

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7387432号
(P7387432)

(45)発行日 令和5年11月28日(2023.11.28)

(24)登録日 令和5年11月17日(2023.11.17)

(51)国際特許分類 F I
G 0 6 Q 30/0601(2023.01) G 0 6 Q 30/0601

請求項の数 27 (全23頁)

(21)出願番号	特願2019-535764(P2019-535764)	(73)特許権者	522008263
(86)(22)出願日	平成29年12月28日(2017.12.28)		オブセック・オンライン・リミテッド
(65)公表番号	特表2020-504879(P2020-504879 A)		イギリス・NE38・0AD・ワシントン・フェニックス・ロード・40
(43)公表日	令和2年2月13日(2020.2.13)	(74)代理人	100108453
(86)国際出願番号	PCT/US2017/068675		弁理士 村山 靖彦
(87)国際公開番号	WO2018/125984	(74)代理人	100110364
(87)国際公開日	平成30年7月5日(2018.7.5)		弁理士 実広 信哉
審査請求日	令和2年11月24日(2020.11.24)	(74)代理人	100133400
審判番号	不服2022-19021(P2022-19021/J 1)		弁理士 阿部 達彦
審判請求日	令和4年11月25日(2022.11.25)	(72)発明者	メアリー・ヴィ・ジェンキンズ
(31)優先権主張番号	62/440,798		アメリカ合衆国・アイダホ・83687
(32)優先日	平成28年12月30日(2016.12.30)		・ナンパ・ノース・ディーンミア・ループ・16927
(33)優先権主張国・地域又は機関		合議体	
	最終頁に続く		最終頁に続く

(54)【発明の名称】 ネットワーク化環境における不正コンテンツに関連するデータを収集するためのシステムおよび方法

(57)【特許請求の範囲】

【請求項1】

不正コンテンツを識別するために、ネットワーク化されたコンテンツにおける商品識別子を収集し、構文解析し、分析するためのシステムであって、

ネットワーク化環境におけるデータソースに通信可能に接続されたコンピューティングシステムであって、前記データソースが、コンテンツをホストするように構成されている1つまたは複数のリモートサーバを含む、コンピューティングシステムと、

前記コンピューティングシステムにおいて配置されている1つまたは複数のローカルサーバと

を備え、前記1つまたは複数のローカルサーバが、

少なくとも1つの商品識別子に基づいて、前記ネットワーク化環境における前記1つまたは複数のリモートサーバによってホストされた前記コンテンツを収集エンジンを介して検索することであって、前記検索を実行するために前記収集エンジンは、1つまたは複数の検索語から複数のクエリを構築し、前記収集エンジンによって構築された前記複数のクエリの少なくとも1つは、国際取引商品番号(GTIN)をマーケットプレイス固有識別子に変換し、かつウェブサイト内の特定のウェブページをターゲットにするための追加のクエリ構文を含む、検索することと、

前記コンテンツの検索に回答して、検索結果の組を受け取ることであって、各検索結果が、前記コンテンツにおいて識別された商品に関連している、受け取ることと、

前記データソースから検索結果の前記組を収集することと、

検索結果の前記組における各検索結果から複数の商品識別子を抽出することによって、前記複数の商品識別子が、検索結果ごとに、少なくとも1つの抽出されたGTIN、1つの抽出されたブランド名、1つの抽出された在庫管理単位(SKU)、1つの抽出された製品名、および1つの抽出された製品説明を含む、抽出することと、

検索結果の前記組における検索結果ごとに、前記抽出された商品識別子が、合法のブランド名の商品識別子に対応しているか否かに基づいて、前記抽出されたGTINが、合法であるかまたは不正であるかを分析することによって、前記抽出された商品識別子が合法なブランド名の商品識別子に対応するか否かは、前記SKUが正しいか否かに少なくとも基づき、前記抽出された商品識別子が合法なブランド名の商品識別子に対応するか否かは、前記製品名が正しいか否かに少なくとも基づき、かつ、前記抽出された商品識別子が合法なブランド名の商品識別子に対応するか否かは、前記製品説明が正しいか否かに少なくとも基づく、分析することと、

10

前記分析に基づいて、検索結果の前記組における各検索結果に合法または不正とタグ付けすることと

を行うように構成されている、システム。

【請求項2】

前記1つまたは複数のローカルサーバは、

前記抽出されたGTINに含まれているGS1事業者プリフィックスを検索すること、GS1登録事業者情報検索サービスを検索すること、エンティティのデータベースを前記エンティティのアプリケーションプログラミングインターフェース(API)を介して検索すること、またはブランドGTINの独立データベースを検索すること

20

のうちの少なくとも1つによって、前記抽出されたGTINが、前記ブランド名に基づいて、合法であるかまたは不正であるかを分析すること

を行うようにさらに構成されている、請求項1に記載のシステム。

【請求項3】

前記1つまたは複数のローカルサーバが、

前記検索結果のうちの最初の検索結果について、前記複数の商品識別子のうちの対応する商品識別子が、前記検索結果のうちの前記最初の検索結果に含まれている1つまたは複数のあらかじめ定義された商品識別子に対応しているかどうかを決定することと、

30

前記複数の商品識別子のうちの前記対応する商品識別子が、前記1つまたは複数のあらかじめ定義された商品識別子に対応しているかどうかに基づいて、前記検索結果のうちの前記最初の検索結果に合法または不正とタグ付けすることと

を行うようにさらに構成されている、請求項1に記載のシステム。

【請求項4】

前記1つまたは複数のローカルサーバが、

検索結果ごとに前記複数の商品識別子を分析して、合法のブランド名の前記商品識別子に対応していない商品識別子を識別することと、

前記識別された商品識別子の前記検索結果に不正とタグ付けすることと
を行うようにさらに構成されている、請求項1に記載のシステム。

40

【請求項5】

前記1つまたは複数のローカルサーバが、ウェブサイトおよびアプリケーションの直接検索、クエリ構築、ならびに前記ウェブサイトの電子カタログにより、製品情報を収集するようにさらに構成されている、請求項1に記載のシステム。

【請求項6】

前記ネットワーク化環境における前記1つまたは複数のリモートサーバが、ウェブサーバであり、前記1つまたは複数のリモートサーバによってホストされた前記コンテンツが、ウェブページを含むウェブサイトである、請求項1に記載のシステム。

【請求項7】

50

前記1つまたは複数のローカルサーバが、不正とタグ付けされた1つまたは複数の結果に関連している前記コンテンツの削除通知を、前記コンテンツをホストするように構成されている1つまたは複数のリモートサーバに伝達するようにさらに構成されている、請求項1に記載のシステム。

【請求項 8】

前記1つまたは複数のローカルサーバが、

検索結果の前記組における各検索結果からの前記複数の商品識別子の抽出にตอบสนองして、検索結果の前記組について複数のレコードをデータベースに作成することによって、前記データベースに作成された前記複数のレコードの各レコードが、検索結果の前記組における結果に対応する、作成することと、

前記データベースに作成された前記複数のレコードの対応するレコードに各結果から抽出された前記複数の商品識別子を記憶することと
を行うようにさらに構成されている、請求項1に記載のシステム。

【請求項 9】

前記複数のレコードおよび前記複数の商品識別子を表示するように構成されたユーザインターフェースをさらに備える、請求項8に記載のシステム。

【請求項 10】

不正コンテンツを識別するために、ネットワーク化されたコンテンツにおける商品識別子を収集し、構文解析し、分析するための方法であって、前記方法は、ネットワーク化環境におけるデータソースに通信可能に接続されたコンピューティングシステムを介して実施され、前記データソースは、コンテンツをホストするように構成されている1つまたは複数のリモートサーバを含み、1つまたは複数のローカルサーバが、前記コンピューティングシステムにおいて配置されており、前記方法は、

前記1つまたは複数のローカルサーバによって、収集エンジンによって、少なくとも1つの商品識別子に基づいて、前記ネットワーク化環境における前記1つまたは複数のリモートサーバによってホストされた前記コンテンツを検索するステップであって、前記検索を実行するために前記収集エンジンは、1つまたは複数の検索語から複数のクエリを構築し、前記収集エンジンによって構築された前記複数のクエリの少なくとも1つは、国際取引商品番号(GTIN)をマーケットプレイス固有識別子に変換し、かつウェブサイト内の特定のウェブページをターゲットにするための追加のクエリ構文を含む、ステップと、

前記1つまたは複数のローカルサーバによって、前記コンテンツの検索にตอบสนองして、検索結果の組を受け取るステップであって、各検索結果が、前記コンテンツにおいて識別された商品に関連している、ステップと、

前記1つまたは複数のローカルサーバによって、前記データソースから検索結果の前記組を収集するステップと、

前記1つまたは複数のローカルサーバによって、検索結果の前記組における各検索結果から複数の商品識別子を抽出するステップであって、前記複数の商品識別子が、検索結果ごとに、少なくとも1つの抽出されたGTIN、1つの抽出されたブランド名、1つの抽出された在庫管理単位(SKU)、1つの抽出された製品名、および1つの抽出された製品説明を含む、ステップと、

前記1つまたは複数のローカルサーバによって、検索結果の前記組における検索結果ごとに、前記抽出された商品識別子が、合法のブランド名の商品識別子に対応しているか否かに基づいて、前記抽出されたGTINが、合法であるかまたは不正であるかを分析するステップであって、前記抽出された商品識別子が合法的なブランド名の商品識別子に対応するか否かは、前記SKUが正しいか否かに少なくとも基づき、前記抽出された商品識別子が合法的なブランド名の商品識別子に対応するか否かは、前記製品名が正しいか否かに少なくとも基づき、かつ、前記抽出された商品識別子が合法的なブランド名の商品識別子に対応するか否かは、前記製品説明が正しいか否かに少なくとも基づく、ステップと、

前記1つまたは複数のローカルサーバによって、前記分析に基づいて、検索結果の前記組における各検索結果に合法または不正とタグ付けするステップと

10

20

30

40

50

を含む、方法。

【請求項 1 1】

前記1つまたは複数のローカルサーバによって、
前記抽出されたGTINに含まれているGS1事業者プリフィックスを検索すること、
GS1登録事業者情報検索サービスを検索すること、
エンティティのデータベースを前記エンティティのアプリケーションプログラミング
インターフェース(API)を介して検索すること、または
ブランドGTINの独立データベースを検索すること
のうちの少なくとも1つによって、前記抽出されたGTINが、前記ブランド名に基づいて、
合法であるかまたは不正であるかを分析するステップをさらに含む、請求項10に記載の方法。

10

【請求項 1 2】

前記1つまたは複数のローカルサーバによって、前記検索結果のうちの最初の検索結果
について、前記複数の商品識別子のうちの対応する商品識別子が、前記検索結果のうちの
前記最初の検索結果に含まれている1つまたは複数のあらかじめ定義された商品識別子に
対応しているかどうかを決定するステップと、
前記1つまたは複数のローカルサーバによって、前記複数の商品識別子のうちの前記対
応する商品識別子が、前記1つまたは複数のあらかじめ定義された商品識別子に対応して
いるかどうかに基づいて、前記検索結果のうちの前記最初の検索結果に合法または不正と
タグ付けするステップと
をさらに含む、請求項10に記載の方法。

20

【請求項 1 3】

前記1つまたは複数のローカルサーバによって、検索結果ごとに前記複数の商品識別子
を分析して、合法のブランド名の前記商品識別子に対応していない商品識別子を識別する
ステップと、
前記1つまたは複数のローカルサーバによって、前記識別された商品識別子の前記検索
結果に不正とタグ付けするステップと
をさらに含む、請求項10に記載の方法。

【請求項 1 4】

前記1つまたは複数のローカルサーバによって、ウェブサイトおよびアプリケーション
の直接検索、クエリ構築、ならびに前記ウェブサイトの電子カタログにより、製品情報を
収集するステップをさらに含む、請求項10に記載の方法。

30

【請求項 1 5】

前記ネットワーク化環境における前記1つまたは複数のリモートサーバが、ウェブサー
バであり、前記1つまたは複数のリモートサーバによってホストされた前記コンテンツが
、ウェブページを含むウェブサイトである、請求項10に記載の方法。

【請求項 1 6】

前記1つまたは複数のローカルサーバによって、不正とタグ付けされた1つまたは複数の
結果に関連している前記コンテンツの削除通知を、前記コンテンツをホストするように構
成されている1つまたは複数のリモートサーバに伝達するステップをさらに含む、請求項1
0に記載の方法。

40

【請求項 1 7】

前記1つまたは複数のローカルサーバによって、検索結果の前記組における各検索結果
からの前記複数の商品識別子の抽出に回答して、検索結果の前記組について複数のレコー
ドをデータベースに作成するステップであって、前記データベースに作成された前記複数
のレコードの各レコードが、検索結果の前記組における結果に対応する、ステップと、
前記1つまたは複数のローカルサーバによって、前記データベースに作成された前記複数
のレコードの対応するレコードに各結果から抽出された前記複数の商品識別子を記憶す
るステップと
をさらに含む、請求項10に記載の方法。

50

【請求項 18】

ユーザインターフェースを使用してさらに実施され、前記複数のレコードおよび前記複数の商品識別子を前記ユーザインターフェースにおいて表示するステップをさらに含む、請求項17に記載の方法。

【請求項 19】

不正コンテンツを識別するために、ネットワーク化されたコンテンツにおける商品識別子を収集し、構文解析し、分析するための命令を記憶する非一時的コンピュータ可読記憶媒体であって、前記命令は、実行されると、

1つまたは複数のローカルサーバによって、収集エンジンを介して、少なくとも1つの商品識別子に基づいて、ネットワーク化環境における1つまたは複数のリモートサーバによってホストされたコンテンツを検索することであって、前記検索を実行するために前記収集エンジンは、1つまたは複数の検索語から複数のクエリを構築し、前記収集エンジンによって構築された前記複数のクエリの少なくとも1つは、国際取引商品番号(GTIN)をマーケットプレイス固有識別子に変換し、かつウェブサイト内の特定のウェブページをターゲットにするための追加のクエリ構文を含む、検索することと、

10

前記1つまたは複数のローカルサーバを介して、前記コンテンツの検索に回答して、検索結果の組を受け取ることであって、各検索結果が、前記コンテンツにおいて識別された商品に関連している、受け取ることと、

前記1つまたは複数のローカルサーバを介して、データソースから検索結果の前記組を収集することと、

20

前記1つまたは複数のローカルサーバを介して、検索結果の前記組における各検索結果から複数の商品識別子を抽出することであって、前記複数の商品識別子が、検索結果ごとに、少なくとも1つの抽出されたGTIN、1つの抽出されたブランド名、1つの抽出された在庫管理単位(SKU)、1つの抽出された製品名、および1つの抽出された製品説明を含む、抽出することと、

前記1つまたは複数のローカルサーバを介して、検索結果の前記組における検索結果ごとに、前記抽出された商品識別子が、合法のブランド名の商品識別子に対応しているか否かに基づいて、前記抽出されたGTINが、合法であるかまたは不正であるかを分析することであって、前記抽出された商品識別子が合法的なブランド名の商品識別子に対応するか否かは、前記SKUが正しいか否かに少なくとも基づき、前記抽出された商品識別子が合法的なブランド名の商品識別子に対応するか否かは、前記製品名が正しいか否かに少なくとも基づき、かつ、前記抽出された商品識別子が合法的なブランド名の商品識別子に対応するか否かは、前記製品説明が正しいか否かに少なくとも基づく、分析することと、

30

前記1つまたは複数のローカルサーバを介して、前記分析に基づいて、検索結果の前記組における各検索結果に合法または不正とタグ付けすることと
を行う、非一時的コンピュータ可読記憶媒体。

【請求項 20】

命令をさらに記憶し、前記命令は、実行されると、前記1つまたは複数のローカルサーバを介して、

前記抽出されたGTINに含まれているGS1事業者プリフィックスを検索すること、

40

GS1登録事業者情報検索サービスを検索すること、

エンティティのデータベースを前記エンティティのアプリケーションプログラミングインターフェース(API)を介して検索すること、または

ブランドGTINの独立データベースを検索すること

のうちの少なくとも1つによって、前記抽出されたGTINが、前記ブランド名に基づいて、合法であるかまたは不正であるかを分析する、請求項19に記載の非一時的コンピュータ可読記憶媒体。

【請求項 21】

命令をさらに記憶し、前記命令は、実行されると、

前記1つまたは複数のローカルサーバを介して、前記検索結果のうちの最初の検索結果

50

について、前記複数の商品識別子のうちの対応する商品識別子が、前記検索結果のうちの前記最初の検索結果に含まれている1つまたは複数のあらかじめ定義された商品識別子に対応しているかどうかを決定することと、

前記1つまたは複数のローカルサーバを介して、前記複数の商品識別子のうちの前記対応する商品識別子が、前記1つまたは複数のあらかじめ定義された商品識別子に対応しているかどうかに基づいて、前記検索結果のうちの前記最初の検索結果に合法または不正とタグ付けすることと

を行う、請求項19に記載の非一時的コンピュータ可読記憶媒体。

【請求項22】

命令をさらに記憶し、前記命令は、実行されると、

前記1つまたは複数のローカルサーバを介して、検索結果ごとに前記複数の商品識別子を分析して、合法のブランド名の前記商品識別子に対応していない商品識別子を識別することと、

前記1つまたは複数のローカルサーバを介して、前記識別された商品識別子の前記検索結果に不正とタグ付けすることと

を行う、請求項19に記載の非一時的コンピュータ可読記憶媒体。

【請求項23】

命令をさらに記憶し、前記命令は、実行されると、前記1つまたは複数のローカルサーバを介して、ウェブサイトおよびアプリケーションの直接検索、クエリ構築、ならびに前記ウェブサイトの電子カタログにより、製品情報を収集する、請求項19に記載の非一時的コンピュータ可読記憶媒体。

【請求項24】

前記ネットワーク化環境における前記1つまたは複数のリモートサーバが、ウェブサーバであり、前記1つまたは複数のリモートサーバによってホストされた前記コンテンツが、ウェブページを含むウェブサイトである、請求項19に記載の非一時的コンピュータ可読記憶媒体。

【請求項25】

命令をさらに記憶し、前記命令は、実行されると、前記1つまたは複数のローカルサーバを介して、不正とタグ付けされた1つまたは複数の結果に関連している前記コンテンツの削除通知を、前記コンテンツをホストするように構成されている1つまたは複数のリモートサーバに伝達する、請求項19に記載の非一時的コンピュータ可読記憶媒体。

【請求項26】

命令をさらに記憶し、前記命令は、実行されると、

前記1つまたは複数のローカルサーバを介して、検索結果の前記組における各検索結果からの前記複数の商品識別子の抽出にตอบสนองして、検索結果の前記組について複数のレコードをデータベースに作成することであって、前記データベースに作成された前記複数のレコードの各レコードが、検索結果の前記組における結果に対応する、作成することと、

前記1つまたは複数のローカルサーバを介して、前記データベースに作成された前記複数のレコードの対応するレコードに各結果から抽出された前記複数の商品識別子を記憶することと

を行う、請求項19に記載の非一時的コンピュータ可読記憶媒体。

【請求項27】

命令をさらに記憶し、前記命令は、実行されると、前記複数のレコードおよび前記複数の商品識別子をユーザインターフェースにおいて表示する、請求項26に記載の非一時的コンピュータ可読記憶媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本出願は、参照によってその全体が本明細書に組み込まれている2016年12月30日に出願された米国特許仮出願第62/440,798号に対する優先権を主張するものである。

10

20

30

40

50

【背景技術】

【0002】

膨大な量のデジタルコンテンツが、インターネットなどのネットワーク化環境を介してアクセス可能である。このコンテンツは、複数のデータチャンネルおよび/またはデータソースにわたって広がり、ますます多くのコンテンツが、日々、利用可能になってきている。このコンテンツは大半が合法であるものの、コンテンツの一部には、不正または偽造であるものもある。

【発明の概要】

【課題を解決するための手段】

【0003】

本開示の実施形態によれば、不正コンテンツを識別するために、ネットワーク化されたコンテンツにおける商品識別子(item identifier)を収集し、構文解析し、分析するためのシステムが提供される。このシステムは、ネットワーク化環境におけるデータソースに通信可能に接続されたコンピューティングシステムを含む。データソースは、コンテンツをホストするように構成されている1つまたは複数のリモートサーバを含む。システムはまた、コンピューティングシステムにおいて配置されている1つまたは複数のローカルサーバを含む。1つまたは複数のローカルサーバは、少なくとも1つの最初の商品識別子に基づいて、ネットワーク化環境における1つまたは複数のリモートサーバによってホストされたコンテンツを検索するようにプログラミングされている。1つまたは複数のローカルサーバはまた、コンテンツの検索に回答して、検索結果の組を受け取るようにプログラミングされ、各検索結果は、コンテンツにおいて識別された商品に関連している。1つまたは複数のローカルサーバは、データソースから検索結果の組を収集するようにさらにプログラミングされている。1つまたは複数のローカルサーバはまた、検索結果の組における各検索結果から複数の商品識別子を抽出するようにプログラミングされている。複数の商品識別子は、検索結果ごとに、少なくとも1つのGTINおよび1つのブランド名を含む。1つまたは複数のローカルサーバは、検索結果の組における検索結果ごとに、GTINが、ブランド名に基づいて、合法であるかまたは不正であるかを分析するようにさらにプログラミングされている。1つまたは複数のローカルサーバはまた、分析に基づいて、検索結果の組における各検索結果に合法または不正とタグ付けするようにプログラミングされている。

【0004】

本開示の実施形態によれば、不正コンテンツを識別するために、ネットワーク化されたコンテンツにおける商品識別子を収集し、構文解析し、分析するための方法が提供される。この方法は、ネットワーク化環境におけるデータソースに通信可能に接続されたコンピューティングシステムを使用して実施され、データソースは、コンテンツをホストするように構成されている1つまたは複数のリモートサーバを含み、1つまたは複数のローカルサーバが、コンピューティングシステムにおいて配置されている。方法は、1つまたは複数のローカルサーバによって、少なくとも1つの最初の商品識別子に基づいて、ネットワーク化環境における1つまたは複数のリモートサーバによってホストされたコンテンツを検索するステップを含む。方法はまた、1つまたは複数のローカルサーバによって、コンテンツの検索に回答して、検索結果の組を受け取るステップを含み、各検索結果は、コンテンツにおいて識別された商品に関連している。方法は、1つまたは複数のローカルサーバによって、データソースから検索結果の組を収集するステップをさらに含む。方法はまた、1つまたは複数のローカルサーバによって、検索結果の組における各検索結果から複数の商品識別子を抽出するステップを含む。複数の商品識別子は、検索結果ごとに、少なくとも1つのGTINおよび1つのブランド名を含む。方法は、1つまたは複数のローカルサーバによって、検索結果の組における検索結果ごとに、GTINが、ブランド名に基づいて、合法であるかまたは不正であるかを分析するステップをさらに含む。方法はまた、1つまたは複数のローカルサーバによって、分析に基づいて、検索結果の組における各検索結果に合法または不正とタグ付けするステップを含む。

【0005】

10

20

30

40

50

実施形態の任意の組合せ形態および/または置換形態が想到される。他の目的および特徴が、添付の図面とともに検討される次の詳細な説明から明らかになる。しかしながら、図面は、単に例示として設計されているにすぎず、本開示の限定の定義として設計されているわけではないことを理解されたい。

【0006】

図面では、同様の参照数字は、限定的でも包括的でもない実施形態の様々な図全体を通じて同様のパーツを示している。

【図面の簡単な説明】

【0007】

【図1】本開示の実施形態によるネットワーク化環境における様々なデータチャンネルに関連している異なるデータソースから不正コンテンツを収集し、構文解析し、分析し、その除去を容易にするための一例示的不正コンテンツ検出エンジンのブロック略図である。

10

【図2】本開示の実施形態による一例示的コンピューティングデバイスのブロック略図である。

【図3】本開示の実施形態によるインターネット上の不正コンテンツを収集し、構文解析し、分析し、その除去を容易にするためのネットワーク化環境の一例である。

【図4】本開示の実施形態による不正コンテンツ検出エンジンを使用してネットワークコンテンツにおいて識別された商品に関連している商品識別子を構文解析し、カテゴリ化するための一例示的方法を示すフローチャートである。

【図5】本開示の実施形態によるネットワーク化環境における不正コンテンツを収集し、構文解析し、分析し、その除去を容易にするための一例示的方法を示すフローチャートである。

20

【発明を実施するための形態】

【0008】

本開示の例示的な実施形態は、不正コンテンツを識別するために、ネットワーク化環境上のデジタルコンテンツにおいて識別された商品に関連している商品識別子を収集し、構文解析し、分析し、最終的には、ネットワーク化環境から不正コンテンツを除去するためのシステム、方法、および非一時的コンピュータ可読媒体に関する。システムおよび方法は、収集エンジンと、抽出エンジンと、タグ付けエンジンと、分析エンジンとを含む不正コンテンツ検出エンジンを含む。不正コンテンツ検出エンジンは、ネットワーク化環境におけるデータソースに通信可能に接続され得る。データソースは、コンテンツをホストするように構成されている1つまたは複数のリモートサーバを含むことができる。1つの非限定適用例においては、不正コンテンツ検出エンジンは、ブランド保護を目的として、不正コンテンツを識別し除去するために、ネットワーク化環境を検索するように構成され得る。

30

【0009】

一例示的实施形態では、収集エンジンは、ネットワーク化環境におけるデータソースからコンテンツを収集するように構成されている。具体的には、収集エンジンは、検索語および/または商品識別子に基づいて、データソースにおけるコンテンツを検索する。商品識別子は、万国製品コード(UPCコード:universal product code)、国際標準図書番号(ISBN:international standard book number)、および欧州統一商品番号(EAN:European article number)を含む国際取引商品番号(GTIN:global trade item number)と、ブランド名と、モデル番号との組合せ、ならびに製品リストを標準化するためにインターネット検索サイト、オンラインマーケットプレイス、および/またはオンラインオークションサイトによって使用される他の標準化された識別子を含むことができる。GTINは、取引商品、製品、またはサービスを識別するのに使用されるグローバルに一意的な番号である。データソースには、限定するものではないが、製品リストを含む、インターネット、マーケットプレイス/オークションウェブサイトが挙げられる。収集エンジンは、データソースにおける検索語の直接検索を使用して、および/または既存の商用サイトカタログ構造を利用して、コンテンツを検索する。たとえば、収集エンジンは、1つまたは複数のデータソースにわたって複数の製品リストにおける特定のGTINまたはキーワード文字列を検索するよう

40

50

に構成され得る。収集エンジンは、データソースの検索に回答して、検索結果の組またはリストを返すことができる。たとえば、収集エンジンは、特定のGTINまたはキーワード文字列に関連しているウェブページおよび/または製品のリストを返してもよい。

【0010】

抽出エンジンは、収集エンジンによって返された結果の組における各結果から商品識別子を抽出または構文解析する。抽出エンジンは、結果ごとに、データベースエントリまたはレコードを作成し、結果から抽出された各商品識別子は、レコードのフィールドに対応する。一例示的实施形態では、各レコードは、コンテンツから構文解析された少なくとも1つの抽出されたGTINおよび1つの抽出されたブランド名を含む。抽出エンジンは、各結果に含まれている任意の追加の商品識別子をさらに抽出し、または構文解析し、それらの追加の商品識別子をデータベースの中にフィールドとして記憶する。各商品識別子は、抽出エンジンが、構文解析された商品識別をカテゴリ化し、それを正しいフィールドの中に入れることを可能にする、認識された商品識別子カテゴリである。たとえば、抽出エンジンは、抽出された名称をブランド名としてカテゴリ化し、その名称をブランド名に対応するフィールドの中に入れるように構成されている。

10

【0011】

いくつかの事例では、抽出エンジンは、ある結果から1つまたは複数の商品識別子を識別し、認識し、および/またはカテゴリ化することができない場合がある。非限定例として、抽出エンジンは、ある結果においてリスト化された欧州統一商品番号(EAN)を認識することも、またはカテゴリ化することもできない場合がある。そのような事例では、アナリストは、その結果をレビューし、EANをEAN商品識別子としてカテゴリ化することができる。新規にカテゴリ化された商品識別子は、今後の抽出中に使用するために、既知の商品識別子とともに記憶される。

20

【0012】

各レコードについて、タグ付けエンジンは、抽出されたGTINが、抽出されたブランド名に対して合法であるかどうかを決定する。一例示的实施形態では、タグ付けエンジンは、GTINに含まれているGS1事業者プリフィックス(GS1 company prefix)を検索すること、および/またはGS1登録事業者情報検索サービス(GEPIR:Global Electronic Party Information Registry)においてGTINを検索すること、および/またはエンティティのデータベースにおけるGTINをエンティティのアプリケーションプログラミングインターフェース(API:application programming interface)を介して検索すること、および/またはブランドGTINの独立データベースにおいてGTINを検索することによって、抽出されたGTINに関連している合法的ブランド名を決定する。いくつかの実施形態では、エンティティは、GTINに関連している商品を所有する、および/またはその商品に関心がある事業者あるいは企業とすることができる。GS1事業者プリフィックスは、ブランドを識別するすべての登録済みUPC/EANに含まれている桁数である。GEPIRは、バーコード/GTIN、および/または事業者名、および/またはブランドを検証するように構成されたデータベースである。タグ付けエンジンは、抽出されたブランド名に対する合法的ブランド名をレビューし、抽出されたGTINが、抽出されたブランド名に対して合法であるかどうかに基づいて、各レコードに合法または不正とタグ付けする。

30

40

【0013】

分析エンジンは、結果から構文解析された商品識別子をさらに分析して、不正ネットワークコンテンツを識別および/または検出する。このステップは、抽出されたGTINが抽出されたブランド名に対して合法でないことをタグ付けエンジンがすでに決定している場合には、スキップされ得る。一例示的实施形態では、分析エンジンは、抽出された商品識別子が、ブランド名に関連しているあらかじめ定義された商品識別子に対応しているかどうかに基づいて、各レコードに合法または不正とタグ付けする。たとえば、分析エンジンは、抽出されたブランド名とあらかじめ定義されたブランド識別子との間の不整合を識別することができる。別の実施形態では、分析エンジンは、分析エンジンが商品識別子における不正特性を識別および/または検出するかどうかに基づいて、各レコードに合法または不

50

正とタグ付けする。不正特性のいくつかの例には、つづりが間違っている単語およびつづりが間違っているブランド名、正しくない最小在庫管理単位(SKU:stock keeping unit)、または正しくない製品説明を挙げることができる。

【0014】

いくつかの実施形態では、不正コンテンツ検出エンジンは、検出された不正コンテンツの自動削除を開始する除去エンジンをさらに含む。一旦、レコードが不正とタグ付けされると、除去エンジンは、不正コンテンツの削除を自動的に開始する。たとえば、除去エンジンは、デジタルミレニアム著作権法(DMCA:Digital Millennium Copyright Act)通知をコンテンツのホストまたはオーナーに送信することができる。DMCA通知は、不正のコンテンツまたは製品に関連している挿入情報を含むあらかじめ定義された通知を含むことができる。別の例では、除去エンジンは、削除通知をアプリケーションプログラミングインターフェース(API)を介してコンテンツのホストまたはオーナーに伝達することができる。

10

【0015】

システムの非限定例には、偽造である製品を売りに出したウェブページをホストするリモートサーバと通信する不正コンテンツ検出エンジンが含まれる。たとえば、ウェブページ上に示されている製品のブランドは「Samsung」でありながら、製品のウェブページ上にリストされているGTINは、ブランド「Apple」に関連付けられている場合があり、結果的にブランドの混乱が生じる。収集エンジンは、コンテンツを検索し、典型的には、結果の組における結果としてウェブページを収集する。たとえば、収集エンジンは、ウェブページのHTMLをデータベースにダウンロードすることができる。抽出エンジンは、ウェブページから商品識別子(すなわち、GTINおよびブランド名)を構文解析し、タグ付けエンジンは、GTINがブランド名に対して合法でないことを決定する。次いで、除去エンジンは、削除通知をウェブページのホストに送信し、ウェブページにおいて売りに出された製品が不正であることが識別される。

20

【0016】

本明細書に記載の方法およびシステムは、ブランド保護の目的として、インターネット検索および電子製品カタログからの製品情報の効率的かつ効果的な集約を可能にする。このシステムは結果的に、より完全、効率的なデータ読出し、および不正製品の入手可能性を減少させるためのより実行可能なデータをもたらす。これに関して、本開示の例示的な諸実施形態は、不正コンテンツを含んでいる可能性のある、増加しつつあるオンラインマーケットプレイスおよびウェブページを収集し、これらの環境から不正コンテンツを迅速に除去するための効率的かつ効果的なツールを提供する。

30

【0017】

図1は、本開示の実施形態によるインターネット上の、または他の任意のネットワーク化環境における様々なデータチャンネルに関連している異なるデータソース102から不正コンテンツを収集し、構文解析し、分析し、その除去を容易にするための一例示的な不正コンテンツ検出エンジン100のブロック略図である。エンジン100は、ユーザインターフェース110、収集エンジン115、抽出エンジン120、タグ付けエンジン125、および分析エンジン130を含む。

【0018】

一例示的な実施形態では、収集エンジン115は、ウェブおよび/またはダークウェブを巡回すること、ウェブページ(マーケットプレイス/オークションウェブページを含む)を検索するための検索エンジンおよび/またはAPIを収集すること、モバイルアプリケーションデータを検索すること、ならびに/あるいはネットワーク化環境における他の任意のコンテンツを検索することによって、不正コンテンツのオンラインコンテンツを検索するように構成されている。収集エンジン115は、商品識別子、キーワード文字列、またはそれらの組合せに基づいて、ネットワーク化環境におけるコンテンツを検索する。たとえば、収集エンジン115は、GTINおよび/または検索語に基づいて、オンラインマーケットプレイスにおけるウェブページを検索することができる。追加の事例では、収集エンジン115は、アプリケーションプログラミングインターフェース(API)を介して、1つまたは複数のデータソ

40

50

ースと通信する。

【0019】

一例示的实施形態では、収集エンジン115は、インターネットコンテンツに含まれている商品のGTINを利用して、存続可能な製品リストを提案することができる。これにより、キーワード多様性またはタイトル多様性に起因して、存続可能な製品リストを検出および収集できないことがなくなり、結果として、より完全、効率的なデータ読出しがもたらされる。一代替的实施形態では、収集エンジン115は、1つまたは複数のグラフィカルユーザインターフェース114を介して、1人または複数人のユーザ104によって入力された1つまたは複数の検索語(たとえば、キーワード)に基づいて、1つまたは複数のクエリ(たとえば、データベース、API、またはウェブベースのクエリ)を生成あるいは構築する。1つの例として、収集エンジン115は、検索語の単一の組からいくつかのクエリを構築することができる。各クエリは、検索エンジンおよび/またはアプリケーションプログラミングインターフェース(API)に固有とすることができる。

10

【0020】

収集エンジン115は、同様のコンテンツの様々なデータソースの並行検索を容易にするようにプログラミングされ得る。クエリは、構造化クエリ言語(SQL:Structured Query Language)、コンテキストクエリ言語(CQL:Contextual Query Language)、プロプライエタリクエリ言語(proprietary query language)、ドメイン固有クエリ言語、および/または他の任意の適切なクエリ言語など、1つまたは複数のクエリ言語を使用して生成あるいは構築され得る。いくつかの実施形態では、収集エンジン115は、Java(登録商標)、C、C++、Perl、およびRubyなど、1つまたは複数のプログラミング言語もしくはスクリプトを使用して、1つまたは複数のクエリを生成あるいは構築することができる。

20

【0021】

収集エンジン115は、検索エンジンおよび/またはAPIにより、各GTINおよび/またはキーワードクエリを実行することができる。これにより、インターネットコンテンツおよび/またはネットワーク化環境における他の任意のコンテンツを返すことができる。たとえば、収集エンジン115は、写真を含む物品について作成されたウェブページ(たとえば、製品ページ)、および物品に関する他の商品識別子(たとえば、GTIN、説明、仕様書、寸法など)を収集することができる。

【0022】

1つの例としては、収集エンジン115を実行すると、1つまたは複数のデータソースで1つまたは複数のウェブサーバによってホストされた1つまたは複数のインターネットドメインから1つまたは複数のウェブページを返すことができる。いくつかの実施形態では、検索結果は、リストとして返すことができ、収集エンジン115は、エンジン100によって処理されることになる結果の量を制限することができる。一非限定例としては、収集エンジン115は、たとえば検索結果においてリストされた最初の100のウェブページ(もしくは最初の10ページの検索結果)、または任意の適切な量の結果を選定してもよい。エンジン100によって処理されるように選定される結果の量は、エンジン100によって、および/またはエンジン100のユーザ104によって指定され得る。

30

【0023】

例示的な実施形態では、収集エンジン115を介して返された結果は、フェッチされ、記憶デバイスにダウンロードされ、収集されたデータセット117として記憶される。たとえば、各結果(たとえば、各ウェブページ)は、ファイルまたは他のデータ構造として記憶され得る。いくつかの事例では、結果のうちの1つまたは複数は、それを読み出すデータソース上にあるのと同じフォーマットで記憶されてよい。たとえば、ウェブページは、それらのネイティブテキストベースのマークアップ言語(たとえば、HTMLおよびXHTML)で記憶され得る。いくつかの事例では、結果のうちの1つまたは複数は、それを読み出すデータソース上に記憶されているフォーマットとは異なるフォーマットで記憶されてよい。

40

【0024】

収集エンジン115を介して返されたウェブページのうちの少なくとも1つは、カタログ

50

モデルを利用するウェブページからのものとするができる。たとえば、収集エンジン115は、購入可能な物品/サービスのリストを含む1つまたは複数のマーケットプレイスウェブサイトのウェブページを検索することができる。そのようなマーケットプレイスウェブサイトでは、複数のサードパーティ出品者が、同じ物品またはサービスを販売することが可能な場合が多く、購入者は、その物品またはサービスをどの出品者から購入すべきかを選択できる。そのような事例では、マーケットプレイスウェブサイトの中には、各出品者によって売りに出された各物品/サービス(たとえば、製品)の別個のウェブページを含む場合があるものもあれば、他のマーケットプレイスウェブサイトの中には、カタログモデルの利用が可能なものもある。

【0025】

カタログモデルを利用するオンラインマーケットプレイスの1つの非限定例は、Amazonによるamazon.comであり、ここでは、製品ページ(たとえば、Amazonマーケットプレイスにおいて販売されている特定の製品のウェブページ)は、同じ製品の多数の他の出品者を識別することができ、および/または同じ製品を販売する出品者のリストへのリンクを提供することができる。この例では、各出品者は、自分自身の製品ページを受け取ることはない。そうではなく、製品ページは、デフォルト出品者を識別し、ユーザは、製品の他の出品者を閲覧するためには、他の出品者のリストへの1つまたは複数のリンク(たとえば、製品を新品として販売する出品者のための「新品」リンク、製品を中古として販売する出品者のための「中古」リンク、製品を修理済みとして販売する出品者のための「修理済み」リンク)を選定しなくてはならない。収集エンジン115は、カタログモデルを使用するウェブサイトを識別し、サードパーティ出品者についての商品識別子をフェッチするように構成されている。たとえば、収集エンジン115は、Amazonの製品広告APIを利用するなど、既存のショッピングサイトカタログ構造を利用する検索エンジンまたはAPIを使用して、製品リストについて検索するようにプログラミングされ得る。

【0026】

いくつかの事例では、マーケットプレイスウェブサイトは、マーケットプレイスにおいて利用可能な各GTINを別個の一意のマーケットプレイス固有識別子に割り当てることができる。このマーケットプレイス固有識別子は、マーケットプレイスウェブサイトによって使用されて、GTINの代わりにマーケットプレイスウェブサイト上の製品を一意に識別することができる。マーケットプレイスウェブサイトは、これらのマーケットプレイス固有識別子をそれらのウェブページおよび/またはユニフォームリソースロケータ(URL:uniform resource locator)に組み込むことができる。マーケットプレイス固有識別子を利用するオンラインマーケットプレイスの一例が、製品にAmazon標準識別番号(ASIN:Amazon Standard Identification Number)を使用する、Amazonによるamazon.comである。ASINは、製品詳細欄における製品ページ上に、ページ自体のURLにリストされる。しかしながら、amazon.comカタログに新規ページを作成する際、出品者には、業界標準の商品識別子、通常はGTINを使用することが要求される。1つの実施形態では、収集エンジン115は、マーケットプレイス固有識別子を利用する特定のマーケットプレイスをターゲットにする場合、GTINをマーケットプレイス固有識別子に変換するように構成されている。たとえば、1つの実施形態では、収集エンジン115は、ウェブサイト(たとえば、オンラインマーケットプレイスウェブサイト)内の特定のウェブページをターゲットにするために、追加のクエリ構文とともにマーケットプレイス固有識別子を含むように構成されている。

【0027】

抽出エンジン120は、収集されたデータセット117を構文解析し、収集されデータセット117における各結果(たとえば、各ウェブページおよび関連のメタデータ)から商品識別子を抽出する。一例示的实施形態では、結果のうちの1つまたは複数は、少なくとも1つの抽出されたGTINおよび1つの抽出されたブランド名を含むことができる。抽出エンジン120が各結果から商品識別子を抽出すると、抽出エンジン120は、商品識別子の商品識別子データベース135(たとえば、リレーショナルデータベースまたはNoSQLデータベース)を構築する。抽出エンジン120は、レコードおよび関連の商品識別子を商品識別子データベ

10

20

30

40

50

ース135に追加する。たとえば、抽出エンジン120は、収集されたデータセット117における結果ごと(たとえば、ウェブページごと)に1つのレコードを作成し、ある結果から抽出された各商品識別子は、そのレコードのフィールドに対応することができる。各商品識別子は、商品識別子カテゴリ(すなわち、ブランド名、製品名、UPC、EANなど)であり、それにより、抽出エンジンは、構文解析された商品識別子を正しくカテゴリ化し、それらを正しいフィールドに入れることが可能になる。

【0028】

結果から抽出された商品識別子は、各レコードが合法コンテンツに対応しているかまたは不正コンテンツに対応しているかを評価する際に有用であり得る情報を含む。結果から抽出され、商品識別子データベース135におけるフィールドに記憶されている商品識別子は、たとえば、製品名、製品説明、出品者名、GTINまたは他の商品識別子、出品者の住所、出品者が製品を出荷する先の住所、出品者レビュー、および/または結果のタイトル(ウェブページのタイトル)などのテキスト;価格、購入可能な製品の数量、製品寸法、および/またはマーケットプレイス固有の識別子などの数字;製品画像、ロゴ、および/またはイラストなどの画像;動画および/または音声などの他の媒体;ウェブページ用のドメインの登録者名;ウェブページをホストするネームサーバ;ならびにHTMLページソースコード、XMLファイル、およびJavaScript(登録商標)などの生データとすることができる。

【0029】

収集されたデータセット117における結果から商品識別子を抽出するには、抽出エンジン120は、たとえば、自然言語処理、機械学習、類似性測定、画素マッチング法および/またはパターンマッチング法を含む画像マッチング法を使用して、結果における商品識別子を識別することができる。抽出エンジン120は、エンティティの1つまたは複数のオントロジーを利用して、結果に含まれているエンティティ(たとえば、出品者)を導き出し、および/または識別することができる。様々なアルゴリズムおよび/または技法が、抽出エンジン120によって利用され得る。たとえば、Baeza-Yates-Gonnetなど、ファジーテキストパターンマッチングのためのアルゴリズムは単一の文字列に使用され得、ファジーAho-Corasickは複数の文字列マッチングに使用され得、教師付きまたは教師なしの文書分類のためのアルゴリズムは、ファジーAho-Corasickなどの複数の文字列ファジーテキストパターンマッチングアルゴリズムを使用して、ならびに潜在ディリクレ配分(LDA:Latent Dirichlet Allocation)および階層ディリクレ過程(HDP:Hierarchical Dirichlet Process)などのトピックモデルを使用してテキスト文書を数値ベクトルに変形後に、使用され得る。

【0030】

一代替的实施形態では、収集エンジン115によってコンテンツをダウンロードして、収集されたデータセット117を作成するのではなく、収集エンジン115が、コンテンツ(すなわち、ウェブページ)を識別し、抽出エンジン120が、コンテンツから直接、商品識別子を構文解析する。抽出エンジン120は、上述の商品識別子を使用して商品識別子データベース135を作成する。

【0031】

一旦、商品識別子が抽出エンジン120によって抽出され、商品識別子データベース135に記憶されると、タグ付けエンジン125は、抽出されたGTINが、抽出されたブランド名に対して合法であるかどうかをレコードごとに決定する。タグ付けエンジン125は、GTINに含まれているGS1事業者プリフィックスを検索すること、および/またはGS1登録事業者情報検索サービス(GEPIR)を検索すること、および/またはエンティティのデータベースをエンティティのAPIを介して検索すること、および/またはブランドGTINの独立データベースを検索することにより、GTINに関連している合法的ブランド名を決定する。タグ付けエンジン125は、抽出されたGTINが、抽出されたブランド名に対して合法であるかどうかに基づいて、各レコードに合法または不正であるとタグ付けする。

【0032】

分析エンジン130は、合法と識別された、商品識別子データベース135の中のレコード

10

20

30

40

50

に関する商品識別子を(たとえば、消費識別子データベース135の中のレコードのタグ付けに基づいて)レビューする。たとえば、商品識別子データベース135の中のレコードの合法である(たとえば、偽造ではない)とのタグ付けに回答して、分析エンジン130の例示的な実施形態は、抽出され構文解析された商品識別子を分析して、各識別子とそのブランドのあらかじめ定義されたおよび/または既知の商品識別子と対応しているかどうかを決定する。一非限定例としては、分析エンジン130は、ブランドの抽出された製品寸法がブランドのあらかじめ定義されたおよび/または既知の製品寸法と対応しているかどうかを分析することができる。

【0033】

一追加の実施形態では、分析エンジン130は、抽出された商品識別子を分析して、疑わしいコンテンツ(たとえば、不正製品)を識別する。インターネット検索およびショッピングサイトにおいてリストされている製品情報は、出品者から入力またはアップロードされたデータから提供される場合が多い。いくつかの既知の事例では、不正または偽造製品を販売する出品者は、不十分なまたは正しくない商品識別子、たとえば、つづりが間違っているブランド名および/または製品名、正しくない製品数量、正しくない製品寸法、正しくないGTIN、誤った最小在庫管理単位(SKU)、あるいは正しくないまたは大きくつづりが間違っている製品説明を含んでいるデータを提供する。いくつかの実施形態では、ユーザ104は、分析の実施およびパラメータを制御するために、1つまたは複数のグラフィカルユーザインターフェース114を介して分析エンジン130と対話することができる。たとえば、ユーザ104は、商品識別子データベース135の中のレコードについてあらかじめ定義されたおよび/または既知の商品識別子を指定することができる。別の実施形態では、分析エンジン130は、商品識別子データベース135のレコードを自律的に分析するように構成されている。たとえば、分析エンジン130は、疑わしいコンテンツを識別するために、1つまたは複数の機械学習アルゴリズムを利用して、商品識別子データベース135の中のレコードを分析するように構成され得、この場合、機械学習アルゴリズムは、訓練データのコーパスを用いて訓練され得る。

【0034】

いくつかの実施形態では、合法的タグオプションが、商品識別子データベース135の中の各レコードに関連付けられる。タグ付けエンジン125および分析エンジン130は、レコード(そのため、その関連のウェブページ)が合法コンテンツまたは不正コンテンツに対応していることを示す合法タグまたは不正タグをレコードに付すことができる。

【0035】

追加の実施形態では、除去エンジン119が、検出された不正コンテンツおよび/または製品の自動削除を開始する。一旦、結果が不正とタグ付けされる、または決定されると、除去エンジン119は、不正コンテンツの削除要求を開始する。たとえば、除去エンジン119は、構造化されたファイルまたは電子メールを生成するために、収集されたデータセット117および/または商品識別子データベース135からレコードを読み出すことによって、デジタルミレニウム著作権法(DMCA)通知を生成することができる。通知が生成された後、除去エンジン119は、その通知をコンテンツのホストまたはオーナーに送信することができる。別の例では、除去エンジン119は、削除通知をコンテンツのホストまたはオーナーにAPIを介して伝達する。

【0036】

例示的な実施形態では、ユーザインターフェース110が、たとえば商品識別子データベース135の閲覧を用いて、検索からのレコードのリストを含むように1つまたは複数のグラフィカルユーザインターフェース(GUI: graphical user interface)114を生成することができ、レコードは、収集されたデータセット117から抽出された商品識別子のうちの1つまたは複数に基づいて、1つまたは複数のグラフィカルユーザインターフェース114においてグループ分けされ得る。1つの非限定例としては、不正コンテンツとタグ付けされた商品識別子データベース135に関連しているレコードが、グラフィカルユーザインターフェース114に示される。

10

20

30

40

50

【 0 0 3 7 】

ユーザインターフェース110は、提示/可視化エンジン112、および1つまたは複数のグラフィカルユーザインターフェース114を含む。提示エンジン112は、エンジン100において実施される1つまたは複数のサービスおよび/またはエンジン間のインターフェースを提供するように構成され得る。データを受け取ると同時に、提示エンジン112は、グラフィカルユーザインターフェース114のうちの1つまたは複数を作成するように、および1つまたは複数のグラフィカルユーザインターフェース114においてデータをレンダリングするように実行され得る。1つまたは複数のグラフィカルユーザインターフェース114は、ユーザ104がエンジン100と対話することを可能にすることができ、ユーザ104に情報を表示するためのデータ出力領域、ならびにユーザ104から情報を受け取るためのデータエントリフィールドを含むことができる。データ出力領域のいくつかの例には、限定するものではないが、テキスト、グラフィックス(たとえば、グラフ、地理上のもしくはその他の地図、および画像など)、および/または他の任意の適切なデータ出力領域を挙げることができる。データエントリフィールドのいくつかの例には、限定するものではないが、テキストボックス、チェックボックス、ボタン、ドロップダウンメニュー、および/または他の任意の適切なデータエントリフィールドを挙げることができる。

10

【 0 0 3 8 】

ユーザインターフェース110は、1つまたは複数のローカルサーバおよび/あるいは1つまたは複数のユーザコンピューティングデバイスによって実行されるエンジン100の実施形態によって生成され得る。ユーザインターフェース110は、本明細書に説明したようにコンテンツ(たとえば、インターネットコンテンツ)から抽出された商品識別子をレンダリングするように構成され得、本明細書に説明したように商品識別子データベース135のレコードに記憶され得、レコードは、コンテンツの検索から生じた結果(たとえば、ウェブページ)ごとに作成される。ユーザインターフェース110は、ユーザ104がコンテンツから抽出された商品識別子とそれを通じて対話することができるインターフェースを提供することができる。たとえば、ユーザインターフェース110は、収集エンジン115および抽出エンジン120を介して集められたウェブページから抽出された商品識別子の構造化された構成を提供するように構成され得る。

20

【 0 0 3 9 】

一非限定例として、ユーザインターフェース110は、商品識別子データベース135の中のレコードについてのエントリもしくは行を含むリストまたはテーブルを提供することができ、レコードについてのエントリまたは行に、そのレコードに関連している商品識別子を置くことができる。別の非限定例としては、ユーザインターフェース110は、商品識別子データベース135の中のレコードについての領域もしくはタイトルを含む2次元の配列またはタイル状構成を提供することができ、各領域またはタイルに、その領域またはタイルに対応するレコードに関連している商品識別子を置くことができる。1つの非限定例としては、ユーザインターフェース110は、収集エンジン115を介して集められたウェブページについてのエントリのリストを含むことができる。たとえば、行は、ウェブページに対応する、商品識別子データベース135の中のレコードに関連付けることができる。各行は、ウェブページから抽出された商品識別子を含むことができ、各列は、ウェブページから抽出された商品識別子のカテゴリまたはタイプを含むことができる。たとえば、列についての商品識別子カテゴリは、ウェブページから抽出されたタイトル、ウェブページ上に提示された製品に割り当てられたGTIN、ウェブページ上に提示された製品の価格、ウェブページが不正コンテンツ検出エンジンの一実施形態によって収集されたときを示す検出日付、ウェブページを介して製品を販売するエンティティに関連しているエンティティ名、ウェブページを介して製品を販売するエンティティに関連している住所、ウェブページを介して製品を販売するエンティティに関連しているレーティング、出品者がウェブページ上に提示された製品を出荷することになる先の住所、およびウェブページをホストしているドメイン/マーケットプレイス名を含むことができる。

30

40

【 0 0 4 0 】

50

行、および/または行における商品識別子は、商品識別子を修正するために、および/または1つまたは複数のアクションを行うために、ユーザ104がリストと対話できるようにユーザ104によって選択可能とすることができる。たとえば、抽出エンジン120が、ある結果から1つまたは複数の商品識別子を構文解析することができない場合、アナリストは、その結果をレビューし、1つまたは複数の商品識別子を行に入力することができる。次いで、入力された商品識別子は、製品が合法であるかまたは不正であるかを決定する際にタグ付けエンジン125および分析エンジン130によって使用され得る。

【0041】

エンジン100は、ネットワーク化環境におけるコンテンツを収集エンジン115が再クエリする頻度をユーザ104が指定することを可能にするために、再収集頻度オプションをさらに含むことができる。たとえば、ユーザ104は、収集エンジン115が、毎時、毎日、毎週、毎月、および四半期ごとなどに検索することを指定することができる。

【0042】

図2は、本開示の実施形態による一例示的コンピューティングデバイスのブロック略図である。本実施形態では、コンピューティングデバイス200は、エンジン100の動作および/または機能のうちの1つまたは複数を実行するように、ならびにインターネットまたは他のネットワーク化環境上の不正コンテンツの検出および除去を容易にするようにプログラミングおよび/または構成されているサーバとして構成されている。コンピューティングデバイス200は、例示的な実施形態を実施するための1つまたは複数のコンピュータ実行可能命令もしくはソフトウェアを記憶するための1つまたは複数の非一時的コンピュータ可読媒体を含む。非一時的コンピュータ可読媒体には、限定するものではないが、1つまたは複数のタイプのハードウェアメモリおよび非一時的有形媒体(たとえば、1つまたは複数の磁気記憶ディスク、1つまたは複数の光学ディスク、1つまたは複数のフラッシュドライブ)などを挙げることができる。たとえば、コンピューティングデバイス200に含まれているメモリ206は、エンジン100またはその一部分の例示的な実施形態を実施するためのコンピュータ可読およびコンピュータ実行可能命令またはソフトウェアを記憶することができる。

【0043】

コンピューティングデバイス200はまた、メモリ206に記憶されたコンピュータ可読およびコンピュータ実行可能命令またはソフトウェア、ならびにシステムハードウェアを制御するための他のプログラムを実行するために、構成設定可能なおよび/またはプログラマブルのプロセッサ202と関連のコア204、オプションで、1つまたは複数の追加の構成設定可能なおよび/またはプログラマブルのプロセッサ202'と関連のコア204'(たとえば、複数のプロセッサ/コアを有するコンピュータシステムの場合)を含む。プロセッサ202およびプロセッサ202'はそれぞれ、単一のコアプロセッサであっても、または複数のコア(204および204')プロセッサであってもよい。

【0044】

仮想化がコンピューティングデバイス200において採用されてもよく、それにより、コンピューティングデバイスにおけるインフラストラクチャおよびリソースは、動的に共有され得るようになる。複数のプロセッサにおいて動作する処理を、この処理が、複数のコンピューティングリソースではなく、たった1つのコンピューティングリソースを使用しているように見えるように扱うために、ならびに/またはエンジン100に関連している機能および動作を行うようにコンピューティングリソースを割り当てるために、1つまたは複数の仮想マシン214を設けることができる。複数の仮想マシンはまた、1つのプロセッサとともに使用することも、またはいくつかのプロセッサにわたって分散することもできる。

【0045】

メモリ206は、コンピュータシステムメモリ、またはDRAM、SRAM、およびEDO RAMなどのランダムアクセスメモリを含むことができる。メモリ206は、同様に、他のタイプのメモリ、またはそれらの組合せを含むことができる。

【0046】

10

20

30

40

50

コンピューティングデバイス200はまた、データ、ならびに本明細書に記載のエンジン100の例示的な実施形態を実施するように処理デバイス202によって実行可能であるコンピュータ可読命令および/またはソフトウェアを記憶するためのハードドライブ、CD-ROM、大容量フラッシュドライブあるいは他のコンピュータ可読媒体など、1つまたは複数の記憶デバイス224を含むことができる。

【0047】

コンピューティングデバイス200は、1つまたは複数のネットワーク、たとえば、ローカルエリアネットワーク(LAN)、広域ネットワーク(WAN)、あるいは限定するものではないが、標準電話回線、LANもしくはWANリンク(たとえば、802.11、T1、T3、56kb、X.25)、ブロードバンド接続(たとえば、ISDN、フレームリレー、ATM)、ワイヤレス接続(セルラ基地局経路を含む)、コントローラエリアネットワーク(CAN)、または上記のうちのいずれかもしくはすべての何らかの組合せを含む多様な接続によるインターネットと1つまたは複数のネットワークデバイス222を介してインターフェースをとるよう構成されているネットワークインターフェース212を含むことができる。ネットワークインターフェース212は、通信ができ本明細書に記載の動作を行うことができる任意のタイプのネットワークとのコンピューティングデバイス200のインターフェースをとるのに適切な内蔵ネットワークアダプタ、ネットワークインターフェースカード、PCMCIAネットワークカード、カードバスネットワークアダプタ、ワイヤレスネットワークアダプタ、USBネットワークアダプタ、モデム、または他の任意のデバイスを含むことができる。図2に示すコンピューティングデバイス200はサーバとして実装されているが、コンピューティングデバイス200の例示的な実施形態は、ワークステーション、デスクトップコンピュータ、またはワイヤレス通信かもしくはワイヤード通信かのいずれかによって他のデバイスと通信することができる、ならびに本明細書に説明の動作を行うのに十分なプロセッサパワーおよびメモリ容量を有する他の形態のコンピューティングまたは通信デバイスなど、任意のコンピュータシステムとすることができる。

【0048】

コンピューティングデバイス200は、任意のUnixベースのサーバアプリケーション、Linux(登録商標)ベースのサーバアプリケーション、任意のプロプライエタリサーバアプリケーションを含むサーバアプリケーションのバージョンのうちのいずれか、またはコンピューティングデバイス200において動作することができ本明細書に説明の動作を行うことができる他の任意のサーバアプリケーションなど、任意のサーバアプリケーション216を実行することができる。コンピューティングデバイスにおいて動作することができるサーバアプリケーションの一例には、Apacheサーバアプリケーションが挙げられる。

【0049】

図3は、本開示の実施形態によるインターネットまたは他のネットワーク化環境上の不正コンテンツの検出および監視を容易にするための一例示的なネットワーク化環境300である。環境300は、任意のネットワークとすることができる通信ネットワーク340を介して、1つまたは複数の(ローカル)サーバ321~323を含むリモートコンピューティングシステム320に動作可能に接続されたユーザコンピューティングデバイス310~312を含み、このネットワーク340を介して、情報がネットワークに通信可能なように接続されたデバイス間で伝送され得る。たとえば、通信ネットワーク340は、インターネット、イントラネット、仮想プライベートネットワーク(VPN)、広域ネットワーク(WAN)、およびローカルエリアネットワーク(LAN)などとすることができる。環境300は、通信ネットワーク340を介してサーバ321~323ならびにユーザコンピューティング310~312に動作可能に接続され得るレポジトリまたはデータベース330を含むことができる。データベース330がサーバ321~323のうちの1つまたは複数に組み込み可能であり、それによりこれらのサーバのうちの1つまたは複数がデータベースを含むことができることを当業者は認識するであろう。一例示的な実施形態では、エンジン100の実施形態は、サーバ321~323のうちの1つまたは複数によって単独でまたは共同で実施されても、ユーザコンピューティングデバイスのうちの1つまたは複数(たとえば、ユーザコンピューティングデバイス312)によつ

10

20

30

40

50

て実施されても、および/またはサーバ321～323とユーザコンピューティングデバイスとの間に分散されてもよい。

【0050】

ユーザコンピューティングデバイス310～312は、サーバ321～323のうちの1つまたは複数によって実施されるエンジン100との対話を容易にするようにユーザによって動作可能である。例示的な実施形態では、ユーザコンピューティングデバイス(たとえば、ユーザコンピューティングデバイス310～311)は、サーバ321～323のうちの1つまたは複数と対話するようにプログラミングおよび/または構成されているクライアント側アプリケーション315を含むことができる。1つの実施形態では、ユーザコンピューティングデバイス310～311によって実施されるクライアント側アプリケーション315は、エンジン100のGUIをホストする1つまたは複数のウェブページにナビゲートすることができるウェブブラウザとすることができる。いくつかの実施形態では、ユーザコンピューティングデバイス310～311のうちの1つまたは複数によって実施されるクライアント側アプリケーション315は、1つまたは複数のサーバ(たとえば、サーバ321、322、および/または323と対話するためのユーザインターフェースを可能にするアプリケーション)によって実施されるエンジン100との対話を可能にするためのエンジン100に固有のアプリケーションとすることができる。

10

【0051】

1つまたは複数のサーバ321～323(および/またはユーザコンピューティングデバイス312)は、通信ネットワーク340を介して利用可能なコンテンツについて検索するようにエンジン100を実行することができる。たとえば、エンジン100は、データソース350、360、および370の検索を容易にするようにプログラミングされ得、これらのデータソース350、360、および370はそれぞれ、コンテンツをホストし通信ネットワーク340を介してそのコンテンツを利用可能にするようにプログラミングされている1つまたは複数の(リモート)サーバ380を含むことができる。一非限定例としては、サーバ380は、エンジン100によって生成された1つまたは複数のクエリを使用して、1つまたは複数の検索エンジンおよび/またはAPIを介して検索され得るウェブサイトホストするように構成されているウェブサーバとすることができる。たとえば、データソース350、360、および/または370のうちの少なくとも1つは、オンラインマーケットプレイスウェブサイトを提供することができる。

20

30

【0052】

データベース330は、エンジン100によって使用される情報を記憶することができる。たとえば、データベース330は、本明細書に説明するように、クエリ、エンジン100による抽出された商品識別子データセット、エンジン100に関連しているタグ、および/またはエンジン100の実施形態によって使用され得る適切な他の任意の情報/データを記憶することができる。さらには、データベース330は、収集されたデータセット(すなわち、収集されたデータセット117)を記憶し、および/または商品識別子データベース(すなわち、商品識別子データベース135)を含むことができる。

【0053】

図4は、本開示の実施形態により実施される不正コンテンツ検出エンジンを使用して商品識別子を構文解析し、カテゴリ化するための一例示的方法400である。動作402においては、不正コンテンツ検出エンジン(すなわち、エンジン100)は、1つまたは複数のデータソースにおいてコンテンツ検索を行う。動作403においては、不正コンテンツ検出エンジンは、コンテンツ検索に回答して、複数の検索結果を返す。簡単にするために、図4は、単に結果404、結果405、および結果406の3つの返された結果のみを示している。動作407においては、不正コンテンツ検出エンジンは、結果404、結果405、および結果406から商品識別子を抽出または構文解析する。

40

【0054】

動作408においては、結果404、結果405、および結果406は、レコード409としてデータベースに記憶される。各商品識別子は、その商品識別子カテゴリに従って、適切なデ

50

ータフィールドの中に入れられる。たとえば、不正コンテンツ検出エンジンは、結果404からブランド名、製造業者部品番号(MPN:manufacturer part number)、および万国製品コード(UPC)を、ならびに結果405からブランド名およびMPNを構文解析する。ブランド名は、ブランド名データフィールド420の中に入れられ、MPNは、MPNデータフィールド422の中に入れられ、UPCは、UPCデータフィールド424の中に入れられる。不正コンテンツ検出エンジンは、結果406について、いずれの商品識別子のカテゴリ化もできていない。

【0055】

動作412においては、不正コンテンツ検出エンジンは、結果406に、アナリストによってレビューされるとタグ付けする。動作414においては、アナリストは、結果406を分析して、未知の商品識別子をカテゴリ化する。動作416においては、アナリストは、商品識別子として欧州統一商品番号(EAN)を識別し、カテゴリ化する。動作416においては、不正コンテンツ検出エンジンは、EANをレコード406に関連しているEANデータフィールド426に入れる。商品識別子としてのEANは、今後の抽出に向けて、既知の商品識別子とともに記憶される。次いで、結果は、本明細書に説明しているように、タグ付けエンジンに送信されてレビューされる。

10

【0056】

図5は、本開示の実施形態による不正コンテンツ検出エンジン(すなわち、不正コンテンツ検出エンジン100)の一実装形態を実施するコンピューティングシステムにおける1つまたは複数のローカルサーバを介して実施される一例示的処理1100を示すフローチャートである。ステップ1102においては、ネットワーク化環境における1つまたは複数のリモートサーバによってホストされたコンテンツ(たとえば、ウェブページ)が、たとえば、1つまたは複数のローカルサーバによって実施される検索により収集される。検索は、GTINおよび/または1つまたは複数の検索語に基づいてよい。たとえば、1つまたは複数のローカルサーバは、1つまたは複数のリモートサーバによってホストされたコンテンツを検索するために、検索エンジンまたはアプリケーションプログラミングインターフェースのうちの少なくとも1つへの入力としてGTINを使用することができる。あるいは、エンジン100の一実装形態を実行する1つまたは複数のローカルサーバは、1つまたは複数の検索語に基づいて、1つまたは複数のクエリを生成することができ、1つまたは複数のクエリは、1つまたは複数のリモートサーバによってホストされたコンテンツを検索するために、検索エンジンまたはアプリケーションプログラミングインターフェースのうちの少なくとも1つへの入力を形成することができる。検索からの結果が、1つまたは複数のローカルサーバによってフェッチされ、記憶デバイスにダウンロードされる。例示的な実施形態では、1つまたは複数のリモートサーバは、ウェブサーバとすることができ、1つまたは複数のサーバによってホストされたコンテンツは、ウェブページを含むウェブサイトを含むことができる。

20

30

【0057】

ステップ1104においては、1つまたは複数のローカルサーバは、1つまたは複数のリモートサーバによってホストされたコンテンツの検索に回答して返された結果の組における各結果(たとえば、ウェブページ)から商品識別子を抽出する。一例示的な実施形態では、抽出された商品識別子は、GTINおよびブランド名を含む。ステップ1105においては、1つまたは複数のローカルサーバは、結果の組からの商品識別子の抽出に回答して、レコードを結果(たとえば、ウェブページ)ごとにデータベースに作成する。各結果から抽出された商品識別子は、それらのそれぞれのレコードに関連しているデータフィールドの中に記憶される。

40

【0058】

ステップ1106においては、1つまたは複数のローカルサーバは、抽出されたGTINが、抽出されたブランド名に対して合法であるかどうかに基づいて、各レコードに合法または不正とタグ付けする。1つまたは複数のローカルサーバが特定のGTINについて検索した事例では、1つまたは複数のローカルサーバは、特定のGTINが、抽出されたブランド名に対

50

して合法であるかどうかに基づいて、各レコードに合法または不正とタグ付けする。

【0059】

ステップ1108においては、1つまたは複数のローカルサーバは、各レコードに係る抽出および構文解析された商品識別子を分析して、不正製品をさらに識別する。一例示の実施形態では、1つまたは複数のローカルサーバは、1つまたは複数の商品識別子が、抽出されたブランド名に関連しているあらかじめ定義された商品識別子に対応しているかどうかを決定する。追加の実施形態では、1つまたは複数のローカルサーバは、不十分なまたは正しくない商品識別子について、商品識別子を分析する。ステップ1110においては、1つまたは複数のローカルサーバは、任意の識別された不正製品または偽造製品の除去を開始する。

10

【0060】

例示的なフローチャートは、本明細書に例示を目的として提供されており、方法の非限定例である。例示的な方法が、例示的なフローチャートに示されているものよりも多くのステップを含んでも、またはより少ないステップを含んでもよく、また例示的なフローチャートにおけるステップを例示のフローチャートに示されている順序とは異なる順序で行ってもよいことを当業者は認識するであろう。

【0061】

本明細書に開示の主題の特定の実施形態の前述の説明は、例示および説明の目的で提示されたものであり、本明細書に記載の主題の範囲を限定するように意図するものではない。他の様々な実施形態、修正形態、および適用例が、前述の説明および添付の図面から当業者には明らかになることが完全に企図される。したがって、他のそのような実施形態、修正形態、および適用例が、付随の特許請求の範囲内に入ることが意図される。さらには、本明細書に説明した実施形態、修正形態、および適用例は、特定の環境の文脈にあり、本明細書に記載の主題は、それらに限定するものではなく、任意の数の他の形、環境、および目的で有益に適用され得ることを当業者であれば認識するであろう。したがって、添付の特許請求の範囲は、本明細書に開示の新奇な特徴および技法の全幅および趣旨を考慮して解釈すべきである。

20

【符号の説明】

【0062】

- 100 不正コンテンツ検出エンジン
- 102 データソース
- 104 ユーザ
- 110 ユーザインターフェース
- 112 提示/可視化エンジン
- 114 グラフィカルユーザインターフェース
- 115 収集エンジン
- 117 データセット
- 119 除去エンジン
- 120 抽出エンジン
- 125 タグ付けエンジン
- 130 分析エンジン
- 135 商品識別子データベース
- 200 コンピューティングデバイス
- 202 プロセッサ、処理デバイス
- 202' プロセッサ
- 204 コア
- 204' コア
- 206 メモリ
- 212 ネットワークインターフェース
- 214 仮想マシン

30

40

50

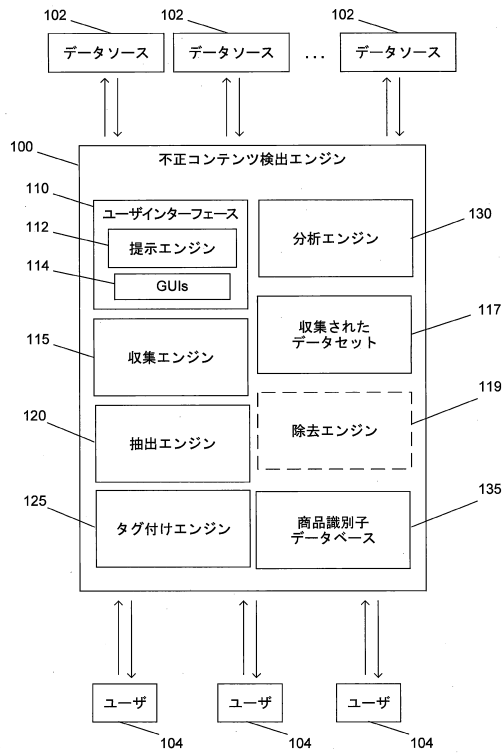
- 216 サーバアプリケーション
- 222 ネットワークデバイス
- 224 記憶デバイス
- 300 ネットワーク化環境
- 310、311、312 ユーザコンピューティングデバイス
- 315 クライアント側アプリケーション
- 320 リモートコンピューティングシステム
- 321、322、323 サーバ
- 330 レポジトリまたはデータベース
- 340 通信ネットワーク
- 350、360、370 データソース
- 380 サーバ
- 400 方法
- 404、405、406 結果
- 409 レコード
- 420 ブランド名データフィールド
- 422 MPNデータフィールド
- 424 UPCデータフィールド
- 426 EANデータフィールド
- 1100 方法

10

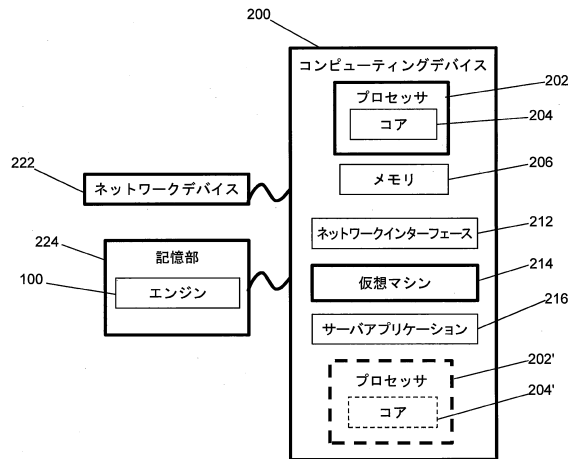
20

【図面】

【図 1】



【図 2】

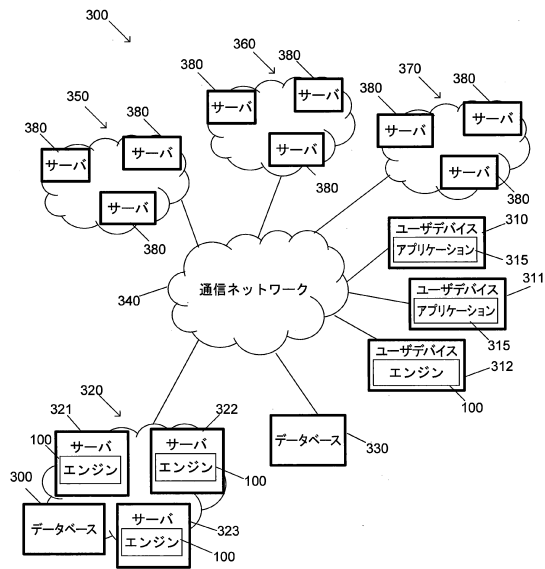


30

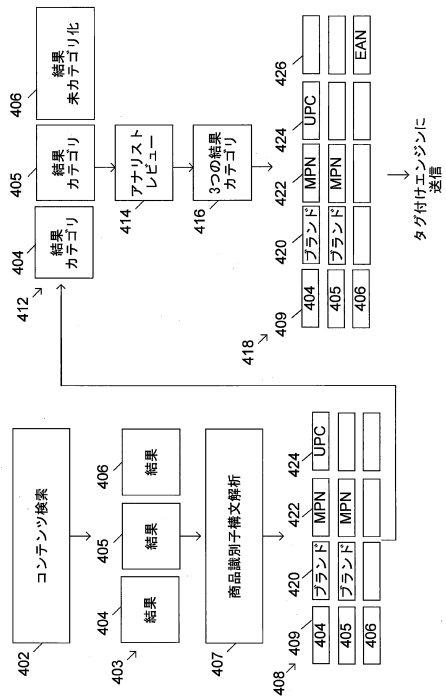
40

50

【図3】



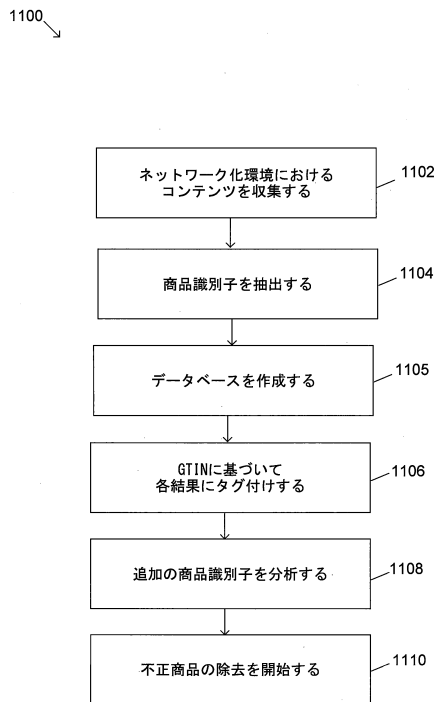
【図4】



10

20

【図5】



30

40

50

