

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5655056号
(P5655056)

(45) 発行日 平成27年1月14日(2015.1.14)

(24) 登録日 平成26年11月28日(2014.11.28)

(51) Int.Cl. F I
G O 6 F 17/21 (2006.01) G O 6 F 17/21 5 7 0 U

請求項の数 14 (全 17 頁)

(21) 出願番号	特願2012-502130 (P2012-502130)	(73) 特許権者	500046438
(86) (22) 出願日	平成22年3月19日 (2010. 3. 19)		マイクロソフト コーポレーション
(65) 公表番号	特表2012-521608 (P2012-521608A)		アメリカ合衆国 ワシントン州 9805
(43) 公表日	平成24年9月13日 (2012. 9. 13)		2-6399 レッドモンド ワン マイ
(86) 国際出願番号	PCT/US2010/028026		クロソフト ウェイ
(87) 国際公開番号	W02010/111146	(74) 代理人	100140109
(87) 国際公開日	平成22年9月30日 (2010. 9. 30)		弁理士 小野 新次郎
審査請求日	平成25年3月12日 (2013. 3. 12)	(74) 代理人	100075270
(31) 優先権主張番号	12/410, 477		弁理士 小林 泰
(32) 優先日	平成21年3月25日 (2009. 3. 25)	(74) 代理人	100101373
(33) 優先権主張国	米国 (US)		弁理士 竹内 茂雄
		(74) 代理人	100118902
			弁理士 山本 修
		(74) 代理人	100153028
			弁理士 上田 忠

最終頁に続く

(54) 【発明の名称】 辞書を分配するためのプログラム

(57) 【特許請求の範囲】

【請求項 1】

コンピュータ実装辞書システムであって、

辞書を異なる複数のアプリケーションによって使用可能なユニバーサルフォーマットにフォーマットするためのフォーマットコンポーネントと、

該フォーマット済みの辞書についてのコンテンツをネットワーク上の他の利用可能な複数の辞書のソースから引き込む辞書アーキテクチャと、

前記辞書がアクセスされたことに応答して、前記異なる複数のアプリケーションの各々に特化して使用できるように、前記引き込んだフォーマット済み辞書をコンパイルするためのコンパイラコンポーネントと、

前記コンパイルしたフォーマット済みの辞書のフォーマットをサポートする前記異なる複数のアプリケーションの内、互換性のあるアプリケーションの各々により前記コンパイルしたフォーマット済みの辞書を共有する分配コンポーネントとを備え、

前記異なる複数のアプリケーションが、スピーチディクテーションアプリケーション、スペルチェックアプリケーション、言語分析アプリケーション、およびインプットメソッドアプリケーションを含むことを特徴とする、システム。

【請求項 2】

前記辞書が異なるアプリケーションによってアクセスできるように中央に配置されていることを特徴とする請求項 1 に記載のシステム。

【請求項 3】

前記ユニバーサルフォーマットはXMLであることを特徴とする請求項1に記載のシステム。

【請求項4】

請求項1記載のシステムであって、更に、テキストデータ、言語データ、又はディクテーションデータの内の少なくとも1つを含むように前記フォーマット済みの辞書を編纂するためのオーサリングツールを備えることを特徴とする、システム。

【請求項5】

請求項1記載のシステムにおいて、前記コンパイラコンポーネントが、前記言語分析アプリケーションによって使用できるように前記フォーマット済みの辞書をコンパイルした後に、該コンパイルしたフォーマット済みの辞書がダウンロードされ、前記言語分析アプリケーションで使用できるように該言語分析アプリケーションに登録されることを特徴とする、システム。

10

【請求項6】

請求項1記載のシステムにおいて、前記コンパイラコンポーネントが前記スペルチェックアプリケーションによって使用できるように前記フォーマット済みの辞書をコンパイルした後に、該コンパイルしたフォーマット済みの辞書がダウンロードされ、前記スペルチェックアプリケーションで使用できるように該スペルチェックアプリケーションに登録されることを特徴とする、システム。

【請求項7】

請求項1記載のシステムにおいて、前記コンパイラコンポーネントが前記スピーチディクテーションアプリケーションによって使用できるように前記フォーマット済みの辞書をコンパイルした後に、該コンパイルしたフォーマット済みの辞書がダウンロードされ、前記スピーチディクテーションアプリケーションで使用できるように該スピーチディクテーションアプリケーションに登録されることを特徴とする、システム。

20

【請求項8】

請求項1記載のシステムにおいて、前記コンパイルしたフォーマット済みの辞書がダウンロードされ、アプリケーションで使用できるように該アプリケーションに登録され、該アプリケーションが前記インプットメソッドアプリケーションであることを特徴とする、システム。

【請求項9】

請求項1記載のシステムにおいて、前記コンパイルしたフォーマット済みの辞書が、ウェブサイト、電子メッセージ又はファイル共有アプリケーションの内の少なくとも1つを介して共有されることを特徴とする、システム。

30

【請求項10】

コンピュータ実装辞書システムであって、
辞書を異なる複数のアプリケーションによって使用可能なユニバーサルフォーマットにフォーマットするためのフォーマットコンポーネントと、

テキストデータ、言語データ、又はディクテーションデータの内の少なくとも1つを含むように前記フォーマット済みの辞書を編纂するためのオーサリングツールと、

前記フォーマット済みの辞書についてのコンテンツをネットワーク上の他の利用可能な複数の辞書のソースから引き込む辞書アーキテクチャと、

40

前記辞書への要求に回答して、前記異なる複数のアプリケーションの各々に特化して使用できるように、前記引き込んだフォーマット済みの辞書をコンパイルし、また、前記引き込んだフォーマット済みの辞書が予め定めた基準を満足することを確実にするために、地政学、正確さおよびパフォーマンスのセットの内少なくとも1つに関するチェックを実行する、コンパイラコンポーネントと、

前記コンパイルしたフォーマット済みの辞書を、前記異なる複数のアプリケーションの内少なくとも1つのアプリケーションにより使用される少なくとも1つの他のコンパイルしたフォーマット済みの辞書と同期するための分配コンポーネントとを備え、

前記異なる複数のアプリケーションが、スピーチディクテーションアプリケーション、

50

スペルチェックアプリケーション、言語分析アプリケーション、およびインプットメソッドアプリケーションを含み、前記コンパイラコンポーネントが、前記言語分析アプリケーション、前記スペルチェックアプリケーション、および前記スピーチディクテーションアプリケーションによって使用できるように前記フォーマット済みの辞書をコンパイルすることを特徴とする、システム。

【請求項 1 1】

請求項 1 0 記載のシステムにおいて、前記フォーマットコンポーネント、オーサリングツール、およびコンパイラコンポーネントが前記異なる複数のアプリケーションによってアクセスできるように中央に配置されることを特徴とする、システム。

【請求項 1 2】

請求項 1 0 記載のシステムにおいて、前記コンパイルしたフォーマット済みの辞書が、前記言語分析アプリケーション、スペルチェックアプリケーション、またはスピーチディクテーションアプリケーションの内少なくとも 1 つによって使用できるようにダウンロードされることを特徴とする、システム。

【請求項 1 3】

前記コンパイルしたフォーマット済みの辞書がダウンロードされ、クライアントアプリケーションでの使用のために該クライアントアプリケーションに登録される請求項 1 0 に記載のシステム。

【請求項 1 4】

請求項 1 0 記載のシステムにおいて、前記コンパイルしたフォーマット済みの辞書が、ウェブサイト、電子メッセージ、又はファイル共有アプリケーションの内少なくとも 1 つを通じて分配および共有されることを特徴とする、システム。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本発明は、アプリケーションのための共有可能な分散型辞書に関する。

【背景技術】

【0 0 0 2】

電子辞書は、IME (input method editor: インプットメソッドエディタ)、スペルチェッカ、手書き認識装置、スピーチディクテーション等のような言語分析を行うアプリケーションのための不可欠な要素であり得る。しかしながら、自然言語とソフトウェアの性質が進化するのに伴い、全ての必要な単語をアプリケーションに含ませることは不可能となる。なぜなら、アプリケーションの性能・パフォーマンスは退化し、製品リリース後に新しい単語が現れ、ある分野に属さないユーザにとって当該分野に特有な用語は正確さが低下する等の理由があるからである。

【発明の概要】

【発明が解決しようとする課題】

【0 0 0 3】

この問題に対処する共通のアプローチ (対応策) は、例えばアドオン辞書のように分野に特有 (特定) 用語を提供すること、又は個人的利用のためのカスタム辞書を作成する機能を提供することがある。現在のアドオン辞書及びカスタム辞書が有する問題は、特定のアプリケーションのための専用 (プロプライエタリ) フォーマットを使用していることである。使用される用語は複数のアプリケーションに亘ってユーザまたはユーザグループにとって共通である (広く使用される) にもかかわらず、上記プロプライエタリフォーマットは他のアプリケーションと共有することが非常に困難であるか不可能である。

【0 0 0 4】

アドオン辞書の他の問題は、辞書のグラニュラリティ (粒度、granularity) である。市場では、辞書は多数のユーザを有する大きなグラニュラリティ分野のために提供されている。なぜなら小さい分野のための辞書は商品化することが困難であるからである。

10

20

30

40

50

【課題を解決するための手段】

【0005】

以下の説明は、本明細書に記載されているいくつかの新規な実施形態の基礎的な理解を提供するための概要である。この「発明の概要」は完璧な概要ではなく、且つ、主要な/重要な(必須の)要素を特定すること若しくは当該要素の範囲を限定することを意図していない。この「発明の概要」の唯一の目的は、後に示される詳細な説明の前置きとして、簡易化された形式でのいくつかの技術思想を示すことである。

【0006】

開示されたアーキテクチャ・構成は、例えばXMLのようなユニバーサルフォーマットの電子辞書を提供する。そのサポートとして、ユーザ環境には、辞書オーサリング(authoring)ツールと共にオンザフライ辞書コンパイラもインストールされる。このモデルによれば、特定の分野及び特定のアプリケーションのために一人のユーザによって作成された辞書は、他のアプリケーションによって容易に再利用され、同じ分野(ドメイン)に属する複数のユーザの間で共有される。さらに、辞書ファイルは当該辞書ファイルをウェブサーバ上に格納すること及び電子メールに添付すること等により、分散することができる。

10

【0007】

上記目的及び関連する目的を達成するために、本明細書には下記の説明及び添付図面に関連していくつかの例示的形態・特徴が記載されている。これら形態・特徴は、本明細書に開示された原理・技術思想が実施される様々な手段・手法を示すものであり、かつ全ての形態・特徴及びこれらと均等のもは特許請求の範囲に記載された発明の範囲内のものである。他の利点・効果及び新規な特徴は図面を考慮したとき下記の詳細な説明から明らかになるだろう。

20

【図面の簡単な説明】

【0008】

【図1】開示されたアーキテクチャ・構成に基づくコンピュータに実装された辞書システムを示す図である。

【図2】図1のシステムにオーサリングツールを追加した別の実施形態のシステムを示す図である。

【図3】異なるアプリケーションにより異なる用途で辞書を使用するためのフロー図である。

30

【図4】ユニバーサルフォーマットに基づいて辞書を共有及び分散するための中央に配置されたサーバシステムを示す図である。

【図5】アプリケーションで使用するために辞書を処理する方法を示す図である。

【図6】図5の方法に加えられる形態・特徴を示す図である。

【図7】特定の目的のために辞書を処理する方法を示す図である。

【図8】分野特定辞書を提供する方法を示す図である。

【図9】開示されたアーキテクチャ・構成に基づき辞書のコンパイル及び共有を実行できるコンピューティングシステムのブロック図である。

【図10】辞書を分散させるためのコンピューティング環境の概略ブロック図である。

40

【発明を実施するための形態】

【0009】

開示されたアーキテクチャ・構成は、例えばXML(extensible markup language)のようなユニバーサルフォーマットの辞書を提供し処理する。前記辞書は、前記ユニバーサルフォーマットで書かれているが、複数の互換性のあるアプリケーションと共に使用するために用いられ、かつ辞書コンパイラを使用してオンザフライ方式でコンパイルされることが可能である。前記辞書は、ウェブサーバ、電子メール、並びに他の適切なデータ伝達技術・手段を用いて共有及び/若しくは分散されることが可能である。一旦クライアントアプリケーションへダウンロードされると、前記辞書は要求元のクライアントアプリケーションへ登録され使用可能となる。このモデルを用いると、

50

特定の分野及び特定のアプリケーションのためにユーザによって作成された辞書は、他のアプリケーションによって容易に再利用されると共に、同じ分野に属する複数のユーザの間で容易に共有される。

【 0 0 1 0 】

なお、クライアントシステムは、辞書のコンパイルを必要としないが、ユニバーサルフォーマットであれば辞書を使用することができる。また、移送方法はコンピュータファイルを移送するのに使用される任意の方法であってよい。さらに、ファイルを別のマシンに移送することも必要としない。上記辞書は複数のアプリケーションに亘る環境において共有可能である。

【 0 0 1 1 】

さらに、開示されたアーキテクチャ・構成は、テキスト検索、ワードブレッキング（文章を単語に分解すること）、テキストマイニング（テキストから質の高い情報を導出すること）、テキストを音声に変換する処理、機械翻訳（1つの言語から他の言語へテキスト又は音声を変換すること）、及び人間の読書（読取り）のための辞書のために用いられることも可能である。

【 0 0 1 2 】

以下、添付図面を参照する。全体を通して同様な参照番号は同様な要素を示すのに使用される。下記の記載において、説明のために多くの具体例・詳細がその十分な理解のために示されている。しかしながら、これらの特定の詳細な設定・数値等がなくても新規な実施形態が実施できることは明らかである。他の例・箇所では、周知の構造及び周知の装置がその説明を容易にするためにブロック図の形式で示されている。その意図は、特許請求の範囲に記載された発明の精神及び範囲内の全ての変形物、均等物及び代替物をカバー（包含）することである。

【 0 0 1 3 】

図1は開示されたアーキテクチャ・構成に基づくコンピュータに実装された辞書システム100を示している。前記システム100は、ユニバーサルフォーマットコンポーネント106によって支援されたユニバーサルフォーマットで辞書104（例えば電子辞書）をフォーマット化するためのフォーマットコンポーネント102を含み、ユニバーサルフォーマットの辞書108は異なるアプリケーションにより使用可能である。前記システム100はまた、辞書104へのアクセスに応じて、前記フォーマット化された辞書108をアプリケーション114で使用するようにするために、当該辞書をコンパイルしてコンパイルされた辞書112にするためのコンパイラコンポーネント110を含む。前記ユニバーサル（汎用）フォーマットは例えばXMLである。

【 0 0 1 4 】

前記辞書104は異なるアプリケーションによるアクセスを可能にするために中央に配置することができる（例えばサーバ）。前記コンパイラコンポーネント110は複数の異なるアプリケーションで使用できるようにするために辞書104の適切なコンパイルを行う。換言すれば、前記コンパイラコンポーネント110は複数の異なるコンパイル用の複数の異なるコンパイラを表している。例えば、前記コンパイラコンポーネント110は言語分析アプリケーションによって使用するために、前記フォーマット化された辞書108をコンパイルするコンパイラを含む。前記コンパイラコンポーネント110はまた、スペルチェックアプリケーションによって使用するために、前記フォーマット化された辞書108をコンパイルするコンパイラを含む。前記コンパイラコンポーネント110はさらに、スピーチディクテーションシステムによって使用するために、前記フォーマット化された辞書108をコンパイルするコンパイラを含む。その他のコンパイラが前記コンパイラコンポーネント110の一部として含まれていてもよい。

【 0 0 1 5 】

前記フォーマット化されコンパイルされた辞書112はダウンロードされ、前記アプリケーション114と共に使用するために前記アプリケーション114へ登録することができる。前記フォーマット化されコンパイルされた辞書112はまた、例えばウェブサイト

10

20

30

40

50

、電子メッセージ又はファイル共有アプリケーションの内の少なくとも1つを用いて共有される。しかしながら、前記アプリケーション114は前記コンパイルされた形式(前記コンパイルされた辞書112)ではなく前記フォーマット化された(ユニバーサルフォーマットの)辞書108を直接利用することができる。

【0016】

図2は図1のシステム100にオーサリングツール202を加えた別の実施形態によるシステム200を示している。前記システム200は、前記ユニバーサルフォーマットコンポーネント106によって支援されたユニバーサルフォーマットで前記辞書104をフォーマット化するための前記フォーマットコンポーネント102を含む。前記フォーマット化された辞書108は異なるアプリケーションによって使用可能である。当該アプリケーションは例えばクライアントアプリケーション及び/又はサーバアプリケーションである。前記システム200はまた、前記辞書104へのアクセス(又は要求)に応じて、前記フォーマット化された辞書108をコンパイルされた辞書112にコンパイルすることにより、前記アプリケーション114の内部で又は前記アプリケーション114と共に使用できるようにするコンパイラコンポーネント110を含む。

10

【0017】

前記オーサリングツール202は、例えばテキストデータ、言語データ、手書きデータ又はディクテーションデータの内の少なくとも1つを含むように前記フォーマット化された辞書108を編纂するのに用いられる。辞書の著者は前記オーサリングツール202を使用して前記ユニバーサルフォーマットの辞書108(例えばXMLフォーマット)を編纂し、又はスクリプト若しくは他の適切なプログラムを使用し辞書ソースを変換する。

20

【0018】

特定のアプリケーションのために作成された辞書は他のアプリケーションによって使用されるための属性を有していない可能性がある。従って、もし著者が辞書を他のアプリケーションで使用したい場合、当該不足する属性を追加することができる。分散型辞書は前記辞書をサポートする他のユーザによって使用されることができる。前記アプリケーション114による前記サポートは前記辞書のための適切なローダ(loader)を含む。

【0019】

さらに、辞書(前記辞書104)を使用するアプリケーション(前記アプリケーション114)は関連する辞書コンパイラ(前記コンパイラ110)を有する。通常、前記コンパイラ110はユーザ環境内ではなくプロダクション(製造)環境内に存在する。ソーステキストから辞書をコンパイルする基本的なメカニズムは十分知られているので、当該コンポーネントを作成する場合、前記アプリケーションのための既存の辞書コンパイラを再利用することにより、入力として前記ユニバーサルフォーマットの辞書108を受信する。前記受信又は指定されたアプリケーション114が前記コンパイルされた辞書112を登録するので、前記辞書は前記アプリケーション114によって使用可能となる。さらに、同じ辞書は当該フォーマットをサポートする互換性のある複数のアプリケーションの間で共有することが可能である。

30

【0020】

換言すれば、前記システム200は、異なるアプリケーションで使用可能なユニバーサルフォーマットの前記辞書104をフォーマット化する前記フォーマットコンポーネント102と、テキストデータ、言語データ若しくはディクテーションデータの内の少なくとも1つを含むように前記フォーマット化された辞書108を編纂する前記オーサリングツール202と、前記辞書104に対する要求に応じて前記フォーマット化された辞書108をコンパイルする前記コンパイラコンポーネント110とを含む、コンピュータに実装された辞書システムである。

40

【0021】

前記フォーマットコンポーネント102、オーサリングツール202及びコンパイラコンポーネント110は異なるアプリケーションによるアクセスができるように中央に配置されている。前記コンパイラコンポーネント110は、言語分析アプリケーション、スペ

50

ルチェックアプリケーション又はスピーチディクテーションシステムの内の少なくとも1つによって使用できるように前記フォーマット化された辞書108をコンパイルする。前記フォーマット化されコンパイルされた辞書112はダウンロードされ、クライアントアプリケーションと共に使用するために前記クライアントアプリケーションへ登録される。前記フォーマット化されコンパイルされた辞書112は例えば、ウェブサイト、電子メッセージ若しくはファイル共有アプリケーションの内の少なくとも1つを用いて分散及び共有される。

【0022】

図3は異なるアプリケーションにより異なる用途で使用される辞書のためのフロー図300を示している。符号302において、単語及び属性を含む辞書ソースが提供される。前記属性は、前記辞書が1つ又は複数のエンドユーザアプリケーションによって使用されるよう指定することができる。符号304において、前記辞書オーサリングツールは前記辞書ソースに変更を加えるために採用されている。符号306において、前記辞書はユニバーサルフォーマットを使用してフォーマット化される。符号308において、分散コンポーネントは前記フォーマット化された辞書を適切なアプリケーションへ分散できるようにしている。

【0023】

例えば、符号310において、インプットメソッドアプリケーション(例えばIME:インプットメソッドエディタ)からの要求に応じて前記分散コンポーネントは前記ユニバーサルフォーマットの辞書を前記インプットメソッドアプリケーションのための辞書コンパイラへ転送する。符号312において、前記インプットメソッドアプリケーションの辞書は前記インプットメソッドアプリケーションと共に使用する準備が整う。同様に、符号314において、スペルチェックアプリケーションからの要求に応じて前記分散コンポーネントは前記ユニバーサルフォーマットの辞書を前記スペルチェックアプリケーションのための辞書コンパイラへ転送する。符号316において、前記スペルチェック辞書は前記スペルチェックアプリケーションと共に使用する準備が整う。さらに、符号318において、スピーチディクテーションアプリケーションからの要求に応じて前記分散コンポーネントは前記ユニバーサルフォーマットの辞書を前記スピーチディクテーションアプリケーションのための辞書コンパイラへ転送する。符号320において、前記スピーチディクテーション辞書は前記スピーチディクテーションアプリケーションと共に使用する準備が整う。

【0024】

図4はユニバーサルフォーマットに基づき辞書を共有及び分散するために中央に配置されたサーバシステム400を示している。前記システム400は、前記辞書104をユニバーサルフォーマットでフォーマット化して、分散コンポーネント404を用いてクライアントシステム402に分散し共有するためのフォーマットコンポーネント102を含む。前記フォーマット化は前記ユニバーサルフォーマットコンポーネント106によって行われる。ファイルの同期は例えば同じ辞書ファイルを格納する複数のサーバの間で行われる。当該ファイルの共有は、多くの異なるクライアントシステム402のために作成及び利用される複数の異なる辞書ファイルのために行うことができる。

【0025】

前記分散コンポーネント404は、辞書104へのアクセス(又は要求)に応じて、前記フォーマット化された辞書108をコンパイルされた辞書へコンパイルするために、及び前記フォーマット化された辞書108を前記クライアントシステム402と共に使用するか前記クライアントシステム402内で使用するために、前記フォーマット化された辞書108を1つ若しくは複数の前記クライアントシステム402へ送信する。前記システム400はまた、例えば前記テキストデータ、言語データ、手書きデータ、ディクテーションデータ、及び/又は他の種類のデータを含むように前記フォーマット化された辞書108を編纂するための前記オーサリングツール202を含むことができる。前記サーバシステム400は例えばインターネット又はエンタープライズ(法人用ネットワーク、企業

10

20

30

40

50

ネットワーク)を介して前記辞書を分散するためのウェブサーバであってもよい。

【0026】

前記インプットメソッドアプリケーションにおいて、用語・語彙の範囲は正確さ及びユーザの経験の向上のために管理される。前記インプットメソッドアプリケーション辞書は、最高頻度の語(単語、用語、ワード)をカバーするように且つ低頻度語の悪影響を受けること無く適切に主要なシナリオに対処できるように構成することができる。

【0027】

ウェブ辞書はユーザ辞書のエリアのほとんどをカバーし、例えば前記分野辞書、新出単語辞書及びコラボレーション辞書も部分的にカバーする。

【0028】

前記開示された辞書アーキテクチャ・構成は主要なネットワーク(例えばインターネット)上で使用可能な辞書ソースからコンテンツを引き込む(取り入れる)ことができる。例えばアジア言語、自動車、印刷、金融、建築、略語、和製英語、翻訳、古語に関する用語・語彙は上記ネットワークから検索され、コンテンツを取得することができる。同様に辞書のコンテンツは繁体字中国語(CHT)、簡体字中国語(CHS)、複合語、古詩、法律、経済、医療等に関するソースから取得することが可能である。

【0029】

前記コンパイラはまた、中央の場所から前記ユニバーサルフォーマットの辞書を受信し、ローカルクライアント(単数または複数)のために前記辞書を自動でコンパイルするクライアントアプリケーションとして利用されることも可能である。

【0030】

辞書GUID(globally unique identifier)を利用することができる。辞書GUIDは辞書の識別子である。辞書が追加され、かつ前記辞書GUIDが既存の辞書の一覧(リスト)の1つに一致した場合、前記既存の辞書は前記追加された辞書に置き換えられる。

【0031】

ウェブ辞書XMLファイルは異なる検索クローラによって認識される特定のストリングを含むことができるため、たとえ前記辞書がウェブ辞書のポータル上になくてもユーザは前記ウェブ辞書を見つけることができる。ウェブ辞書のポータルは前記検索の検索結果ページへのリンクを含むことができる。ウェブ辞書のポータルは例えばオンラインスイート(suite)アプリケーションのシンプルなウェブページである。前記ポータルは利用可能なウェブ辞書の分類された一覧を保持している。

【0032】

コンテンツプロバイダは辞書を追加することを要求できる。地政学、正確さ及びパフォーマンス・性能に関する基本的なチェックは前記辞書が所定の基準を満たすことを保証するために行われる。ウェブ辞書は、最初に前記ポータル上に前記辞書をローディングすることなしに、サービスを始めることができる。さらに、前記ウェブ辞書はリンクの一覧であるため、あらゆるユーザは別のウェブ辞書のポータルを作成することができる。

【0033】

ユーザは辞書のプロパティの「更新」ボタンを選択することにより辞書を更新することができる。更新URLは前記辞書ファイルそのものを直接示すだけでなく前記辞書へのリンクを有し、前記辞書に関する説明を含む一般的なウェブページを示すこともできる。

【0034】

辞書はコード(を用いて)サインされてもよい。セキュリティ上の理由からコードサインされた辞書はたとえ前記辞書GUIDが同じであってもコードサインされていない辞書によって置き換えられることができない。

【0035】

辞書はパフォーマンス・性能又は変換精度に悪影響を及ぼす疑わしいコンテンツを含むことがある。ユーザには、プロパティから辞書を削除するオプションが与えられる。しかしながら、辞書内のエントリが問題かどうかを自動的に見分けることが困難である場合も

10

20

30

40

50

あるので、当該疑わしいエントリをブロックしてもよい。これは利点と信頼性との間のトレードオフである。

【0036】

前記辞書ファイルのためのデジタル署名はコンテンツが変更されていないことを保証するために提供してもよい。署名された辞書ソース及び署名されていない辞書ソースがサポートされる。前記辞書ソースが署名されている場合、前記署名が表示され、ユーザは前記辞書を使用するかどうかを促される。同様に、前記辞書ソースが署名されていないと、前記ユーザが前記辞書を使用するかどうかを選択するように要求するメッセージが表示される。

【0037】

前記デジタル署名は不正変更の防止に使用される。前記署名をチェックすることによってユーザは署名の名前から前記辞書の質を判定することができる。コンテンツプロバイダにとっては、前記署名は攻撃から逃れるため、及び(つまり)辞書コンテンツを変更されることによって会社の評判にダメージを与えられないようにするために使用される。

【0038】

ウェブサイトからのインプットメソッドアプリケーション辞書の登録の1つの例を以下に説明する。ユーザがウェブサイト上のウェブ辞書へのリンクを選択すると、ブラウザはダウンロード確認ダイアログを表示及び実行する。前記インプットメソッドアプリケーションはデジタル署名を確認(チェック)し、前記辞書が署名されていない場合は所定のインジケータを表示し、前記辞書が正しく署名されている場合は異なるインジケータを表示する。前記ユーザは前記署名されていない辞書の使用を継続するか使用しないかを選択することができる。

【0039】

本明細書に含まれている記載は開示されたアーキテクチャ・構成の新規な形態・特徴を実行するための典型的な手法を示す1組のフローチャートである。説明を容易にするために、例えばフローチャート又はフロー図の形式で示されている1つまたは複数の方法は本明細書において一連の動作として図示及び記載されているが、前記方法は記載・図示された動作の順序に限定されない。なぜなら、いくつかの動作は、本発明の教示に基づいて、本明細書に図示・記載されている順序とは異なる順序で及び/又は本明細書に図示・記載されていない他の動作と同時にも行われてもよいからである。当業者であれば例えば別の方法が状態図のように相関する一連の状態又は事象として表されることもあることは容易に理解できるであろう。さらに、新規な実装例として(新規な実施形態において)、前記方法に含まれている動作の全てが必要となるわけではない。

【0040】

図5はアプリケーション用の辞書の処理方法を示している。符号500において、辞書は異なるクライアントアプリケーションによって使用可能なユニバーサルフォーマットでフォーマット化される。符号502において、前記フォーマット化された辞書は前記辞書に対する要求に応じてアプリケーションで使用するためにコンパイルされる。符号504において、前記コンパイルされた辞書は前記アプリケーションによって使用されるためにダウンロードされる。

【0041】

図6は図5の方法に加えることができる形態・特徴を示している。符号600において、前記アプリケーションによって使用されるために前記ダウンロードされた辞書を登録する。符号602において、異なるアプリケーションによる前記辞書の使用のために、当該辞書とは異なるアプリケーションに関連した辞書に属性が追加される。符号604において、前記フォーマット化された辞書はテキストデータ、言語データ、手書きデータ又はディクテーションデータの内の少なくとも1つを含むように編纂される。符号606において、前記辞書は互換性のあるアプリケーションによるXMLフォーマットで分散されるために中央で更新される。符号608において、前記コンパイルされた辞書は、言語分析アプリケーション、スペルチェックアプリケーション又はスピーチディクテーションアプリ

10

20

30

40

50

ケーションの内の少なくとも1つによって使用されるためにダウンロードされる。

【0042】

図7は、特定の目的のための辞書の処理方法を示している。例えば、当該目的は言語分析又はディクテーションである。符号700において、辞書を作成するためにユニバーサルフォーマットが選択される。符号702において、前記辞書は前記特定の目的のために編纂される。符号704において、前記辞書はコンパイルされ、前記目的に関連するユーザへ分配される。分配は、例えば、前記コンパイルされた辞書の前記目的に関連するユーザへの直接送信、ウェブサイト上への掲載、及び/又は前記ユーザによるアクセスが可能な複数のサーバへの分散格納によってなされる。

【0043】

図8は、分野特定辞書を提供する方法を示している。符号800において、前記分野特定辞書はアジア言語分析のためにユニバーサルフォーマット(例えばXML)で編纂される。符号802において、互換性のあるクライアントアプリケーションから前記辞書に対する要求が受信される。符号804において、前記辞書はインプットメソッドアプリケーション(例えばIME)のために自動的にコンパイルされる。符号806において、前記辞書は前記インプットメソッドアプリケーションによって使用するためにエクスポート(expose)される。このエクスポージャ(exposure)は例えば追加(アドイン)モジュールの前記インプットメソッドアプリケーションへのダウンロードとしてであってもよい。これの代替例として、前記エクスポージャは前記インプットメソッドアプリケーションの既存のコンポーネントへのアップデート(更新)としてなされてもよい。符号808において、前記辞書は他の利用者・利用物によって使用するためにエクスポートされる。前記利用者・利用物は他のユーザの同様なインプットメソッドアプリケーション、及び/又は辞書に接続可能な互換性のあるアプリケーションであってもよい。

【0044】

本出願で使用されている用語「コンポーネント」及び「システム」は、コンピュータ関連のエンティティ、又はハードウェア、ハードウェアとソフトウェアの組合せ、ソフトウェア若しくは実行中のソフトウェアを意味することを意図している。例えばコンポーネントは(これらに限定されないが)、プロセッサ上で実行中のプロセス、プロセッサ、ハードディスクドライブ、複数のストレージドライブ(光学記憶媒体、半導体記憶媒体、及び/又は磁気記憶媒体)、オブジェクト、実行ファイル、実行中のスレッド、プログラム、及び/又はコンピュータである。例えば、サーバ上で実行中のアプリケーション及びサーバは共にコンポーネントであってもよい。1つ又は複数のコンポーネントは実行中のプロセス及び/又はスレッド内に存在してよく、コンポーネントは1つのコンピュータでローカライズされ、かつ/又は2つ以上のコンピュータの間で分散されることも可能である。用語「代表的な・例示的な」は例示的なもの、実例又は具体例であることを表すために本明細書で使用されている。「代表的な・例示的な」ものとして本明細書に記載されている全ての形態・特徴又は設計・構造は必ずしも他の形態・特徴又は設計・構造より好ましい又は有利であると解釈される必要はない。

【0045】

次に図9を参照する。図9は開示されたアーキテクチャ・構成に基づき辞書のコンパイル及び共有が可能なコンピューティングシステム900のブロック図を示している。様々な形態・特徴のためのさらなる説明を提供するために、図9及び以下の記載は前記様々な形態・特徴が実装される前記適切なコンピューティングシステム900の要約的で概略的な記載を提供するものである。上記の説明は1つ又は複数のコンピュータ上で実行可能なコンピュータ実行可能命令として説明されているが、当業者であれば新規な実施形態は他のプログラムモジュールとの組合せで、並びに/又はハードウェア及びソフトウェアの組合せとして実装可能であることを認識するだろう。

【0046】

様々な形態・特徴を実装するための前記コンピューティングシステム900は、プロセッシングユニット(単数又は複数)904、システムメモリ906、及びシステムバス9

10

20

30

40

50

08を有するコンピュータ902を含む。前記プロセッシングユニット(単数又は複数)904は、シングルプロセッサ、マルチプロセッサ、シングルコアユニット、及びマルチコアユニットのような様々な市販のプロセッサのいずれであってよい。さらに、当業者であれば、新規な方法は、ミニコンピュータ、汎用コンピュータ、又はパーソナルコンピュータ(例えばデスクトップ、ラップトップ等)、手持式のコンピューティングデバイス、マイクロプロセッサベースの家庭用電化製品又はプログラム可能な家庭用電化製品等を含む他のコンピュータシステム・装置と共に実行されてもよいことが理解できるであろう。当該他のコンピュータシステム・装置の各々は1つ又は複数の関連するデバイスに接続可能である。

【0047】

前記システムメモリ906は揮発性(VOL)メモリ910(例えばランダムアクセスメモリ(RAM))及び不揮発性メモリ(NON-VOL)912(例えばROM、EPROM、EEPROM等)を含む。基本的な入出力システム(BIOS)は不揮発性メモリ912へ格納され、例えば起動時における前記コンピュータ902内のコンポーネントの間のデータ及び信号の通信を支援する基本的なルーティンを含む。前記揮発性メモリ910はデータをキャッシュ保存するためのスタティックRAMのようなハイスピードRAMも含んでいてもよい。

【0048】

前記システムバス908はシステムコンポーネントのためのインターフェイスを提供する。前記システムコンポーネントはプロセッシングユニット(単数又は複数)904及び前記メモリサブシステム906を含む(これらに限定されない)。前記システムバス908は、様々な市販のバスアーキテクチャ・構成のいずれかを使用してメモリバス(メモリコントローラ付きの又はメモリコントローラなしの)、及び周辺機器用バス(例えばPCI、PCIe、AGP、LPC等)とさらに相互接続可能な数種類のバス構造のいずれかであってよい。

【0049】

前記コンピュータ902はさらに、ストレージサブシステム(単数又は複数)914と、当該ストレージサブシステム(単数又は複数)914をシステムバス908及び他の所望のコンピュータコンポーネントへ接続するためのストレージインターフェイス916(単数又は複数)とを含む。前記ストレージサブシステム(単数又は複数)914は例えばハードディスクドライブ(HDD)、磁気フロッピーディスクドライブ(FDD)、及び/又は光学ディスクストレージドライブ(例えばCD-ROMドライブ、DVDドライブ)のうちの1つ若しくは複数を含んでいてもよい。前記ストレージインターフェイス(単数又は複数)916は例えばIDE、ATA、SATA、及びIEEE1394規格のようなインターフェイス技術・手段を含むこともできる。

【0050】

1つ若しくは複数のプログラム及び1つ若しくは複数のデータは前記メモリサブシステム906、取り外し可能なメモリサブシステム918(例えばフラッシュドライブ形式のファクタ手段)、及び/又は前記ストレージサブシステム(単数又は複数)914(例えば光学、磁気、半導体)へ格納されてもよく、当該プログラム及びデータはオペレーティングシステム920、1つ若しくは複数のアプリケーションプログラム922、他のプログラムモジュール924及びプログラムデータ926を含む。

【0051】

前記コンピュータ902がサーバマシンである場合、1つ又は複数のアプリケーションプログラム922、他のプログラムモジュール924及びプログラムデータ926は例えば図1の前記フォーマットコンポーネント102、前記辞書104、前記ユニバーサルフォーマットコンポーネント106、前記フォーマット化された辞書108及び前記コンパイラコンポーネント110、図2のオーサリングツール202、図3のフロー図300で表現されたエンティティ、コンポーネント及びフロー、図4の前記サーバシステム400及び前記クライアントシステム402、並びに図5から図8のフローチャートで表現され

10

20

30

40

50

た前記方法を含んでいてもよい。

【 0 0 5 2 】

一般に、プログラムは、特定のタスクを実行するか特定の抽象データ型を実装するルーティン、方法、データ構造、他のソフトウェアコンポーネントを含む。オペレーティングシステム 9 2 0、アプリケーション 9 2 2、モジュール 9 2 4、及び/又はデータ 9 2 6の全部又は一部は例えば前記揮発性メモリ 9 1 0のようなメモリ内にキャッシュ保存されることも可能である。なお、前記開示されたアーキテクチャ・構成は様々な市販のオペレーティングシステム又はオペレーティングシステムの組合せ（例えば仮想マシンとして）と共に実装されてもよい。

【 0 0 5 3 】

前記ストレージサブシステム（単数又は複数） 9 1 4及びメモリサブシステム（ 9 0 6及び 9 1 8）はデータ、データ構造、コンピュータ実行可能命令等の揮発性ストレージ及び不揮発性ストレージのためのコンピュータ可読媒体として機能する。コンピュータ可読媒体は前記コンピュータ 9 0 2によってアクセス可能な任意の利用可能な・入手可能な媒体であってよく、揮発性媒体及び不揮発性媒体並びに取り外し可能な媒体及び固定の媒体を含む。前記コンピュータ 9 0 2の場合、前記媒体は任意の適切なデジタルフォーマットでデータの記憶が可能である。当業者であれば、他のタイプのコンピュータ可読媒体も使用可能であることは理解できるであろう。当該他のタイプのコンピュータ可読媒体とは、例えば、本明細書において開示されたアーキテクチャ・構成の新規な方法を実行するためのコンピュータ実行可能命令を格納するための Z I Pドライブ、磁気テープ、フラッシュメモリカード、カートリッジ等である。

【 0 0 5 4 】

ユーザはキーボード及びマウス等の外部のユーザ入力デバイス 9 2 8を用いて、前記コンピュータ 9 0 2、プログラム及びデータとインタラクトすることができる。他の外部のユーザ入力デバイス 9 2 8としては、マイクロフォン、 I R（赤外線）リモートコントローラ、ジョイスティック、ゲームパッド、カメラ認識システム、スタイラスペン、タッチスクリーン、ジェスチャーシステム（例えば目の動き、頭の動き等）、及び/又は同様のものが挙げられる。例えば前記コンピュータ 9 0 2が携帯型コンピュータである場合、前記ユーザはタッチパッド、マイクロフォン、キーボード等の内蔵のユーザ入力デバイス 9 3 0を使用して、前記コンピュータ 9 0 2、プログラム、及びデータにインタラクトすることができる。これら及び他の入力デバイスは前記システムバス 9 0 8を経由して、入出力（ I / O）デバイスインターフェイス（単数又は複数） 9 3 2を介して前記プロセッシングユニット（単数又は複数） 9 0 4と接続されているが、パラレルポート、 I E E E 1 3 9 4シリアルポート、ゲームポート、 U S Bポート、 I Rインターフェイス等の他のインターフェイスによって接続されることもできる。前記 I / Oデバイスインターフェイス（単数又は複数） 9 3 2は、プリンタ、オーディオデバイス、カメラデバイス等の出力周辺機器 9 3 4の使用も可能にし、例えばサウンドカード及び/又は内蔵のオーディオ処理手段等である。

【 0 0 5 5 】

1つ又は複数のグラフィックインターフェイス（単数又は複数） 9 3 6（一般的にグラフィックプロセッシングユニット（ G P U）とも呼ばれる）は前記コンピュータ 9 0 2と外部のディスプレイ（単数又は複数） 9 3 8（例えば L C D、プラズマ）及び/又は内蔵のディスプレイ（例えば携帯型コンピュータのための） 9 4 0との間でグラフィック信号及びビデオ信号を提供する。前記グラフィックインターフェイス（単数又は複数） 9 3 6はコンピュータシステムボードの一部として製造されこともある。

【 0 0 5 6 】

前記コンピュータ 9 0 2は1つ又は複数のネットワーク及び/又は他のコンピュータへの有線の/無線の通信サブシステム 9 4 2を介した論理接続を使用してネットワーク環境（例えば I P）内で動作することができる。前記他のコンピュータはワークステーション、サーバ、ルータ、パーソナルコンピュータ、マイクロプロセッサベースの娯楽用電化製

10

20

30

40

50

品、ピアデバイス又は他のコモンネットワークノードを含んでもよく、典型的な場合、前記コンピュータ 902 に関して記載された多くの又は全ての要素を含む。前記論理接続はローカルエリアネットワーク (LAN)、ワイドエリアネットワーク (WAN)、ホットスポット等への有線の / 無線の接続を含んでもよい。LAN ネットワーキング環境及び WAN ネットワーキング環境はオフィス内及び会社内ではありふれたものであり、イントラネットのような企業全体のコンピュータネットワークの構築を可能にし、全ての当該ネットワーク環境はインターネットのようなグローバル通信ネットワークへ接続される。

【0057】

前記コンピュータ 902 はネットワーク環境で使用されると、有線の / 無線のネットワーク、有線の / 無線のプリンタ、有線の / 無線の入力デバイス 944 等と通信するために有線の / 無線の通信サブシステム 942 (例えばネットワークインターフェイスアダプタ、内蔵のトランシーバサブシステム等) を介してネットワークに接続する。前記コンピュータ 902 はモデムを含むか、ネットワークを介して通信を確立するための他の手段を有することができる。ネットワーク環境において、前記コンピュータ 902 に関するプログラム及びデータは分散システムに関連付けられているので、リモートのメモリ / ストレージデバイスへ格納される。図示されたネットワーク接続は例示的なものであり、コンピュータ間の通信リンクを確立する他の手段が使用されてもよい。

10

【0058】

前記コンピュータ 902 は IEEE 802 . x x 標準ファミリのような無線通信技術・手段を使用して有線の / 無線のデバイス若しくは有線の / 無線のエンティティと通信が可能である。当該デバイスは例えばプリンタ、スキャナ、デスクトップコンピュータ及び / 若しくは携帯型コンピュータ、PDA (personal digital assistant)、通信衛星、無線検出可能なタグに関連付けられた装置若しくは店舗 (キオスク、新聞の売店、化粧室) 及び電話器と無線通信 (例えば IEEE 802 . 11 の無線変調技術) することができる無線のデバイスである。当該技術・手段は、少なくともホットスポットへのワイファイ (Wi-Fi: ワイヤレスフィデリティ) 無線技術手段、ワイマックス (WiMax) 無線技術・手段、及びブルートゥース無線技術・手段を含む。従って、前記通信は少なくとも従来のネットワーク又は少なくとも 2 つのデバイスを繋ぐアドホック通信と同様に、事前定義された構造であってもよい。ワイファイネットワークは安全で信頼性が高く高速な無線接続を提供する IEEE 802 . 11 x (a、b、g 等) と呼ばれる無線通信技術・手段を使用する。ワイファイネットワークはコンピュータ同士を接続したり、コンピュータをインターネットへ接続したり、コンピュータを有線ネットワーク (IEEE 802 . 3 に関連した媒体及び手段を使用する) へ接続するために使用される。

20

30

【0059】

次に図 10 を参照する。図 10 には辞書を分散させるためのコンピューティング環境 1000 の概略ブロック図が示されている。前記環境 1000 は 1 つ又は複数のクライアント 1002 を含む。前記クライアント (単数又は複数) 1002 はハードウェア及び / 又はソフトウェア (例えばスレッド、プロセス、コンピュータデバイス) であってもよい。前記クライアント (単数又は複数) 1002 は例えばクッキー (cookie) (単数又は複数) 及び / 又はこれに関連する文脈情報を格納する。

40

【0060】

前記環境 1000 は 1 つ又は複数のサーバ 1004 も含んでいる。前記サーバ (単数又は複数) 1004 はハードウェア及び / 又はソフトウェア (例えばスレッド、プロセス、コンピューティングデバイス) であってもよい。前記サーバ 1004 は例えば前記アーキテクチャ・構成を用いて変換処理 (transformations) を実行するスレッドを格納することができる。クライアント 1002 とサーバ 1004 の間で可能な 1 つの通信は 2 つ以上のコンピュータプロセスの間で送信できるデータパケットを用いて行われてもよい。前記データパケットは例えばクッキー及び / 又は関連する文脈情報を含んでいてもよい。前記環境 1000 は前記クライアント (単数または複数) 1002 と前記サーバ

50

(単数または複数) 1004の間の通信を支援するために用いられることができる通信フレームワーク1006(例えばインターネットのようなグローバル通信ネットワーク)を含む。

【0061】

通信は有線の(光ファイバーを含む)手段及び/又は無線の手段・技術を介して可能となる。前記クライアント(単数又は複数)1002は前記クライアント(単数又は複数)1002に対してローカルな情報(例えばクッキー(単数又は複数)及び/又は関連する文脈情報)を格納するために用いられることが可能な1つ又は複数のクライアントデータストア1008と接続されている。同様に前記サーバ1004は前記サーバ1004に対してローカルな情報を格納するために用いられることが可能な1つ又は複数のサーバデータストア1010と接続されている。

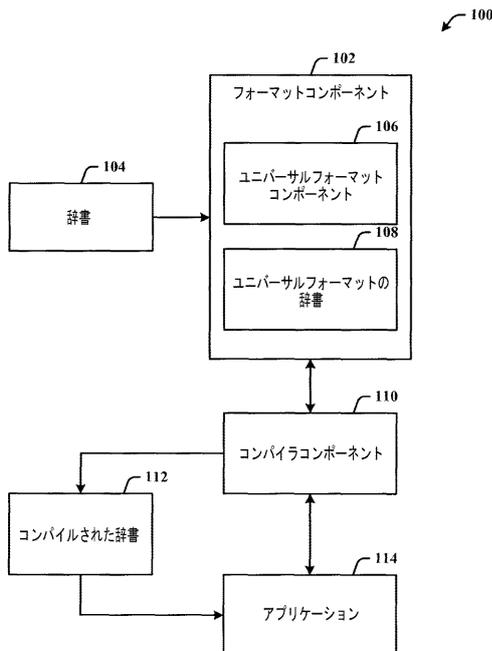
10

【0062】

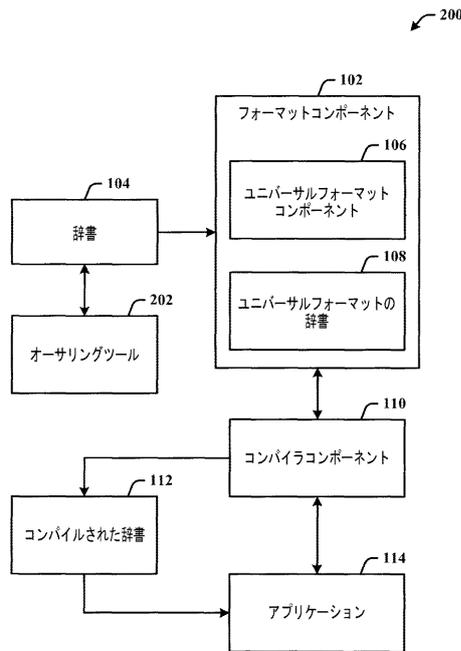
上記の説明はここに開示されたアーキテクチャ・構成の例示的なものを含んでいる。当然ながらコンポーネント及び/又は手法の考えられる全ての組合せを記載することは不可能であるが、当業者であれば多くのさらなる組合せ及び置換が可能であることは認識し得るであろう。従って、新規なアーキテクチャ・構成は特許請求の範囲の精神及び技術的範囲に属する全ての変更、改良、及び変形を包含することを意図している。さらに、本明細書又は特許請求の範囲に使用されている限りにおいて、用語「含む」は、請求項において使用された場合の(移行句としての)「から成る」、「含む」と同じように「他のものもさらに備えることができる」と解釈されることを意図している。

20

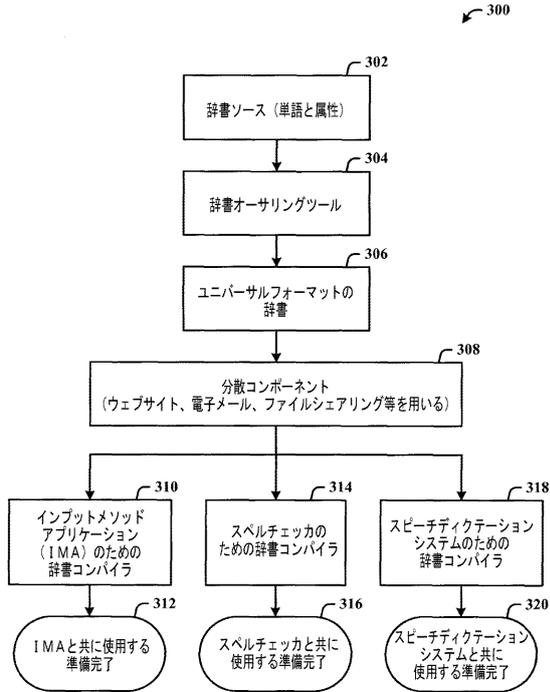
【図1】



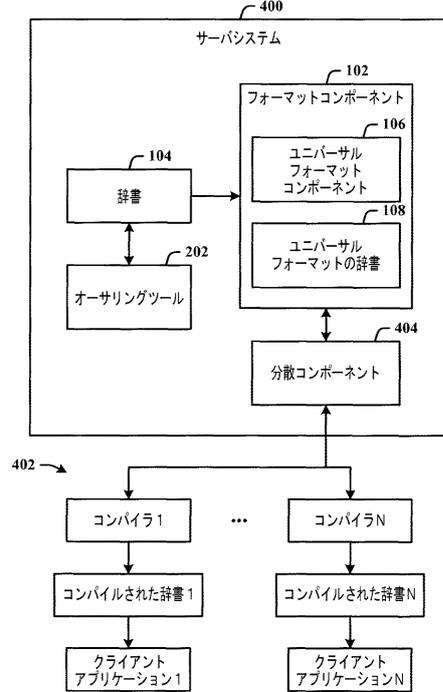
【図2】



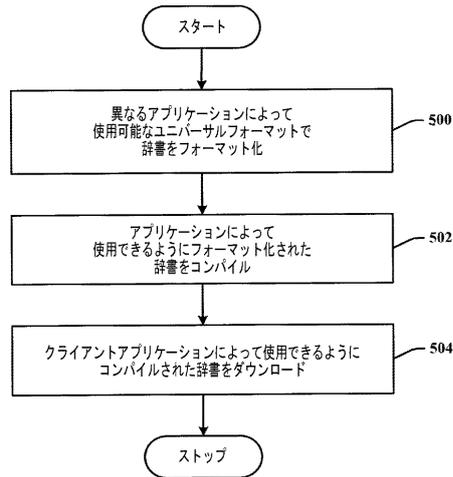
【図3】



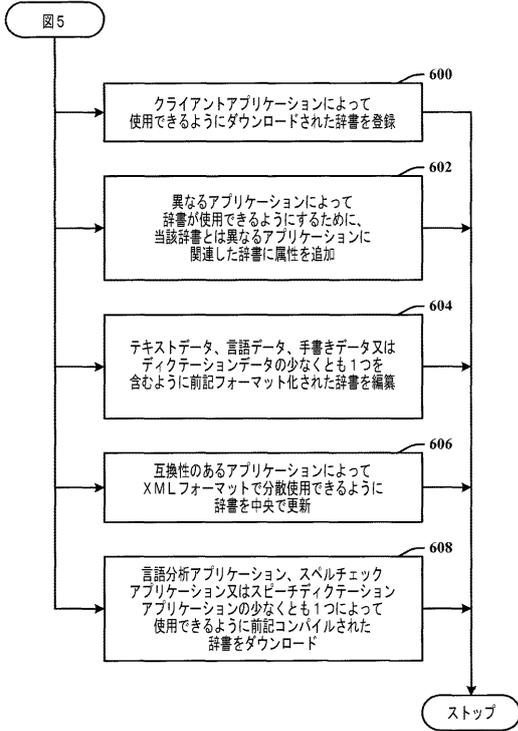
【図4】



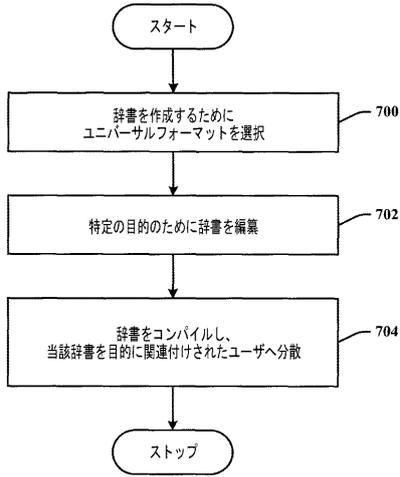
【図5】



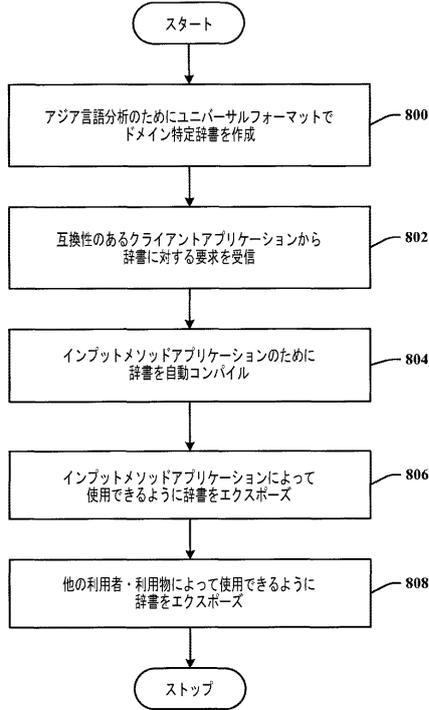
【図6】



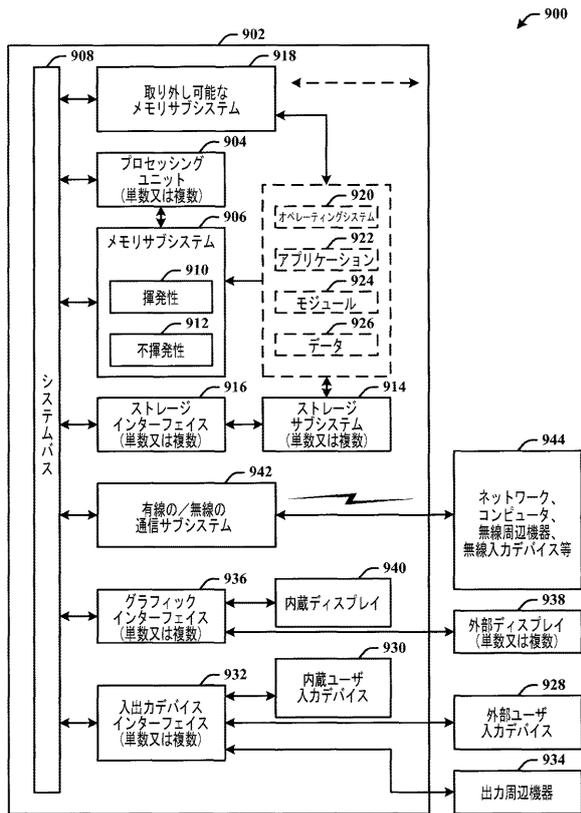
【図7】



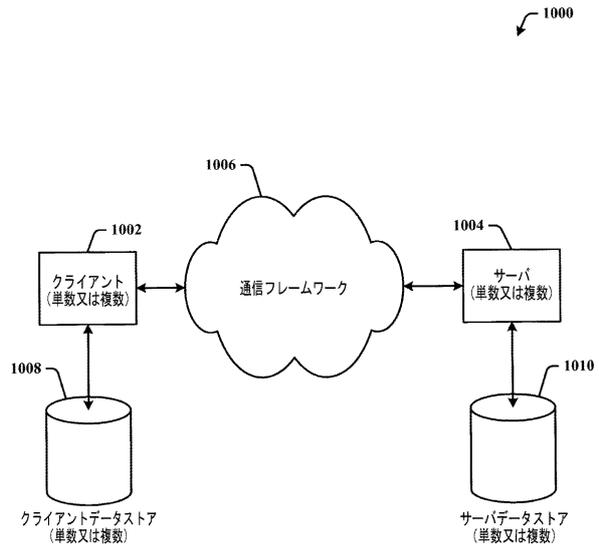
【図8】



【図9】



【図10】



フロントページの続き

- (74)代理人 100120112
弁理士 中西 基晴
- (74)代理人 100147991
弁理士 鳥居 健一
- (74)代理人 100119781
弁理士 中村 彰吾
- (74)代理人 100162846
弁理士 大牧 綾子
- (74)代理人 100173565
弁理士 末松 亮太
- (74)代理人 100138759
弁理士 大房 直樹
- (74)代理人 100091063
弁理士 田中 英夫
- (72)発明者 ケン イタクラ
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション エルシーイー - インターナショナル パテント内

審査官 本郷 彰

- (56)参考文献 特開平08-153090(JP,A)
特開2006-215823(JP,A)
特開2007-156530(JP,A)
特開2007-026155(JP,A)
国際公開第2007/114181(WO,A1)
特開2001-306091(JP,A)
木村 和広,平川 秀樹,ネットワークを利用した辞書サービスシステム,情報処理学会研究報告 Vol.97 No.107,日本,社団法人情報処理学会,1997年11月14日,第97巻 第107号【ISSN】0919-6072,p.27-p.32

- (58)調査した分野(Int.Cl.,DB名)
G06F 17/21