施形態110のシステム。

10

【0615】

実施形態112:構造化プレゼンテーションは、カードのコレクションを含む、実施形態102のシステム。

【0616】

実施形態113:ディスプレイデバイスを備えるクライアントデバイスをインタラクティブに操作し、オペレーションを実行するようにプログラムされた1つまたは複数のコンピュータであって、これらのオペレーションは、構造化プレゼンテーションをディスプレイデバイス上に表示する段階であって、構造化プレゼンテーションは、構造化設計に適合する組織的な構造化された配列で情報を視覚的に提示し、構造化プレゼンテーションは、構造化プレゼンテーションの視覚的提示でインスタンスとインスタンスの属性を特徴付ける値との間の関連付けをインスタンスの識別子と値の配列によって表す、段階と、表示される構造化プレゼンテーションに対するユーザーインタラクションを特徴付ける、構造化プレゼンテーションの第1のインスタンスおよび第1の属性の指定を含むデータを受け取る段階と、データを受け取ったことにตอบสนองしてディスプレイデバイス上に以前には隠されていた検索インターフェイスを表示する段階であって、検索インターフェイスは、電子文書コレクション内の第1のインスタンスの第1の属性を特徴付ける第1の値の位置を識別する情報もしくはインタラクティブ要素を含む、段階とを含む、1つまたは複数のコンピュータを備える、システム。

20

【0617】

発明対象および本明細書で説明されている機能するオペレーションのこれらの実施形態および他の実施形態は、デジタル電子回路で、または本明細書で開示されている構造およびその構造的等価物を含むコンピュータソフトウェア、ファームウェア、もしくはハードウェアで、またはこれらのうちの1つまたは複数のものの組み合わせで実装することができる。本明細書で説明されている発明対象の実施形態は、1つまたは複数のコンピュータプログラム、つまり、データ処理装置による実行のため、またはデータ処理装置のオペレーションを制御するためにコンピュータ記憶媒体上に符号化されたコンピュータプログラム命令からなる1つまたは複数のモジュールとして実装されうる。代替えとして、またはそれに加えて、プログラム命令は、データ処理装置による実行のため好適な受信機装置に伝送する情報を符号化するように生成される、人工的に生成された信号である伝搬信号、例えば、機械で生成された電気、光、または電磁信号に符号化できる。コンピュータ記憶媒体は、機械可読記憶デバイス、機械可読記憶基板、ランダムもしくはシリアルアクセスメモリデバイス、またはこれらのうちの1つもしくは複数のものの組み合わせとすることができる。

30

40

【0618】

「データ処理装置」という用語は、例えば、プログラム可能なプロセッサ、コンピュータ、または複数のプロセッサもしくはコンピュータを含む、データを処理するためのあらゆる種類の装置、デバイス、および機械を包含する。装置は、専用論理回路、例えば、FPGA(フィールドプログラマブルゲートアレイ)、またはASIC(特定用途向け集積回路)を含んでいてもよい。装置は、ハードウェアに加えて、注目しているコンピュータプログラム用の実行環境を作成するコード、例えば、プロセッサファームウェア、プロトコルスタック

50

、データベース管理システム、オペレーティングシステム、またはこれらのうちの1つもしくは複数のものの組み合わせを構成するコードを含むものとしてもよい。

【0619】

コンピュータプログラム(プログラム、ソフトウェア、ソフトウェアアプリケーション、スクリプト、またはコードとも呼ばれる)は、コンパイル言語もしくはインタプリタ言語、または宣言型言語もしくは手続き型言語を含む、任意の形態のプログラミング言語で書かれ、スタンドアロンプログラム、またはモジュール、コンポーネント、サブルーチン、またはコンピューティング環境において使用するのに適している他のユニットを含む、任意の形態で配備されうる。コンピュータプログラムは、ファイルシステム内のファイルに対応するものとしてよいが、そうである必要はない。プログラムは、他のプログラムもしくはデータ(例えば、マークアップ言語文書内に格納される1つまたは複数のスクリプト)を保持するファイルの一部に、注目するプログラム専用の単一ファイル内に、または複数の調整されたファイル(例えば、1つまたは複数のモジュール、サブプログラム、またはコードの一部を格納するファイル)に格納されうる。コンピュータプログラムは、1つのコンピュータ上で、または1つのサイトに配置されるか、または複数のサイトにまたがって分散され、通信ネットワークによって相互接続されている複数のコンピュータ上で実行されるように配備されうる。

10

【0620】

本明細書で説明されているプロセスおよび論理の流れは、入力データを操作し、出力を生成することによって機能を実行するように1つまたは複数のコンピュータプログラムを実行する1つまたは複数のプログラム可能なプロセッサによって実行されうる。プロセスおよび論理の流れも、専用論理回路、例えば、FPGA(フィールドプログラマブルゲートアレイ)、またはASIC(特定用途向け集積回路)によって実行され、また装置も、専用論理回路、例えば、FPGA(フィールドプログラマブルゲートアレイ)、またはASIC(特定用途向け集積回路)によって実装されうる。

20

【0621】

コンピュータプログラムの実行に適しているプロセッサとしては、例えば、汎用マイクロプロセッサ、専用マイクロプロセッサ、および任意の種類デジタルコンピュータの任意の1つまたは複数のプロセッサが挙げられる。一般に、プロセッサは、読み取り専用メモリまたはランダムアクセスメモリまたはその両方から命令およびデータを受け取る。コンピュータの不可欠な要素は、命令を遂行または実行するためのプロセッサならびに命令およびデータを格納するための1つまたは複数のメモリデバイスである。一般に、コンピュータは、データを格納するための1つまたは複数の大容量記憶装置デバイス、例えば、磁気ディスク、光磁気ディスク、または光ディスクも備え、これらからデータを受け取るか、またはこれらにデータを転送するか、またはその両方を行うように動作可能なように結合される。しかし、コンピュータはこのようなデバイスを備える必要はない。さらに、コンピュータは、他のデバイス、例えば、2、3例を挙げると、携帯電話、パーソナルデジタルアシスタント(PDA)、携帯オーディオまたはビデオプレーヤー、ゲーム機、全地球測位システム(GPS)受信機、またはポータブル記憶デバイス(例えば、ユニバーサルシリアルバス(USB)フラッシュドライブ)に埋め込むことができる。

30

40

【0622】

コンピュータプログラムの命令およびデータを格納するのに適したコンピュータ可読媒体は、例えば、半導体メモリデバイス、例えば、EPROM、EEPROM、およびフラッシュメモリデバイス、磁気ディスク、例えば、内蔵ハードディスクまたはリムーバブルディスク、光磁気ディスク、ならびにCD-ROMおよびDVD-ROMディスクを含む、あらゆる形態の不揮発性メモリ、媒体、およびメモリデバイスを含む。プロセッサおよびメモリは、専用論理回路で補完されるか、または専用論理回路に組み込まれうる。

【0623】

ユーザーと情報のやり取りを行うために、本明細書で説明されている発明対象の実施形態は、ユーザーに情報を表示するためのディスプレイデバイス、例えば、CRT(陰極線管)

50

またはLCD(液晶ディスプレイ)モニタ、ならびにユーザーがコンピュータに入力を送るために使用できるキーボードおよびポインティングデバイス、例えば、マウスもしくはトラックボール、を有するコンピュータ上で実装することができる。他の種類のデバイスも、ユーザーと情報をやり取りするために使用することができ、例えば、ユーザーに返されるフィードバックは、任意の形態の感覚フィードバック、例えば、視覚フィードバック、聴覚フィードバック、または触覚フィードバック、とすることができ、ユーザーからの入力は、音響、話し声、または触覚入力を含む、任意の形態で受け取ることができる。それに加えて、コンピュータは、ユーザーが使用するデバイスに文書を送信し、そのデバイスから文書を受け取ることによって、例えば、ウェブページをユーザーのクライアントデバイス上のウェブブラウザに、ウェブブラウザから受け取った要求に回答して送信することによって、ユーザーと情報をやり取りすることができる。

10

## 【0624】

本明細書は、多くの実装固有の詳細事項を含んでいるが、これらは、任意の発明の範囲または請求内容の範囲に対する制限として解釈すべきではなく、むしろ特定の発明の特定の実施形態に特有のものであると思われる特徴の説明として解釈すべきである。別々の実施形態の背景状況において本明細書で説明されているいくつかの特徴も、単一の実施形態において組み合わせて実装されうる。逆に、単一の実施形態の背景状況において説明されているさまざまな特徴は、複数の実施形態で別々に、または好適な部分的組み合わせで、実装することも可能である。さらに、上ではいくつかの特徴を特定の組み合わせで動作するものとして説明することができ、さらには最初にそのようなものとして請求されうるが、請求されている組み合わせから得られる1つまたは複数の特徴は、場合によっては、組み合わせから切り取られ、また請求された組み合わせは、部分的組み合わせまたは部分的組み合わせの変更形態を対象とするものとすることもできる。

20

## 【0625】

同様に、オペレーションは図面内に特定の順序で示されているが、これは、そのようなオペレーションが図示されている特定の順序で、または順番に実行されること、または例示されているオペレーションが、望ましい結果が得られるように実行されることを要求していると理解されるべきではない。さらに、上述の実施形態においてさまざまなシステムコンポーネントが分離しているが、すべての実施形態においてそのような分離が必要とされていると理解されるべきではなく、また説明されているプログラムコンポーネントおよびシステムは、一般的に、単一のソフトウェア製品に一体化されうるか、または複数のソフトウェア製品にパッケージングされうることは理解されるであろう。

30

## 【0626】

本明細書で説明されている発明対象の特定の実施形態が説明された。他の実施形態も、以下の請求項の範囲内に収まる。例えば、請求項に記載のアクションは、異なる順序で実行することができ、それでも、所望の結果が得られる。一例として、付属の図面に示されているプロセスは、所望の結果を得るために、図示されている特定の順序、または順番を必ずしも必要としない。

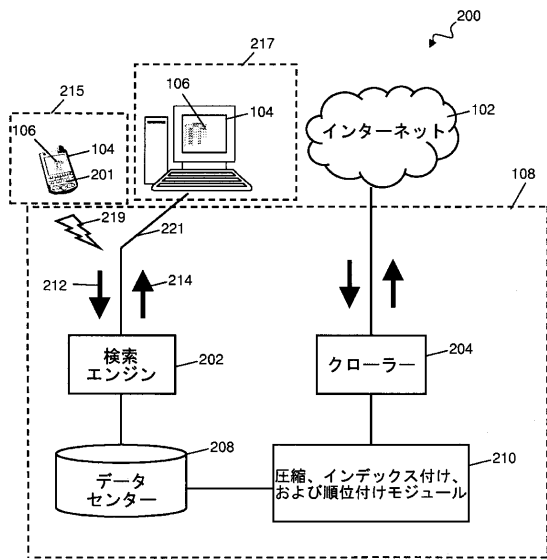
## 【符号の説明】

## 【0627】

- 100 システム
- 102 非構造化電子文書コレクション
- 104 ディスプレイ画面
- 106 構造化プレゼンテーション
- 108 データ通信インフラストラクチャ
- 110、112、114 電子文書

40

【図2】



【図3】

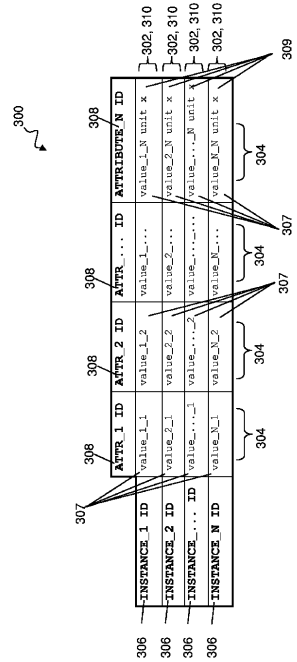


FIG. 3

【図4】

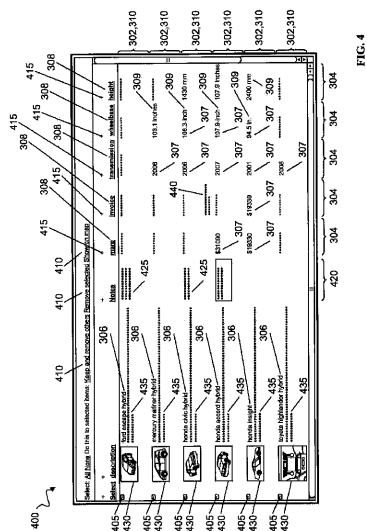


FIG. 4

【図5】

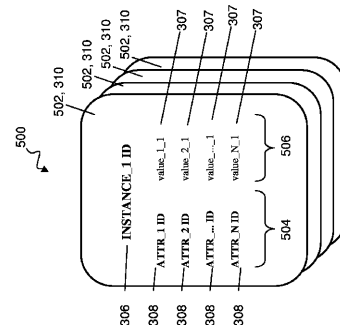
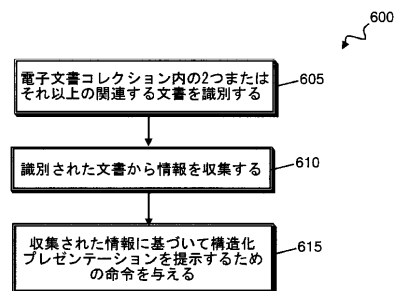
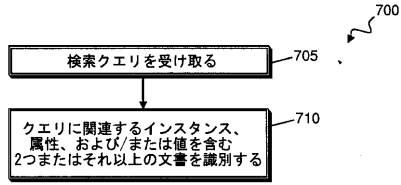


FIG. 5

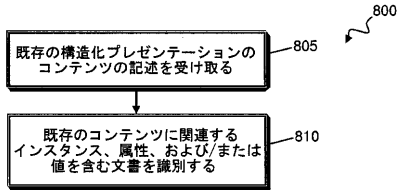
【図6】



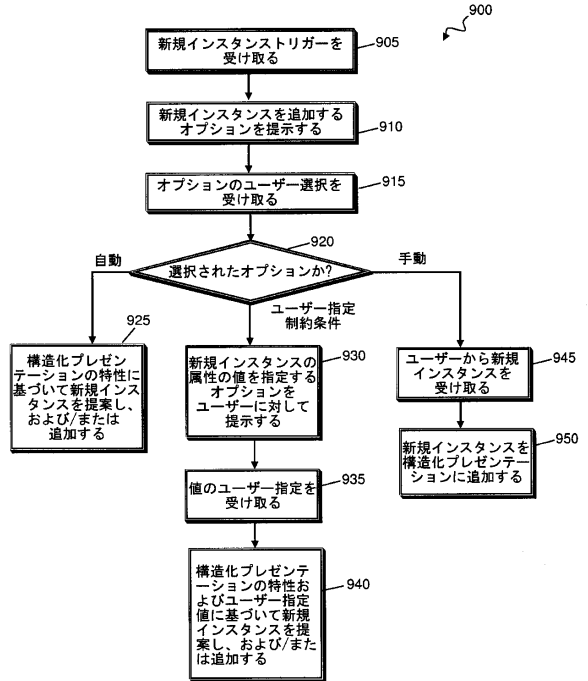
【図7】



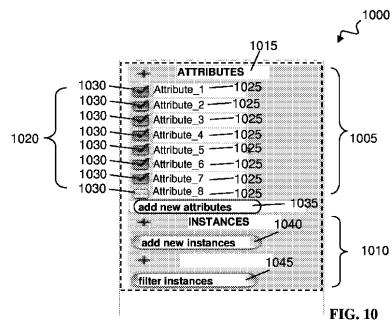
【図8】



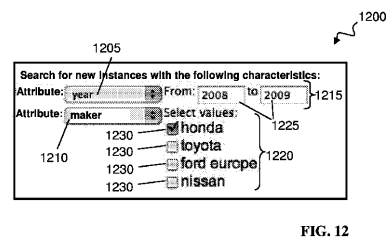
【図9】



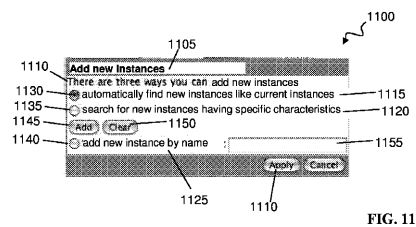
【図10】



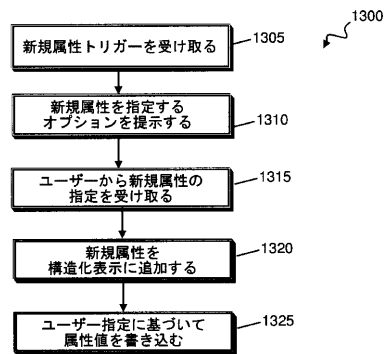
【図12】



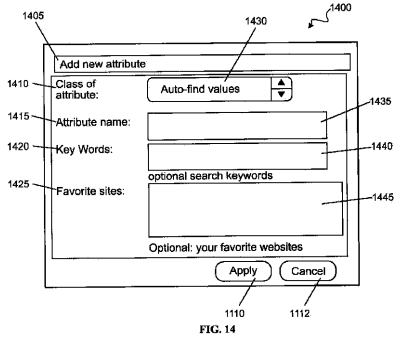
【図11】



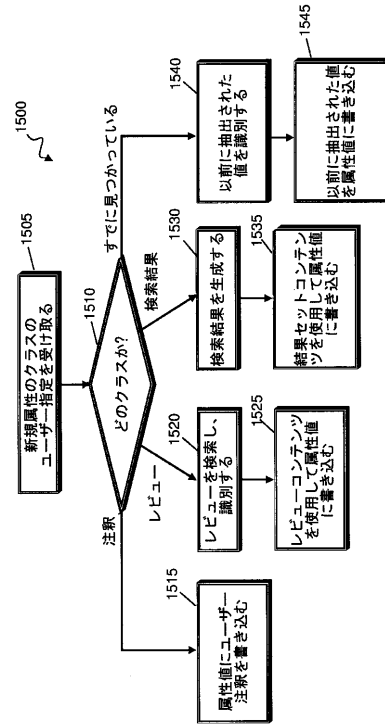
【図13】



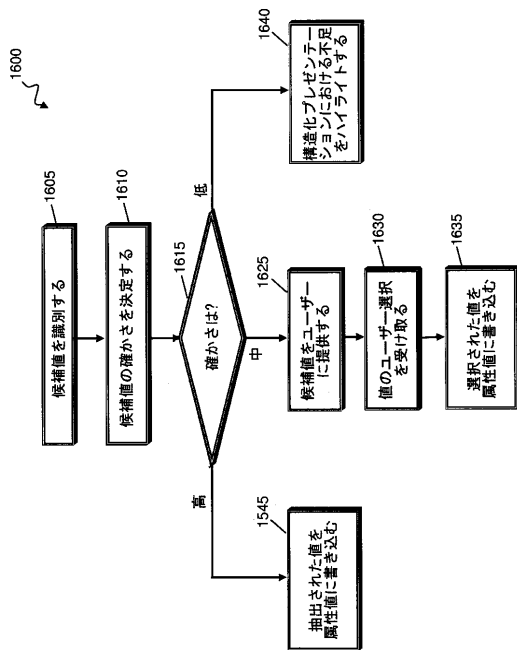
【 図 1 4 】



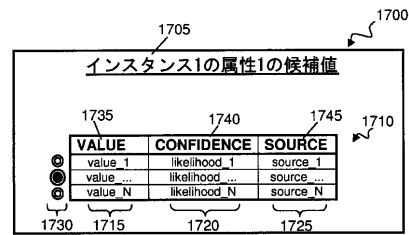
【 図 1 5 】



【 図 1 6 】



【 図 1 7 】



【図18】

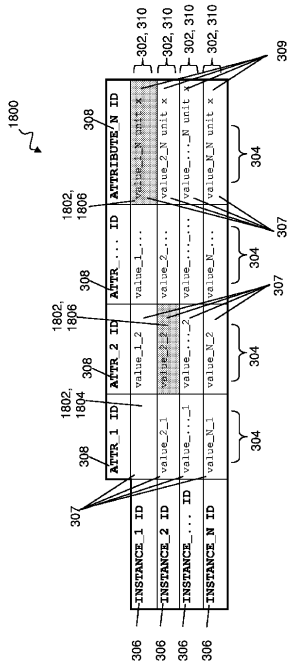
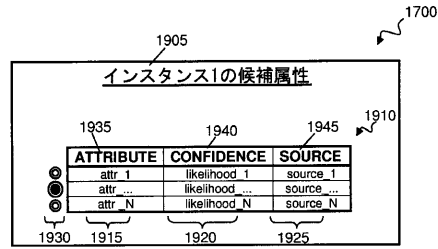
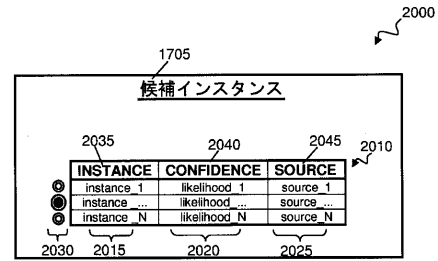


FIG. 18

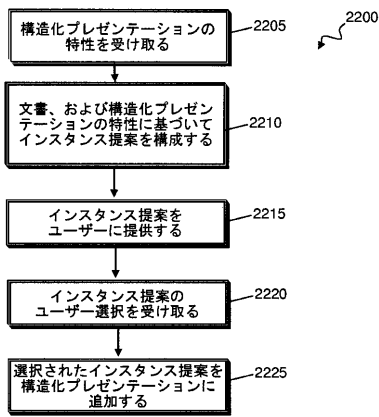
【図19】



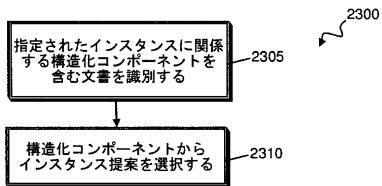
【図20】



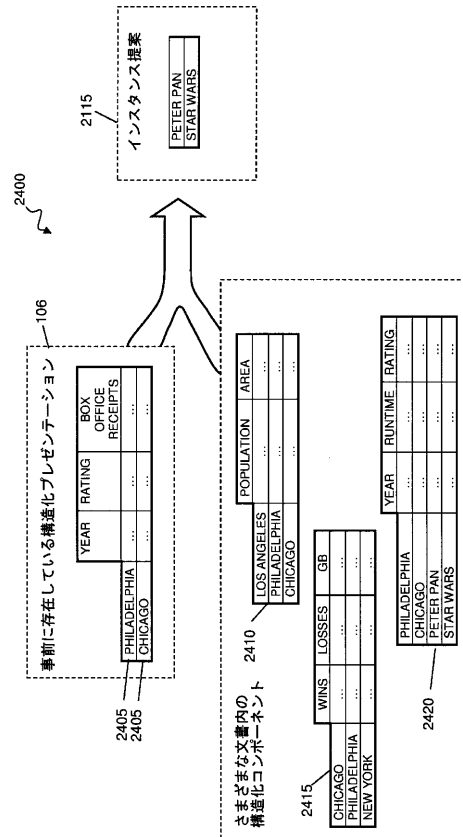
【図22】



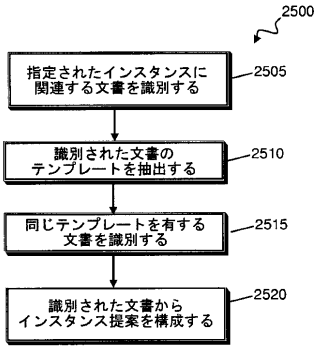
【図23】



【図24】



【図 25】



【図 26】

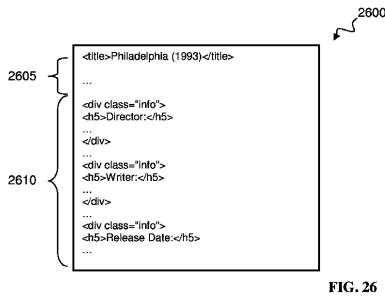
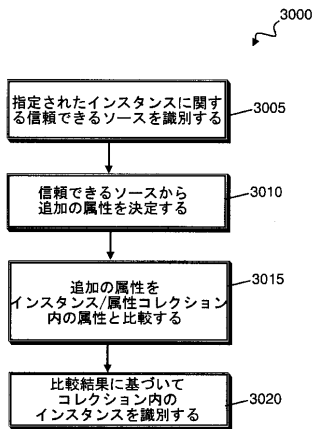
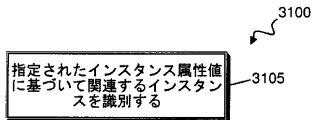


FIG. 26

【図 30】



【図 31】

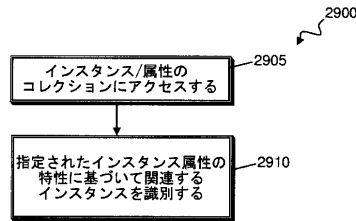


【図 28】

2802	INSTANCE 1	LOCATION 2	ATTR 2	ATTR 3	ATTR 5	
2804	INSTANCE 2	LOCATION 3	ATTR 5	ATTR 6		
2806	INSTANCE 2	LOCATION 3	ATTR 3	ATTR 5	ATTR ...	ATTR_N
2810	INSTANCE 2	LOCATION 4	ATTR 3	ATTR 4		
2812	INSTANCE 2	LOCATION 5	ATTR 5	ATTR 8	ATTR 9	
2814	INSTANCE ...	LOCATION ...	ATTR ...	ATTR ...	ATTR ...	ATTR_N

FIG. 28

【図 29】

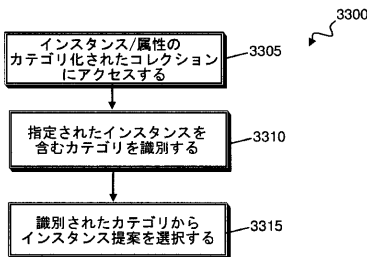


【図 32】

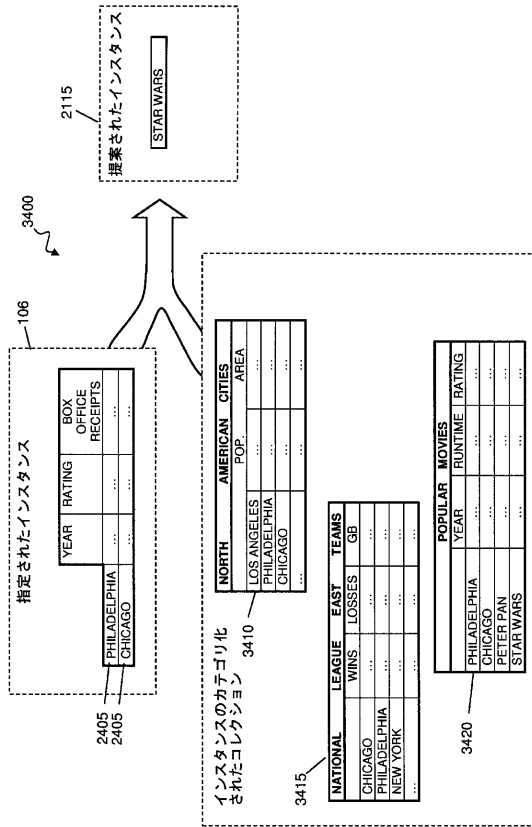
3202	INSTANCE 10	LOCATION 22	ATTR 3	VAL_3_unit_b	ATTR 6	VAL_6_unit_c	...
3204	INSTANCE 11	LOCATION 25	ATTR 3	VAL_8_unit_a	ATTR 6	VAL_6_unit_c	...
3206	INSTANCE 12	LOCATION 32	ATTR 3	VAL_3_unit_a	ATTR 6	VAL_6_unit_f	...
3210	INSTANCE 13	LOCATION 22	ATTR 3	VAL_3_unit_a	ATTR 6	VAL_8_unit_g	...
3212	INSTANCE 14	LOCATION 33	ATTR 3	VAL_4_unit_b	ATTR 6	VAL_..._unit_f	...
3214	INSTANCE ...	LOCATION ...	ATTR 3	VAL ...	ATTR 6	VAL ...	...

FIG. 32

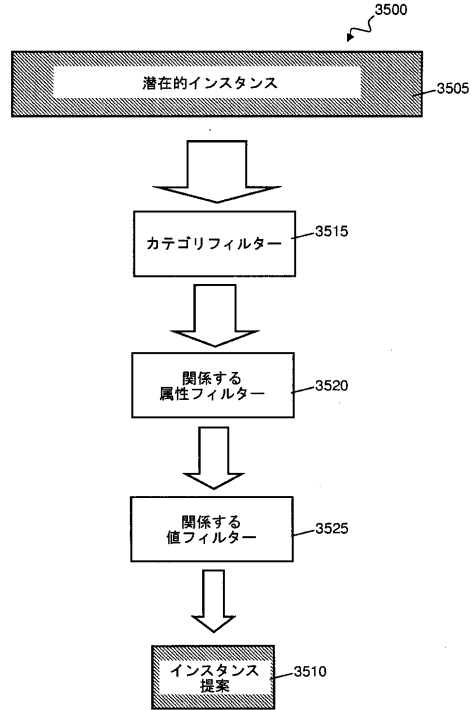
【図 33】



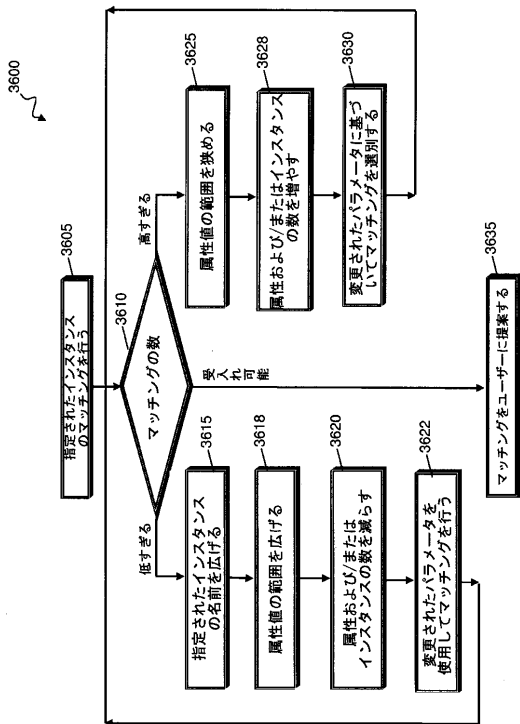
【図34】



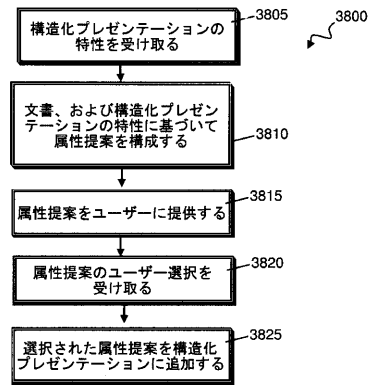
【図35】



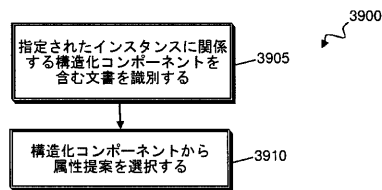
【図36】



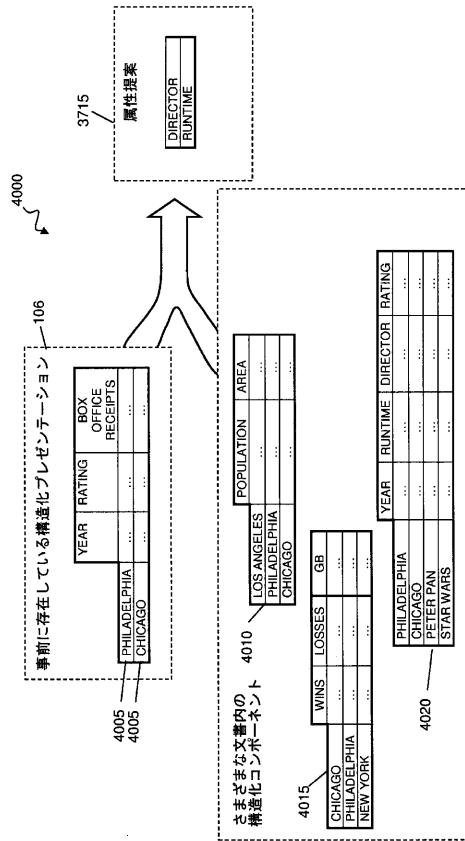
【図38】



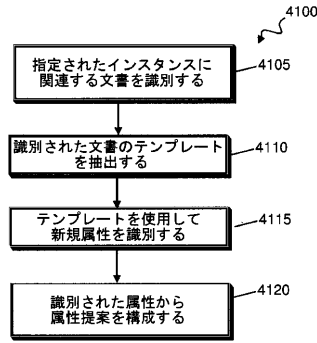
【図39】



【図40】



【図41】



【図42】



FIG. 42

【図44】

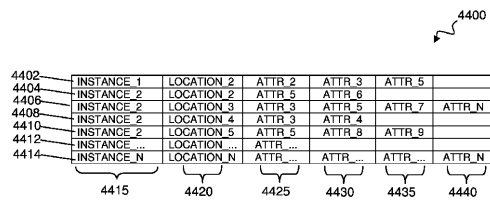
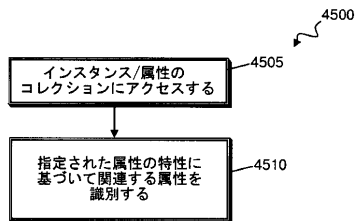
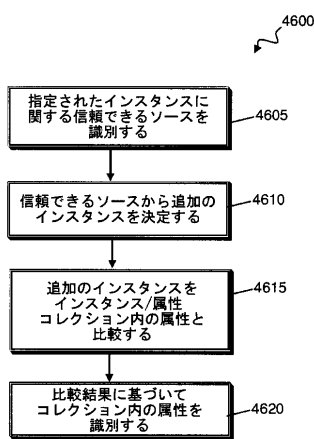


FIG. 44

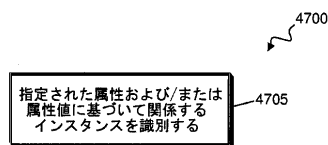
【図45】



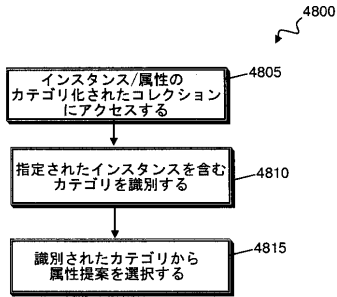
【図46】



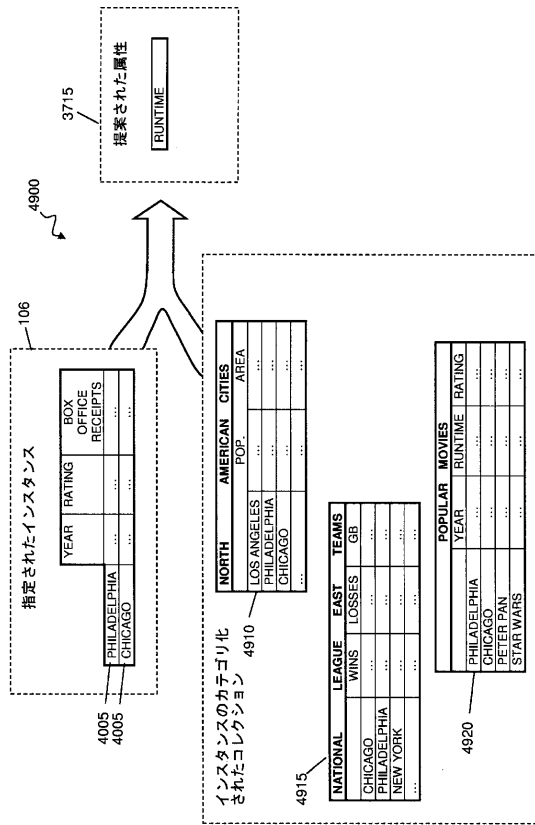
【図47】



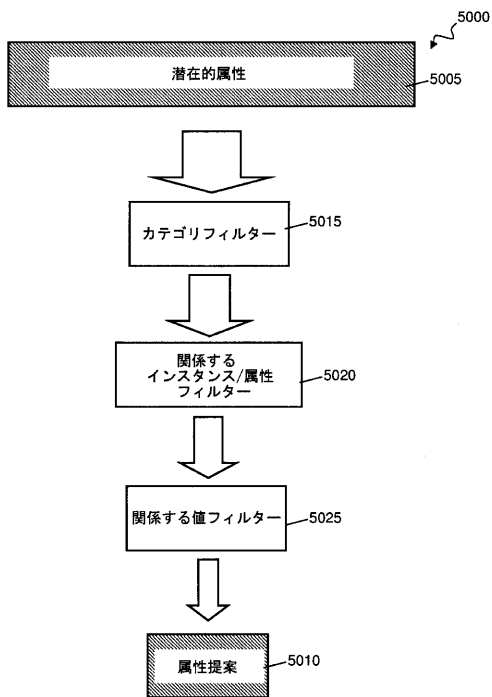
【図48】



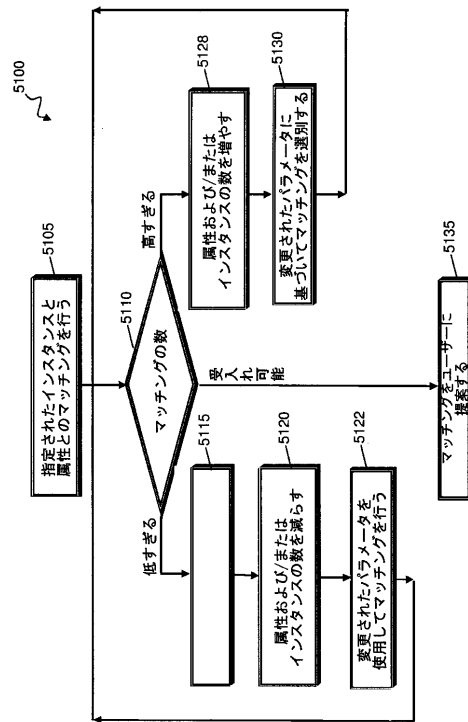
【図49】



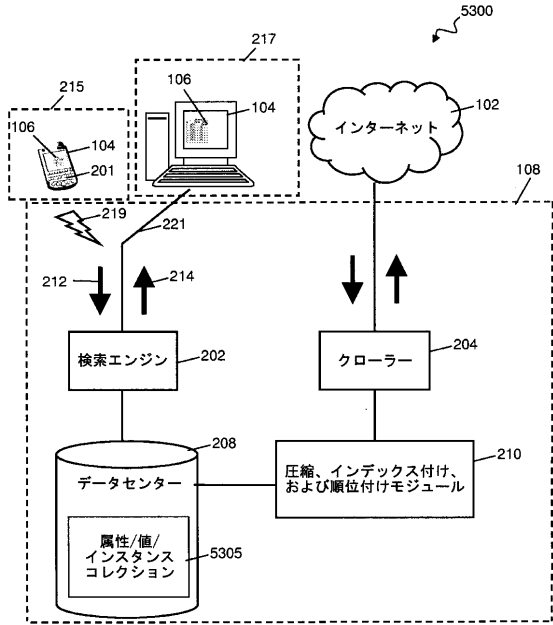
【図50】



【図51】



【図53】

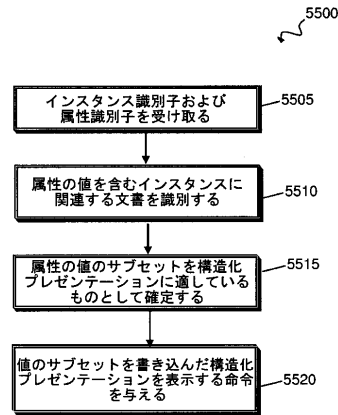


【図54】

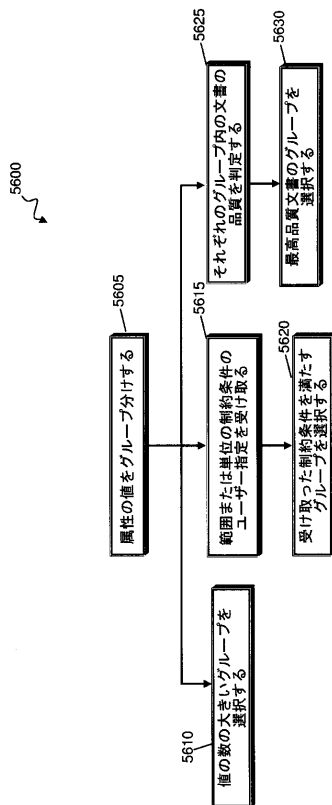
5402	INSTANCE 1	DOC 1	ATTR 1	VALUE 1	ATTR 3	VALUE 3A	ATTR 4	VALUE 4A
5404	INSTANCE 1	DOC 2	ATTR 5	VALUE 5	ATTR 6	VALUE 6	ATTR 7	VALUE 7
5406	INSTANCE 1	DOC 3	ATTR 3	VALUE 3B	ATTR 4	VALUE 4A		
5408	INSTANCE ...	DOC ...	ATTR ...	VALUE ...	ATTR ...	VALUE ...	ATTR N	VALUE N
5410	INSTANCE N	DOC N	ATTR ...	VALUE ...	ATTR ...	VALUE ...	ATTR N	VALUE N

FIG. 54

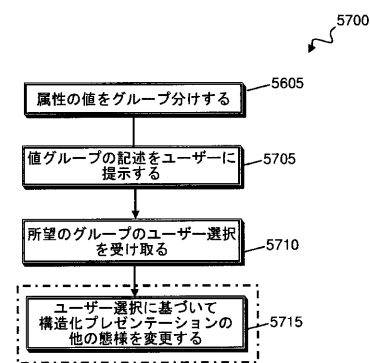
【図55】



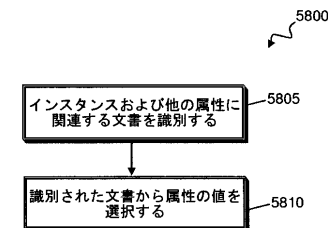
【図56】



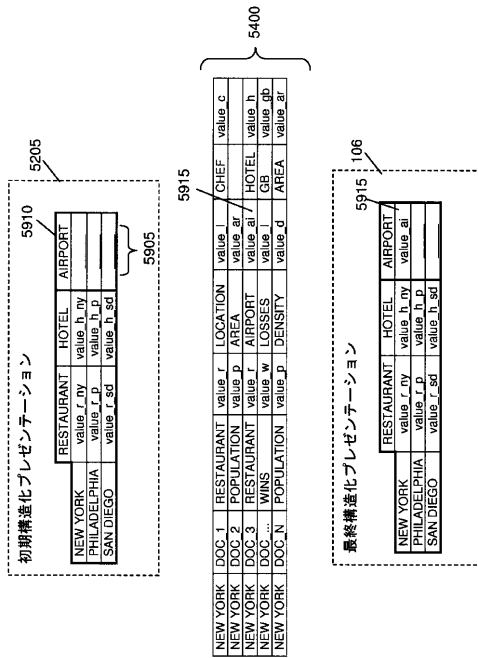
【図57】



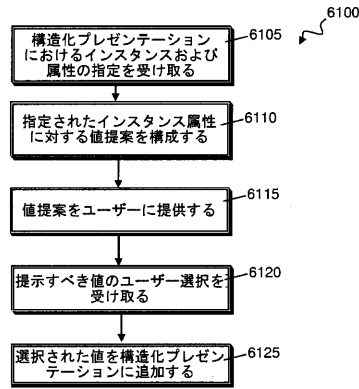
【図58】



【図59】



【図61】



【図62】

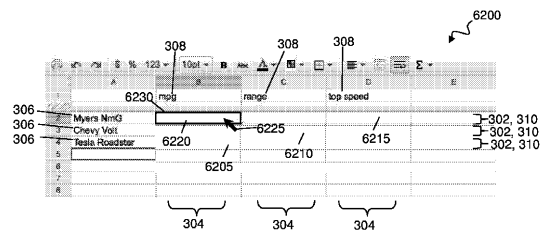


FIG. 62

【図63】

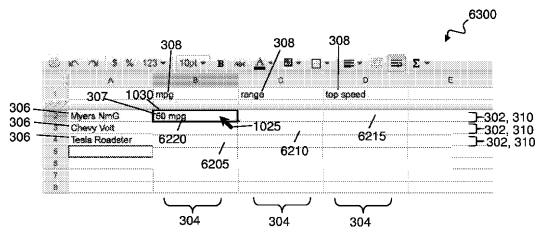


FIG. 63

【図65】

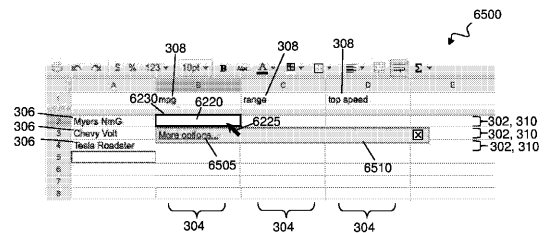


FIG. 65

【図64】

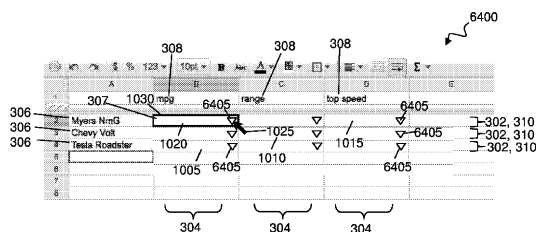


FIG. 64

【図66】

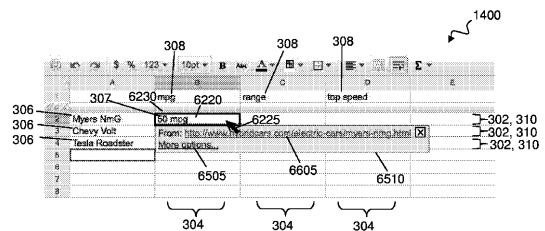
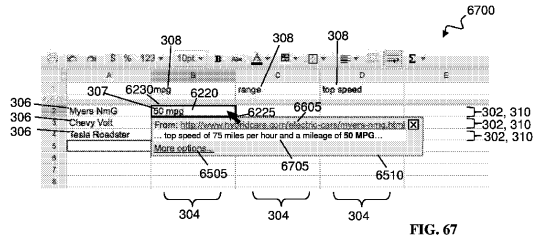
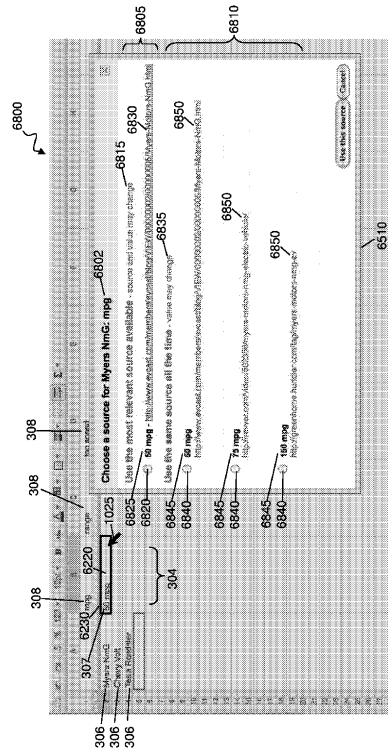


FIG. 66

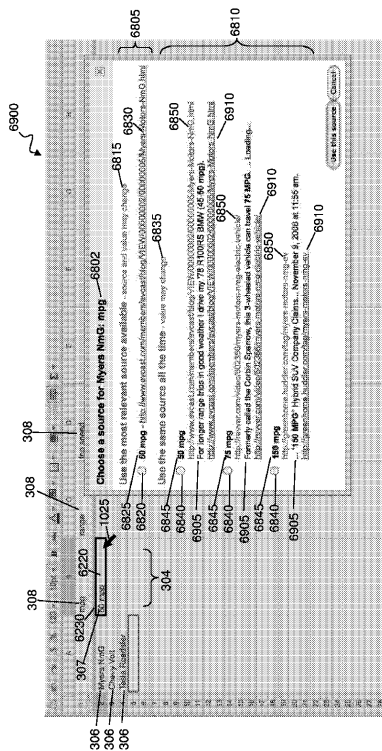
【 67 】



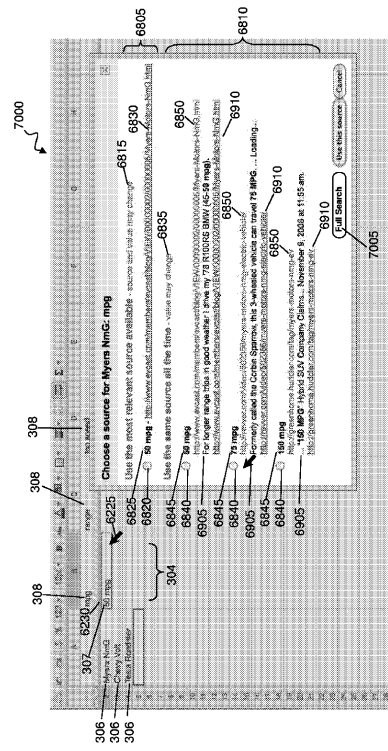
【 68 】



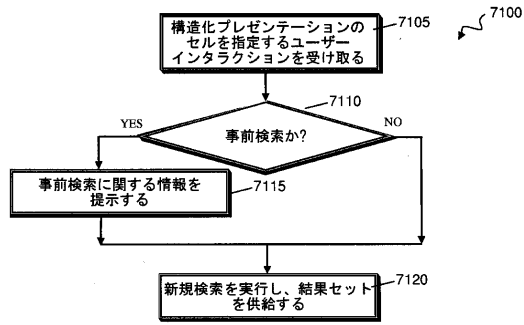
【 69 】



【 70 】



【図71】



【図72】

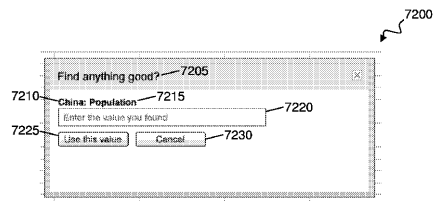


FIG. 72

【図73】

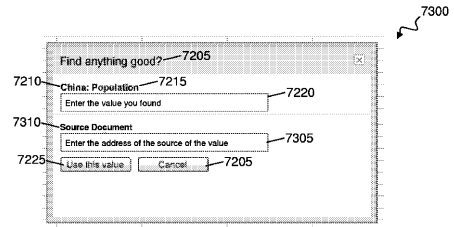


FIG. 73

【図74】

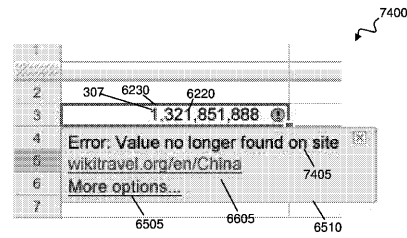
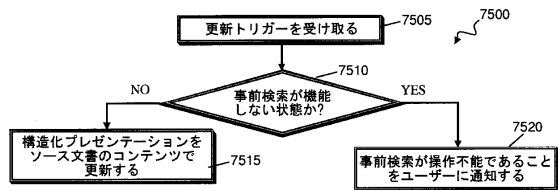
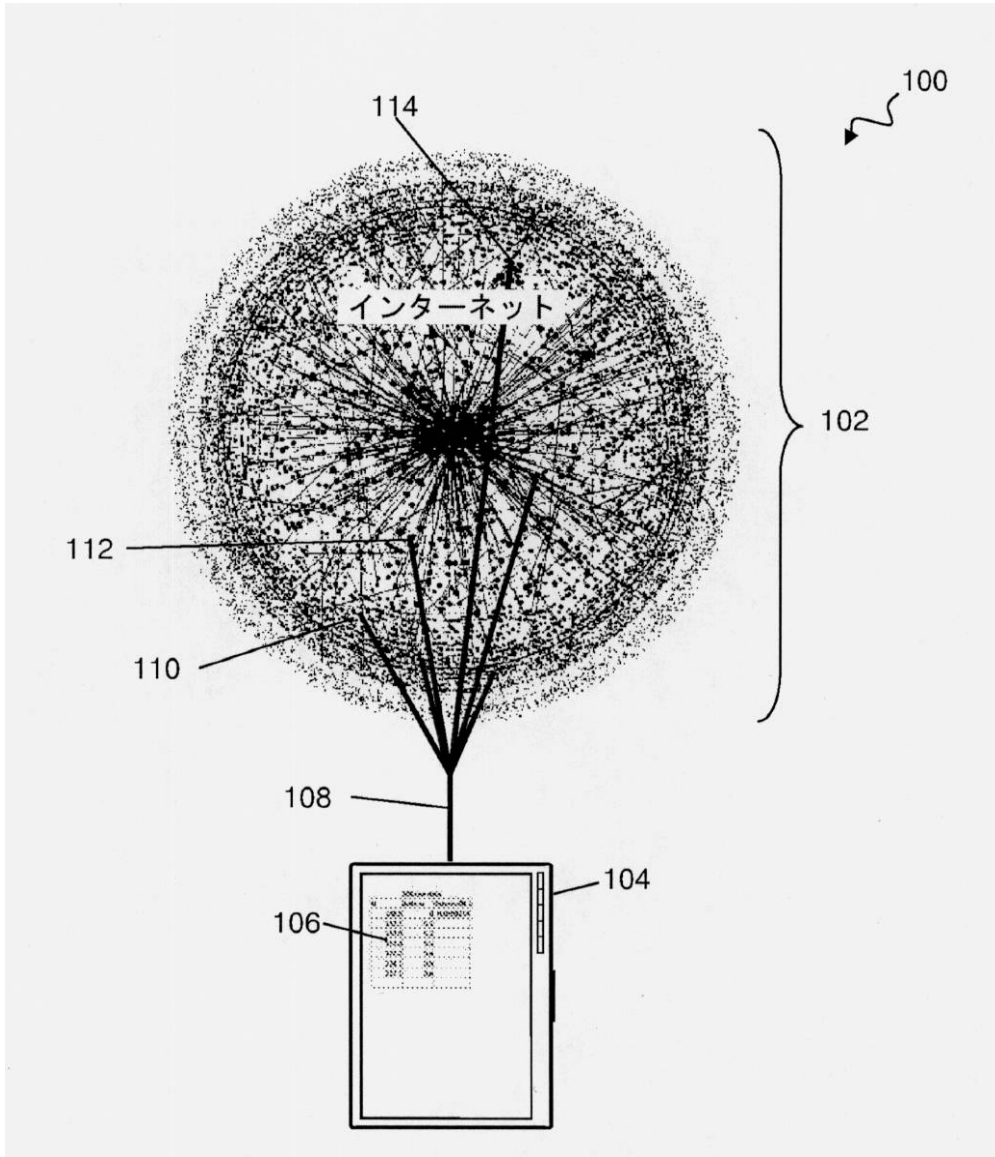


FIG. 74

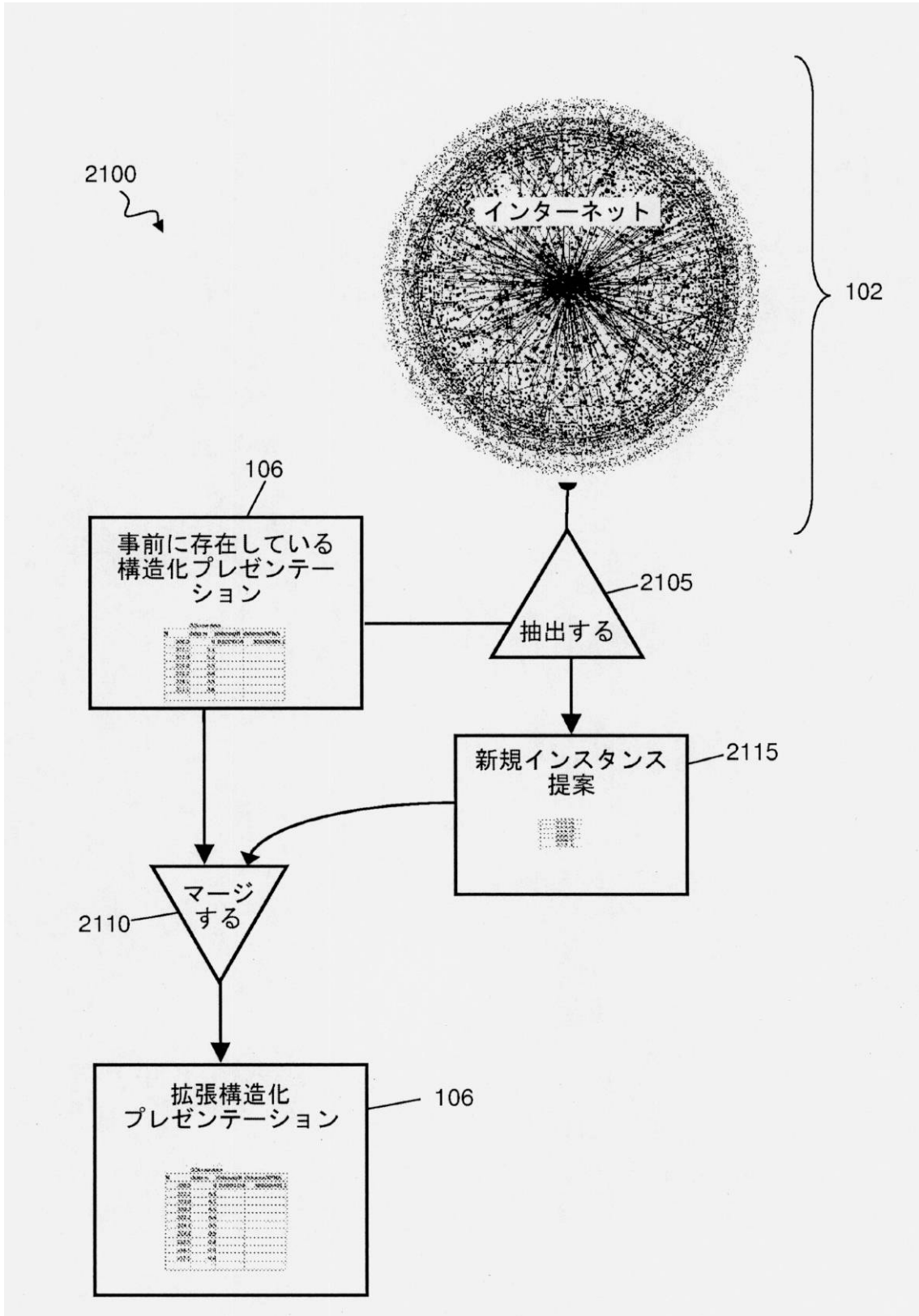
【図75】



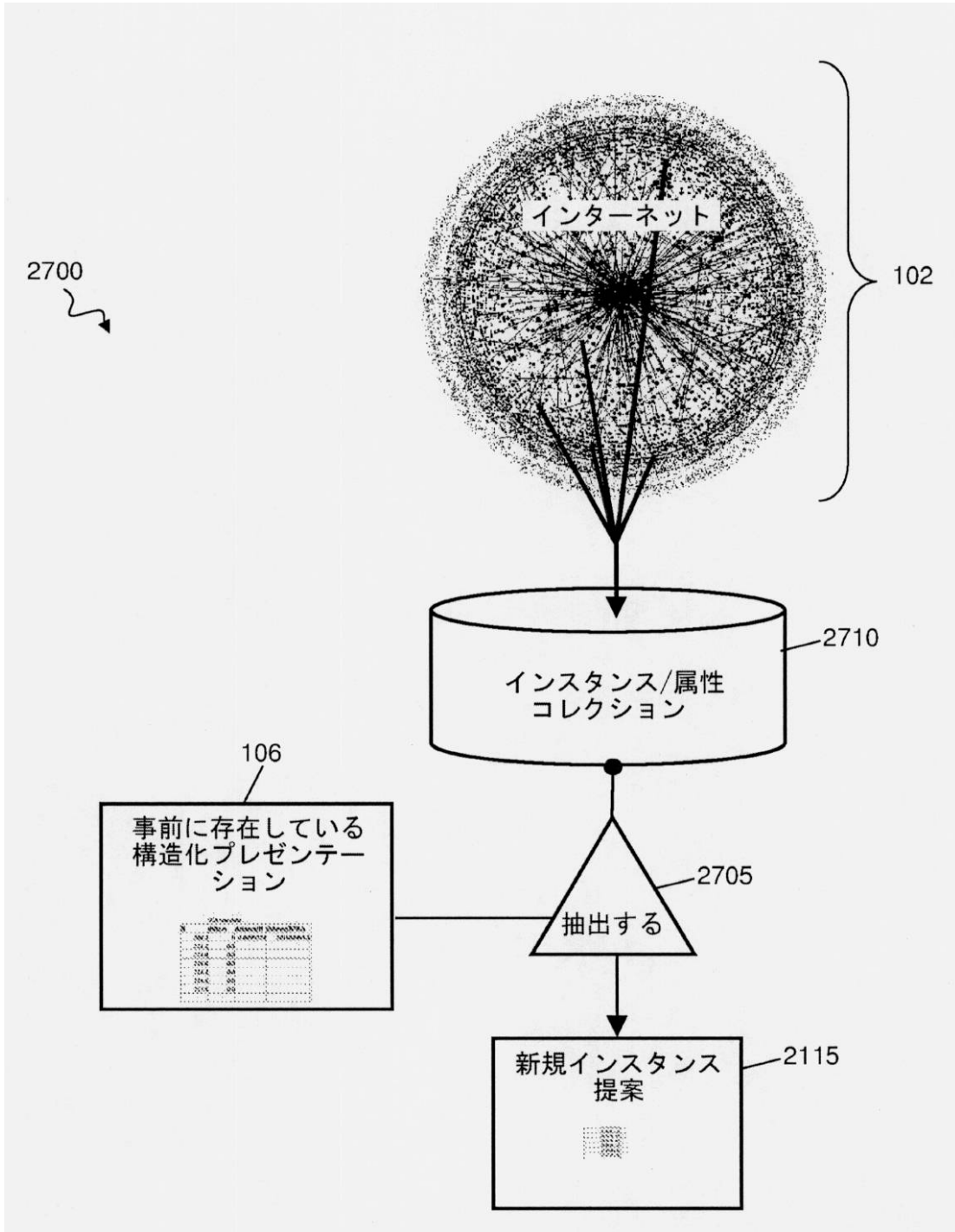
【図1】



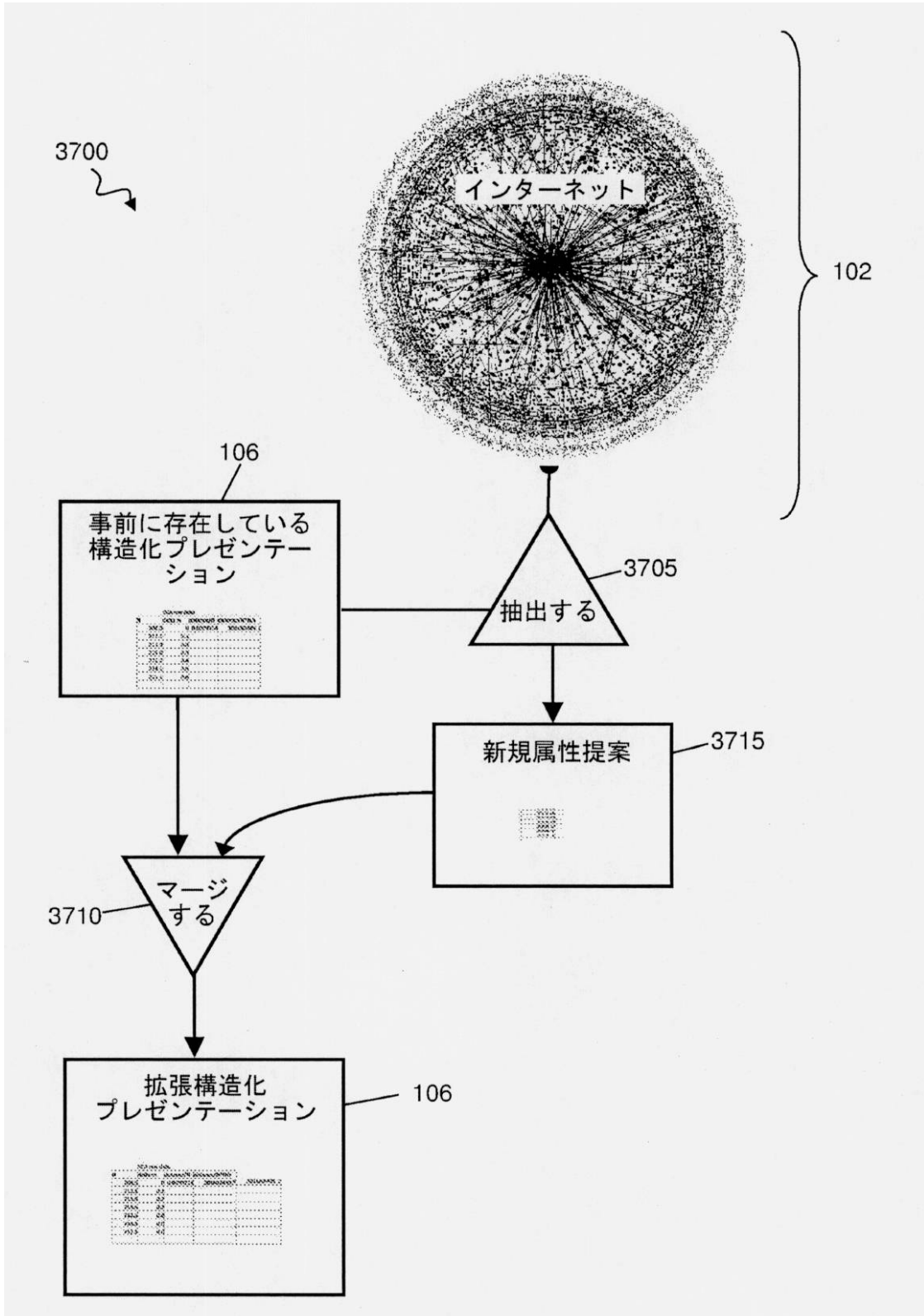
【図 21】



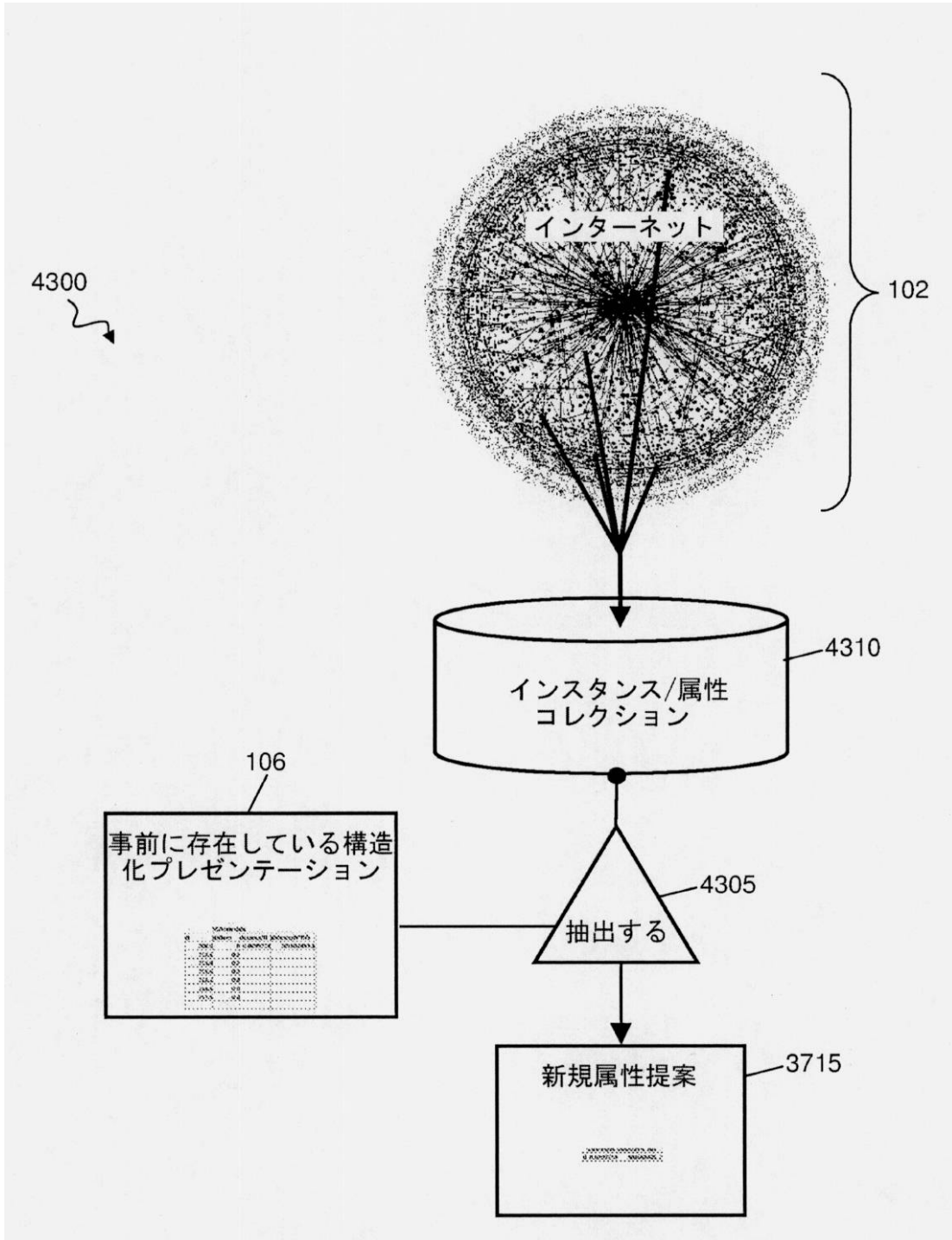
【図 27】



【図 37】

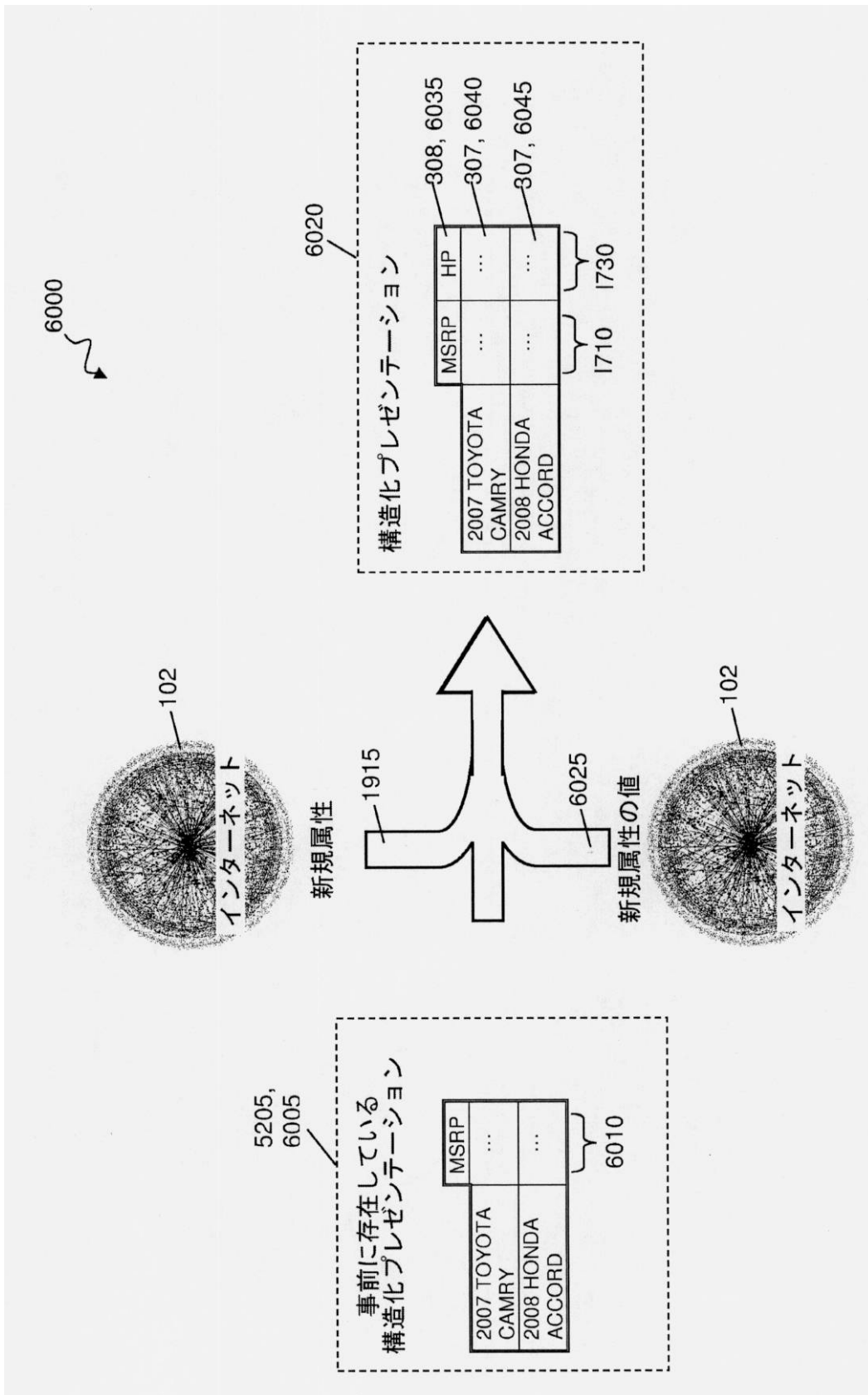


【図43】





【 図 6 0 】



## フロントページの続き

- (31)優先権主張番号 12/355,228  
 (32)優先日 平成21年1月16日(2009.1.16)  
 (33)優先権主張国 米国(US)
- (31)優先権主張番号 12/355,554  
 (32)優先日 平成21年1月16日(2009.1.16)  
 (33)優先権主張国 米国(US)
- (31)優先権主張番号 12/355,607  
 (32)優先日 平成21年1月16日(2009.1.16)  
 (33)優先権主張国 米国(US)
- (72)発明者 ダニエル・エヌ・クアイン  
 アメリカ合衆国・ニューヨーク・10036・ニュー・ヨーク・ウエスト・フォーティエイス・ストリート・235・#29B
- (72)発明者 ダニエル・ロレット  
 アメリカ合衆国・ニューヨーク・10016-7427・ニュー・ヨーク・イースト・トゥエンティナインス・ストリート・10・アパートメント・3F
- (72)発明者 ボグダン・カプリタ  
 アメリカ合衆国・ニューヨーク・10027・ニュー・ヨーク・ウエスト・ワンハンドレッドトゥエンティエス・ストリート・414・アパートメント・504
- (72)発明者 アントネッラ・パヴィーズ  
 アメリカ合衆国・ペンシルヴェニア・19403・ノリスタウン・ウィリアムズ・ウェイ・1707
- (72)発明者 ジェフリー・シー・レイナー  
 アメリカ合衆国・ニューヨーク・10009・ニュー・ヨーク・イースト・セヴンス・ストリート・128・#5
- (72)発明者 アンドリュー・ウィリアム・ホーグ  
 アメリカ合衆国・ニュージャージー・07423・ホー・ホー・カス・エルムウッド・アヴェニュー・45
- (72)発明者 アンソニー・ジェイ・アイウト  
 アメリカ合衆国・ニューヨーク・11024・グレート・ネック・アップルトゥリー・レーン・5
- (72)発明者 ジョン・アレクサンダー・コモロスケ  
 アメリカ合衆国・ニューヨーク・10012・ニュー・ヨーク・ウエスト・フォース・ストリート・140・アパートメント・16

審査官 久々宇 篤志

- (56)参考文献 西野 文人 他, 新聞記事からの人物・企業情報の抽出, 情報処理学会研究報告 Vol. 98 No. 81, 日本, 社団法人情報処理学会, 1998年 9月18日, pp.125-132  
 木村 壘 他, 人物情報辞典の自動生成のための時系列情報自動収集, 電子情報通信学会技術研究報告 Vol. 106 No. 149 DE2006-50~91 データ工学, 日本, 社団法人電子情報通信学会, 2006年 7月 6日, pp.41-46  
 山本 あゆみ 他, ワールドワイドウェブからの人物情報の自動収集, 電子情報通信学会技術研究報告 Vol. 99 No. 534 AI99-77~89 人工知能と知識処理, 日本, 社団法人電子情報通信学会, 2000年 1月13日, pp.93-100

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