

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号  
特許第7541158号  
(P7541158)

(45)発行日 令和6年8月27日(2024.8.27)

(24)登録日 令和6年8月19日(2024.8.19)

(51)国際特許分類	F I
H 0 4 L 67/568 (2022.01)	H 0 4 L 67/568
G 0 6 F 16/90 (2019.01)	G 0 6 F 16/90 1 0 0
H 0 4 L 67/52 (2022.01)	H 0 4 L 67/52
H 0 4 L 67/561 (2022.01)	H 0 4 L 67/561
G 1 0 L 15/10 (2006.01)	G 1 0 L 15/10 5 0 0 T
請求項の数 22 外国語出願 (全43頁)	

(21)出願番号	特願2023-93524(P2023-93524)	(73)特許権者	502208397
(22)出願日	令和5年6月6日(2023.6.6)		グーグル エルエルシー
(62)分割の表示	特願2021-562375(P2021-562375)		Google LLC
	)の分割		アメリカ合衆国 カリフォルニア州 9 4
原出願日	令和1年5月31日(2019.5.31)		0 4 3 マウンテン ビュー アンフィシ
(65)公開番号	特開2023-116607(P2023-116607)		アター パークウェイ 1 6 0 0
	A)		1 6 0 0 Amphitheatre P
(43)公開日	令和5年8月22日(2023.8.22)		arkway 9 4 0 4 3 Mounta
審査請求日	令和5年6月9日(2023.6.9)		in View, CA U . S . A .
(31)優先権主張番号	62/843,890	(74)代理人	100108453
(32)優先日	令和1年5月6日(2019.5.6)		弁理士 村山 靖彦
(33)優先権主張国・地域又は機関	米国(US)	(74)代理人	100110364
			弁理士 実広 信哉
		(74)代理人	100133400
			弁理士 阿部 達彦

最終頁に続く

(54)【発明の名称】 口頭またはタイプされた発話のオンデバイス分析を可能にするためのクライアントデバイスにおけるアシスタントアクションコンテンツのプロアクティブキャッシング

(57)【特許請求の範囲】

【請求項1】

1つまたは複数のプロセッサによって実行される方法であって、  
リモートプロアクティブアシスタントキャッシュエントリを生成するステップであって、  
前記リモートプロアクティブアシスタントキャッシュエントリは、  
1つまたは複数のアシスタント要求を表すアシスタント要求パラメータであって、前記ア  
シスタント要求パラメータが、前記アシスタント要求の1つまたは複数のテキスト表現お  
よび/または前記アシスタント要求の1つまたは複数の意味表現を定義する、アシスタント  
要求パラメータ、および  
前記1つまたは複数のアシスタント要求に応じたアシスタントアクションを実施させるた  
めの自動アシスタントによって解釈可能であるアシスタントアクションコンテンツ  
を定義する、生成するステップと、  
前記リモートプロアクティブアシスタントキャッシュエントリのためのクライアントデバ  
イスのサブセットを選択するステップであって、前記クライアントデバイスのサブセット  
を選択するステップが、前記サブセットの前記クライアントデバイスの各々が1つまたは  
複数の対応する属性を有すると決定することに基づいて前記サブセットを選択するステッ  
プを含む、選択するステップと、  
前記サブセットの前記クライアントデバイスからリモートにある1つまたは複数のリモ  
ートシステムに、  
前記リモートプロアクティブアシスタントキャッシュエントリ、および

10

20

前記リモートプロアクティブアシスタントキャッシュエントリへの前記選択されたサブセットの前記クライアントデバイスの割り当て

を記憶するステップと、

前記記憶するステップの後に、

所与のクライアントデバイスからの送信において、

前記所与のクライアントデバイスにおいて受信した口頭の発話に基づく所与のデバイスアシスタント要求パラメータ、および

前記所与のクライアントデバイスの識別子

を含む所与のデバイスアシスタント要求を受信するステップと、

前記リモートプロアクティブアシスタントキャッシュエントリが前記所与のデバイスアシスタント要求に応じたものであるとの決定するステップであって、前記リモートプロアクティブアシスタントキャッシュエントリが前記所与のデバイスアシスタント要求に応じたものであるとの決定が、

10

前記所与のクライアントデバイスの前記識別子に基づいて、前記所与のクライアントデバイスが、前記リモートプロアクティブアシスタントキャッシュエントリへの前記サブセットの前記クライアントデバイスの前記割り当てに含まれると決定すること、および

前記所与のデバイスアシスタント要求パラメータが、前記リモートプロアクティブアシスタントキャッシュエントリの前記アシスタント要求パラメータに一致すると決定することの両方に基づく、決定するステップと、

前記リモートプロアクティブアシスタントキャッシュエントリが、前記所与のデバイスアシスタント要求に応じたものであるとの決定に応じて、

20

前記リモートプロアクティブアシスタントキャッシュエントリの前記アシスタントアクションコンテンツを使用して、前記所与のデバイスアシスタント要求に応じて前記アシスタントアクションを実施させるステップと、

を含む、方法。

#### 【請求項 2】

前記サブセットの前記クライアントデバイスの各々が1つまたは複数の対応する属性を有すると決定することに基づいて前記サブセットを選択するステップが、

前記1つまたは複数の対応する属性が、前記リモートプロアクティブアシスタントキャッシュエントリの1つまたは複数のキャッシュエントリ属性に対応するとの決定に基づいて前記1つまたは複数の対応する属性を選択するステップを含む、請求項1に記載の方法。

30

#### 【請求項 3】

前記1つまたは複数のキャッシュエントリ属性が、前記1つまたは複数のアシスタント要求に関するエントリを含む、請求項2に記載の方法。

#### 【請求項 4】

前記1つまたは複数の対応する属性が、特定の地理的領域を含み、前記サブセットの前記クライアントデバイスの各々が1つまたは複数の対応する属性を有すると決定することに基づいて前記サブセットを選択するステップが、

前記サブセットの前記クライアントデバイスが前記特定の地理的領域のすべてにあるという決定に基づいて前記サブセットを選択するステップを含む、請求項1に記載の方法。

40

#### 【請求項 5】

記憶するステップのために、前記1つまたは複数のリモートシステムが前記特定の地理的領域にあることに基づいて前記1つまたは複数のリモートシステムを選択するステップをさらに含む、請求項4に記載の方法。

#### 【請求項 6】

前記1つまたは複数の対応する属性が、特定のアプリケーションを含み、前記サブセットの前記クライアントデバイスの各々が1つまたは複数の対応する属性を有すると決定することに基づいて前記サブセットを選択するステップが、

前記サブセットの前記クライアントデバイスのすべてが、インストールされた前記特定のアプリケーションを有すると決定することに基づいて前記サブセットを選択するステップ

50

を含む、請求項1に記載の方法。

【請求項7】

前記1つまたは複数の対応する属性が、特定のバージョンのオペレーティングシステムを含み、前記サブセットの前記クライアントデバイスの各々が1つまたは複数の対応する属性を有すると決定することに基づいて前記サブセットを選択するステップが、  
前記サブセットの前記クライアントデバイスのすべてが、前記特定のバージョンの前記オペレーティングシステムを有すると決定することに基づいて前記サブセットを選択するステップを含む、請求項1に記載の方法。

【請求項8】

前記1つまたは複数の対応する属性が、前記アシスタント要求パラメータおよび前記アシスタントアクションコンテンツを含む任意のローカルプロアクティブアシスタントキャッシュエントリに欠けることを含み、前記サブセットの前記クライアントデバイスの各々が1つまたは複数の対応する属性を有すると決定することに基づいて前記サブセットを選択するステップは、

10

前記サブセットの前記クライアントデバイスが、前記アシスタント要求パラメータおよび前記アシスタントアクションコンテンツを含む任意のローカルプロアクティブアシスタントキャッシュエントリに欠けると決定することに基づいて前記クライアントデバイスの前記サブセットを選択するステップを含む、請求項1に記載の方法。

【請求項9】

前記1つまたは複数の対応する属性が、ストレージ制約を有し、前記サブセットの前記クライアントデバイスの各々が1つまたは複数の対応する属性を有すると決定することに基づいて前記サブセットを選択するステップは、

20

前記サブセットの前記クライアントデバイスの各々が前記ストレージ制約を有するとの決定に基づいて前記クライアントデバイスの前記サブセットを選択するステップを含む、請求項1に記載の方法。

【請求項10】

前記アシスタントアクションコンテンツを使用して、前記所与のデバイスアシスタント要求に応じて前記アシスタントアクションを実施させるステップが、  
前記アシスタントアクションコンテンツをリモートに実行するステップを含み、前記アシスタントアクションコンテンツをリモートに実行することが、前記アシスタントアクションを実施させる、請求項1に記載の方法。

30

【請求項11】

前記アシスタントアクションコンテンツを使用して、前記所与のデバイスアシスタント要求に応じて前記アシスタントアクションを実施させるステップが、  
前記アシスタントアクションコンテンツを、前記所与のデバイスアシスタント要求に応じて前記所与のクライアントデバイスに送信するステップを含み、前記アシスタントアクションコンテンツを前記所与のクライアントデバイスに送信することが、前記所与のクライアントデバイスに、前記アシスタントアクションを実施させるために、前記アシスタントアクションをローカルに実行させる、請求項1に記載の方法。

【請求項12】

40

前記アシスタントアクションのための要求の増大を決定するステップをさらに含み、前記リモートプロアクティブアシスタントキャッシュエントリを生成することが、前記アシスタントアクションのための要求の増大を決定することに応じたものである、請求項1に記載の方法。

【請求項13】

前記アシスタントアクションがエンティティの音楽をストリーミングすることであり、前記1つまたは複数の対応する属性が、前記エンティティへの関連付けを含む、請求項1に記載の方法。

【請求項14】

前記アシスタントアクションのための要求の増大を予測するステップをさらに含み、

50

前記リモートプロアクティブアシスタントキャッシュエントリを生成することが、前記アシスタントアクションのための要求の増大を予測することに応じたものである、請求項1に記載の方法。

【請求項15】

1つまたは複数のプロセッサによって実行される方法であって、  
特定のエンティティに関連するイベントの発生を決定するステップと、  
前記特定のエンティティに関連する前記イベントの発生を決定することに応じて、  
前記特定のエンティティのための1つまたは複数のリモートプロアクティブアシスタントキャッシュエントリを生成するステップであって、前記リモートプロアクティブアシスタントキャッシュエントリの各々が、  
前記特定のエンティティに関する1つまたは複数のそれぞれのアシスタント要求を表すそれぞれのアシスタント要求パラメータであって、前記アシスタント要求パラメータが、前記アシスタント要求の1つまたは複数のテキスト表現および/または前記アシスタント要求の1つまたは複数の意味表現を定義する、アシスタント要求パラメータ、および  
前記特定のエンティティに関するアシスタントアクションを実施させるために解釈可能なそれぞれのアシスタントアクションコンテンツ  
を定義する、生成するステップと、  
サブセットのクライアントデバイスの各々が、前記特定のエンティティに対応する1つまたは複数の対応する属性を有するとの決定に基づいて、前記クライアントデバイスのサブセットを選択するステップと、  
前記サブセットの前記クライアントデバイスからリモートにある1つまたは複数のリモートシステムに、  
前記1つまたは複数のリモートプロアクティブアシスタントキャッシュエントリ、および  
前記1つまたは複数のリモートプロアクティブアシスタントキャッシュエントリへの前記選択されたサブセットの前記クライアントデバイスの割り当て  
を記憶するステップとを含み、  
前記1つまたは複数のリモートプロアクティブアシスタントキャッシュエントリおよび前記割り当てを記憶することは、前記サブセットの前記クライアントデバイスの1つからのものであると決定され、かつ前記リモートプロアクティブアシスタントキャッシュエントリの1つに一致すると決定された将来のアシスタント要求を履行する際に、プロアクティブアシスタントキャッシュエントリを利用させる、方法。

10

20

30

【請求項16】

前記特定のエンティティに関連する前記イベントの発生を決定するステップが、  
前記特定のエンティティに対する要求の増大および/または前記特定のエンティティに関するインターネットコンテンツの増大を決定するステップ  
を含む、請求項15に記載の方法。

【請求項17】

前記特定のエンティティに対する前記1つまたは複数のプロアクティブアシスタントキャッシュエントリのうちの所与の1つを生成するステップが、  
前記特定のエンティティの1つまたは複数の属性に基づいて、前記特定のエンティティに関する前記1つまたは複数のそれぞれのアシスタント要求を生成するステップと、  
前記1つまたは複数のそれぞれのアシスタント要求に基づいて、前記それぞれのアシスタント要求パラメータを生成するステップと  
を含む、請求項15に記載の方法。

40

【請求項18】

特定のエンティティに関連する前記イベントは、前記特定のエンティティに関する変更された属性、または特定のエンティティに関する新しい属性であり、前記特定のエンティティのための前記1つまたは複数のプロアクティブアシスタントキャッシュエントリを生成するステップは、  
前記プロアクティブアシスタントキャッシュエントリのうちの所与の1つに対する前記ア

50

シスタントアクションの実施中に前記属性をレンダリングさせるために、前に生成されたプロアクティブアシスタントキャッシュエントリに対して、それぞれのアクションコンテンツを改変することによって、前記プロアクティブアシスタントキャッシュエントリの所与の1つを生成するステップを含む、請求項15に記載の方法。

【請求項19】

前記クライアントデバイスの前記サブセットを選択するステップは、前記サブセットの前記クライアントデバイスの各々が、前記それぞれのアシスタント要求パラメータおよび任意の前記プロアクティブアシスタントキャッシュエントリの前記それぞれのアシスタントアクションコンテンツを含む任意のローカルプロアクティブアシスタントキャッシュエントリに欠けると決定することにさらに基づく、請求項15に記載の方法。

10

【請求項20】

前記クライアントデバイスの前記サブセットを選択するステップが、前記サブセットの前記クライアントデバイスの各々が、特定の地理的領域にあると決定することにさらに基づき、

記憶するために、前記1つまたは複数のリモートシステムが前記特定の地理的領域にあることに基づいて、前記1つまたは複数のリモートシステムを選択するステップをさらに含む、請求項15に記載の方法。

【請求項21】

請求項1から20のいずれか一項に記載の方法を実行するための1つまたは複数のプロセッサを備えたシステム。

20

【請求項22】

1つまたは複数のプロセッサによって実行されたとき、前記1つまたは複数のプロセッサに、請求項1から20のいずれか一項に記載の方法を実行させる命令を備える、コンピュータプログラム。

【発明の詳細な説明】

【背景技術】

【0001】

人間は、本明細書で「自動アシスタント」と呼ばれる(「デジタルエージェント」、「対話型パーソナルアシスタント」、「インテリジェントパーソナルアシスタント」、「アシスタントアプリケーション」、「会話エージェント」などとも呼ばれる)対話型ソフトウェアアプリケーションとの人間対コンピュータ対話に関与することができる。たとえば、人間(彼らが自動アシスタントと対話するとき、「ユーザ」と呼ばれることがある)は、場合によっては、テキストに変換され、次いで処理され得る口頭の自然言語入力(すなわち、口頭の発話)を使用して、かつ/またはテキスト(たとえば、タイプされた)自然言語入力(すなわち、タイプされた発話)を提供することによって、コマンドおよび/または要求を自動アシスタントに提供し得る。自動アシスタントは、可聴および/または視覚ユーザインターフェイス出力を含み得る、応答ユーザインターフェイス出力を提供することによって、要求に応答する。

30

【0002】

上述のように、多くの自動アシスタントは、口頭の発話を介して対話するように構成される。ユーザは、ユーザが何の情報提供されることに関心を有するか、かつ/またはユーザが実施されることに関心を有するアクションを言語的に示す、口頭の発話を介して、クエリおよび/またはコマンドをクライアントデバイスの自動アシスタントインターフェイスに提出し得る。一般に、口頭の発話は、クライアントデバイスのマイクロフォンによって検出され、オーディオデータとしてキャプチャされる。オーディオデータは、さらなる処理のためにリモートシステムに送信される。リモートシステムは、オーディオデータを処理して、適切な応答を決定し、その応答をクライアントデバイスに送信する。

40

【0003】

リモートシステムの構成要素は、かなりのコンピューティングリソースをオーディオデータの処理に充て、通常ならばクライアントデバイス内でローカルに実装され得るよりも

50

、より複雑な音声認識および意味解析機能性が実装されることを可能にし得る。しかしながら、クライアント-サーバ手法は、ボイス入力を処理するとき、クライアントがオンラインである(すなわち、リモートシステムと通信している)ことを必ず必要とする。様々な状況において、連続的なオンライン接続性は、常にすべてのロケーションにおいて保証されない可能性があり、したがって、クライアント-サーバのボイスベースのユーザインターフェースは、そのデバイスが「オフライン」であり、したがって、オンラインサービスに接続されていないときはいつでも、クライアントデバイス内で無効にされ得る。さらに、クライアント-サーバ手法は、クライアントからリモートシステムの構成要素への高帯域幅オーディオデータの送信を必要とするため、クライアント-サーバ手法は、かなりの帯域幅を消費し得る。帯域幅の消費は、リモートシステムが大量のクライアントデバイスからの要求を処理している一般的な状況において増幅される。またさらに、クライアント-サーバ手法は、ユーザに応答をプロビジョニングする際にかかなりのレイテンシを示すことがあり、これは、ボイスベースのユーザ-クライアント対話を長引かせ、長引いた持続時間にわたってクライアントデバイスのリソースを利用させる可能性がある。

10

**【発明の概要】****【課題を解決するための手段】****【0004】**

本明細書で開示する実装形態は、プロアクティブアシスタントキャッシュエントリの、クライアントデバイスにおけるローカルなプロアクティブキャッシングを通して、かつプロアクティブアシスタントキャッシュエントリのオンデバイス利用を通して、自動アシスタントから応答を取得するために必要とされる時間を低減するのに役立つ。

20

**【0005】**

様々な実装形態では、クライアントデバイス上で動作している自動アシスタントアプリケーションは、ローカルに検出されたオーディオデータを処理するためにオンデバイス音声処理を使用し、オーディオデータによってキャプチャされた口頭の発話(たとえば、ユーザインターフェース入力)に対応する、認識されたテキストを生成し得る。自動アシスタントアプリケーションは、口頭の発話に対応する、ローカルに記憶されたプロアクティブキャッシュエントリを識別するために、認識されたテキストおよび/またはオンデバイス自然言語理解(NLU)によって生成されたNLUデータをさらに利用し得る。識別は、タイプされた句または発話など、他のユーザインターフェース入力に基づいて実施されることも可能である。ローカルに記憶されたプロアクティブキャッシュエントリは、認識されたテキストおよび/または(タイプされたまたは口頭の発話からの)NLUデータがプロアクティブキャッシュエントリのアシスタント要求パラメータにマッチすると決定することによって、識別され得る。自動アシスタントアプリケーションは、次いで、マッチの決定に応じて、識別されたプロアクティブキャッシュエントリのアクションコンテンツを使用して口頭の発話に適切に応答し得る。自動アシスタントアプリケーションの応答は、たとえば、アクションコンテンツ内に含まれたディープリンクを実行すること、コンテンツ内に含まれたテキスト、グラフィックス、および/またはオーディオ(または、オンデバイス音声合成器を使用して応答から変換されたオーディオデータ)をレンダリングすること、および/または周辺デバイスを制御するために、コマンドを(たとえば、WiFiおよび/またはBluetoothを介して)周辺デバイスに送信することを含み得る。さらに、自動アシスタントアプリケーションの応答は、任意で、サーバとの何のライブ通信も伴わずに、または少なくともサーバからの応答を待機することを必要とせずに提供され、それにより、応答が提供され得る時間をさらに低減し得る。

30

40

**【0006】**

本明細書でより詳細に説明するように、クライアントデバイスのローカルプロアクティブキャッシュ内に記憶された、1つまたは複数の(たとえば、すべての)プロアクティブキャッシュエントリは、それらのエントリがユーザ入力に基づくクライアントデバイスにおける最近の要求に応じてプロアクティブキャッシュ内に記憶されないという意味で、「プロアクティブ」である。むしろ、プロアクティブキャッシュエントリは、サーバなどのリモ

50

ートシステムからプロアクティブにプリフェッチされ、次いで、プロアクティブキャッシュのアシスタント要求パラメータに準拠するユーザ入力が一度も受信されていないにもかかわらず、プロアクティブキャッシュ内に記憶され得る。

【0007】

異なるプロアクティブキャッシュエントリが異なるクライアントデバイスに提供されてよく、本明細書で開示する様々な実装形態は、どのプロアクティブキャッシュエントリをどのクライアントデバイスに提供するかを決定する際、リモートシステムによって利用される技法に関する。それらの実装形態のうちいくつかでは、どのプロアクティブキャッシュエントリを所与のクライアントデバイスに(プロアクティブに、または要求に応じて)提供するかを決定する際、リモートシステムは、所与のクライアントデバイスに提供するためのキャッシュエントリのサブセットを候補プロアクティブキャッシュエントリのスーパーセットから選択する。所与のクライアントデバイスの属性および/またはプロアクティブキャッシュエントリの1つまたは複数の属性を考慮に入れる考慮事項を含めて、様々な考慮事項がサブセットを選択する際に利用され得る。クライアントデバイスの/に対する属性は、たとえば、オペレーティングシステムのバージョンレベルに関連するデータ、クライアントデバイスのモデル、どのアプリケーションがクライアントデバイス上にインストールされるか(および、どのバージョンか)、クライアントデバイスのユーザの検索履歴、クライアントデバイスのロケーション履歴、クライアントデバイスの電力モードまたはバッテリーレベルなど、クライアントデバイスの動作に関連する任意のデータを含み得る。

【0008】

一例として、サブセットの1つまたは複数のプロアクティブキャッシュエントリを選択する際、どのアプリケーションが所与のクライアントデバイス上にインストールされるか(また、任意選択で、それらのうちのどれが、好ましいとして暗示的または明示的にフラグ付けされるか)が考慮され、候補プロアクティブキャッシュエントリのアクションコンテンツ(および/またはメタデータ)と比較され得る。たとえば、各々が、同じアシスタント要求パラメータを有するが、異なるアクションコンテンツを有する、複数の候補プロアクティブキャッシュエントリが生成され得る。各々の異なるアクションコンテンツは、特定のアプリケーションに調整され得、たとえば、所与のアクションを実施するために第1の状態に対応する追加のアプリケーションを開かせるために、アシスタントクライアントアプリケーションによってローカルに実行可能である、対応するディープリンクであり得る。たとえば、「サーモスタットスケジュールを調整する」、「サーモスタットのスケジュールを変更する」、および/または「{intent=change/adjust; device=thermostat; property=schedule}」のアシスタント要求パラメータを仮定されたい。そのようなアシスタント要求パラメータの場合、第1のサーモスタット製造会社に対する第1のアプリケーションおよび第2のサーモスタット製造会社に対する第2のアプリケーションなど、適切な応答を実現するために、様々な異なるアプリケーションが利用され得る。第1および第2のプロアクティブキャッシュエントリは、両方とも、同じアシスタント要求パラメータを有するが、異なるディープリンクを含み得る。たとえば、第1のプロアクティブキャッシュエントリは、(実行されたとき、対応するスケジュール変更インターフェースにおいて第1のアプリケーションを開かせる)第1のアプリケーションに対する第1のディープリンクを有してよく、第2のプロアクティブキャッシュエントリは、(実行されたとき、対応するスケジュール変更インターフェースにおいて第2のアプリケーションを開かせる)第2のアプリケーションに対する別個の第2のディープリンクを有してよい。

【0009】

そのような例では、所与のクライアントデバイスは第1のアプリケーションをインストールしているが、第2のアプリケーションはインストールしていないと決定することに基づいて、所与のクライアントデバイスにプロビジョニングするために、(第2ではなく)第1のプロアクティブキャッシュエントリが選択され得る。これは、所与のクライアントデバイスが、アシスタント要求パラメータに準拠する、タイプされたまたは口頭の発話に急速かつ効率的に応答することを可能にするが、これは、所与のクライアントデバイスは、複

10

20

30

40

50

数の異種のディープリンクのうちのどれが利用されるべきかをローカルに決定する必要なしに、そのエントリのアクションコンテンツの唯一の第1のディープリンクを選択することができるためである。さらに、第1および第2のプロアクティブキャッシュエントリのアクションコンテンツは、複数の異種のディープリンクを含むアクションコンテンツと比較して、ストレージスペースの面で、削減される。これは、それらの各々が個々に、対応するローカルプロアクティブキャッシュ内により効率的に記憶され、対応するクライアントデバイスに送信する際により少ない帯域幅を消費することを可能にする。したがって、より少ないコンピュータおよびネットワークリソースが消費され得る。さらに、プロアクティブキャッシュ内のディープリンクの使用は、より少ないユーザ入力およびより少ない処理ステップでアクションの実施を促すことができ、これは、アクションを実施する際にクライアントデバイスによって消費されるリソースを低減し得る。より少ないユーザ入力に対する要件は、器用さが低減したユーザにとっても有益であり得、デバイスの効用を改善し得る。

10

#### 【0010】

別の例として、サブセットの1つまたは複数のプロアクティブキャッシュエントリを選択する際、任意選択で、イベントの規模に応じて、1つまたは複数の決定されたイベントを有するエンティティに対するプロアクティブキャッシュエントリが選択されるより高い可能性があり得る。さらに、イベントの発生の決定に応じて、エンティティに対するプロアクティブキャッシュエントリすら生成され得る。エンティティに対するイベントを決定するいくつかの例は、エンティティに関する要求の増大を決定すること、エンティティに対するインターネットコンテンツの増大を決定すること、および/またはエンティティに関する要求の増大を予測すること、である。

20

#### 【0011】

たとえば、「Jane Doe」に対するアシスタント要求(および/または旧来の検索要求)の量がスパイクすると決定された場合、「Jane Doe」に関するインターネットコンテンツにおけるスパイクとともに、Jane Doeに関するプロアクティブキャッシュエントリが生成され得、かつ/またはローカルプロアクティブキャッシングのために様々なクライアントデバイスに提供されるより高い可能性が(スパイク前よりも高い可能性が)あり得る。「Jane Doe」に対するプロアクティブキャッシュエントリを様々なクライアントデバイスに提供することは、任意選択で、それらの様々なクライアントデバイスが、Jane Doeに関する、対応する属性(たとえば、Jane Doeに対する過去の検索、および/または関係するエンティティ(たとえば、同じタイプの他のエンティティ)、Jane Doeに関連する地理的ロケーション、および/または他の属性)を有すると決定することにさらに基づき得る。

30

#### 【0012】

これらのおよび他の方法で、最終的な要求がより急速に応答されることを可能にするために、エンティティに対するプロアクティブキャッシュエントリは、エンティティに対する連続的な要求(また、関連する帯域幅消費および要求を処理するための関連するサーバサイドプロセッサ消費)を予期する際に提供され得る。さらに、多くのクライアントデバイスに対するプロアクティブキャッシュエントリのプロビジョニングは、比較的低いネットワーク使用の期間中(たとえば、それらのクライアントデバイスがアイドルであり、充電している夜間)に発生し得、より高い使用期間の間にネットワークリソース使用を軽減しながら、低い使用期間においてネットワークリソースの使用を可能にする。言い換えれば、Jane Doeに対するプロアクティブキャッシュエントリは、低い使用期間中にクライアントデバイスにおいて事前に記憶され、次いで、高い使用期間中にクライアントデバイスにおいてタイプされたまたは口頭の発話にローカルに応答するために利用され得る。大量のクライアントデバイスにわたってこれが発生するとき、本明細書で企図されるように、ネットワークリソースの効果的な再分配を可能にし得る。したがって、ネットワークパフォーマンスが改善され得る。さらに、様々な実装形態は「Jane Doe」に対するプロアクティブキャッシュエントリをプロビジョニングするためにクライアントデバイスのサブセットのみを選択し、Jane Doeに関する属性を有するサブセットのクライアントデバイスに基づいて、

40

50



サブセットを提供する。プロアクティブキャッシュエントリを選択してそのサブセットのみに提供することによって、使用帯域幅およびネットワークリソースの削減が達成され得る。そのようなエントリを利用する可能性があるクライアントデバイスにプロアクティブキャッシュエントリをプロビジョニングするためにネットワークリソースを依然として選択的に利用し、それらのクライアントデバイスにおける応答のより高速なプロビジョニングおよびそれらのクライアントデバイスとの低減された対話持続時間を実現しながら、そのような削減が達成され得る。

#### 【0013】

エンティティに対するイベントを決定および利用する別の例として、要求の増大がまだ観測されていないにもかかわらず、Hypothetical Artist(この例では、仮想ミュージカルアーティスト)に関する要求の増大が予測され得る。たとえば、要求の増大は、1つまたは複数の追加システムから、Hypothetical Artistが新しい歌および/または新しいアルバムを発表することが予定されていると決定することに基づいて予測され得る。要求の増大の予測に応じて、Hypothetical Artistに関するプロアクティブキャッシュエントリが生成され得、かつ/またはローカルプロアクティブキャッシングのために様々なクライアントデバイスに提供されるより高い可能性(スパイクの前よりも高い可能性)があり得る。たとえば、「Hypothetical Artistを再生する」、「あるHypothetical Artistを聴く」、および/または「{intent=listen to music; artist= Hypothetical Artist }」のアシスタント要求パラメータを有するプロアクティブキャッシュエントリ。プロアクティブキャッシュエントリは、たとえば、実行されたとき、各々が、Hypothetical Artistからの音楽がストリーミングされ、可聴的にクライアントデバイスにレンダリングされる状態に対応するアプリケーションを開かせる、1つまたは複数のディープリンクなど、アシスタントアクションコンテンツをさらに含み得る。「Hypothetical Artist」に対するプロアクティブキャッシュエントリを様々なクライアントデバイスに提供することは、任意選択で、それらの様々なクライアントデバイスがHypothetical Artistに関する、対応する属性(たとえば、Hypothetical Artistおよび/または関係するミュージカルアーティストからの音楽の過去のストリーミング、前記デバイス上に記憶されたHypothetical Artistからの音楽ファイル、Hypothetical Artistに関する検索履歴、など)を有すると決定することにさらに基づき得る。プロアクティブキャッシュエントリを所与のクライアントデバイスに提供することは、所与のクライアントデバイスがプロアクティブキャッシュエントリのディープリンクのうちの1つに(たとえば、いくつかの実装形態では、唯一のディープリンクに)対応するアプリケーションを有すると決定することに基づいてもよい。

#### 【0014】

各々が、対応するクライアントデバイスにおいてローカルに記憶されたローカルプロアクティブキャッシュに加えて、いくつかの実装形態は、各々がクライアントデバイスのサブセットに対して生成されたりリモートプロアクティブキャッシュをさらに含み得る。リモートプロアクティブキャッシュに対するクライアントデバイスのサブセットは、単一のクライアントデバイスであってよく、または互いと共通する属性を有するそれらのクライアントデバイスに基づいてグループ化されたクライアントデバイスのグループであってよい。たとえば、リモートプロアクティブキャッシュは、単一のクライアントデバイスのみに対するものであってよく、または同じまたは同様の属性を有する1,000個のクライアントデバイスに対するものであってもよい。

#### 【0015】

所与のクライアントデバイスに対する(唯一その所与のクライアントデバイスに対する、またはその所与のクライアントデバイスおよび同じ/同様の属性を備えた他のクライアントデバイスに対する)リモートプロアクティブキャッシュは、所与のクライアントデバイスのローカルプロアクティブキャッシュ内に記憶されたプロアクティブキャッシュエントリに対する追加であるプロアクティブキャッシュエントリを含む(または、それに制限される)。リモートプロアクティブキャッシュのプロアクティブキャッシュエントリは、多くの実装形態では、依然として、すべての利用可能な候補キャッシュエントリのサブセットであ

10

20

30

40

50

る。リモートプロアクティブキャッシュのプロアクティブキャッシュエントリは、所与のクライアントデバイスの属性および/またはプロアクティブキャッシュエントリの属性に基づいて(たとえば、クライアントデバイス属性とプロアクティブキャッシュエントリ属性の比較に基づいて、かつ/またはプロアクティブキャッシュエントリ属性に基づいて)、所与のクライアントデバイスに関連すると決定されたプロアクティブキャッシュエントリを含み得る。しかしながら、リモートキャッシュのプロアクティブキャッシュエントリは、ローカルプロアクティブキャッシュ内のローカルストレージに提供されないプロアクティブキャッシュエントリを含む。ローカルプロアクティブキャッシュ内のローカルストレージに対してそれらを提供しないとの判定は、ローカルプロアクティブキャッシュに対するストレージスペース制限に基づいてよく(たとえば、それらを提供することは、ローカルプロアクティブキャッシュに対するストレージ制限を超えさせる)、またそれらのエントリが所与のクライアントデバイスのローカルプロアクティブキャッシュ内のストレージに提供されるものにあまり関連しないと決定されることに基づいてよい。言い換えれば、所与のクライアントデバイスに対するリモートプロアクティブキャッシュは、ストレージスペース制限に基づいて、かつプロアクティブキャッシュエントリが所与のクライアントデバイスのローカルプロアクティブキャッシュ内のローカルストレージに提供されたものにあまり関連しないと決定されることに基づいて、所与のクライアントデバイスに関連するが、所与のクライアントデバイスのローカルキャッシュ内に記憶されないと決定されたプロアクティブキャッシュエントリを含み得る。

10

## 【 0 0 1 6 】

20

リモートプロアクティブキャッシュは、リモートプロアクティブキャッシュに割り当てられたクライアントデバイスからのアシスタント要求に応答するとき、リモート自動アシスタント構成要素によって利用され得る。たとえば、所与のクライアントデバイスにおいてオーディオデータ内で検出された口頭の発話の場合、所与のクライアントデバイスは、オーディオデータおよび/またはローカルに決定された、認識されたテキストをリモート自動アシスタント構成要素に送信することができる。リモート自動アシスタント構成要素への送信は、任意選択で、何のローカルプロアクティブキャッシュエントリも応答しないとの決定に応じてよく、またはローカルプロアクティブキャッシュエントリが応答するかどうかの決定と並行して発生してよい。送信には、所与のクライアントデバイスの識別子、および所与のクライアントデバイスに対するリモートプロアクティブキャッシュを識別するために利用される識別子が添付されてよい。自動アシスタント構成要素のリモート履行モジュールは、リモートプロアクティブキャッシュエントリがアシスタント要求に応答するかどうかを決定し得る。応答する場合、リモート履行モジュールは、アシスタント要求に応答するためのアシスタントアクションコンテンツを決定するために応答エントリを利用し得る。アシスタントアクションコンテンツは、リモート自動アシスタント構成要素によってリモートに実行され、かつ/またはローカル実行のために所与のクライアントデバイスに送信されてよい。

30

## 【 0 0 1 7 】

リモート自動アシスタント構成要素によるリモートプロアクティブキャッシュの利用は、リモート自動アシスタント構成要素が自動アシスタント要求(たとえば、所与のアクションの実施を要求する自動アシスタントを対象とする、タイプされたまたは口頭の発話)により急速に応答することを可能にする。これは、アシスタントアクションがプロアクティブキャッシュエントリ内でアシスタント要求パラメータに直接的にマッピングされ、アシスタント要求に応じてアシスタントアクションをライブで生成する必要なしに、プロアクティブキャッシュエントリから、アシスタントアクションの効率的な識別を可能にし得る結果であり得る。たとえば、プロアクティブキャッシュエントリ内にアシスタントアクションがなければ、アシスタントアクションは、任意選択で、1つまたは複数の追加のリモートシステムとの通信を通して、オンザフライで生成されなければならないことになり、これは、追加のリモートシステムとの通信の結果として、計算的に負担になり、レイテンシをもたらし得る。したがって、クライアント-サーバラウンドトリップは必要とされるが、

40

50

アシスタント要求を分析する際のリモートプロアクティブキャッシュの利用は、依然として、応答の低減されたレイテンシプロビジョニングを提供し、結果として、ユーザ-アシスタント対話の持続時間を低減する。さらに、いくつかの実装形態では、クライアントデバイスのサブセットに固有のリモートプロアクティブキャッシュは、所与の地理的領域からアシスタント要求を受信するサーバなど、サブセットのクライアントデバイスに地理的に近接している1つまたは複数のサーバにおいて記憶され得る。複数の地理的に離れたサーバ同士の間でのサーバ間通信を必要とせずに、アシスタントアクションはプロアクティブキャッシュからより急速に決定され得るため、これは、アシスタント要求を分析する際のレイテンシをさらに低減し得る。

**【 0 0 1 8 】**

本明細書で開示する様々な実装形態は、少なくとも1つまたは複数のマイクロフォンおよび自動アシスタントアプリケーションを含むクライアントデバイス(たとえば、スマートフォンおよび/または他のクライアントデバイス)を対象とする。自動アシスタントアプリケーションは、クライアントデバイスのオペレーティングシステムの「上部に」インストールされてよく、かつ/またはそれ自体がクライアントデバイスのオペレーティングシステムの部分(または、その全体)を形成し得る。自動アシスタントアプリケーションは、オンデバイス音声認識、オンデバイス自然言語理解、およびオンデバイス履行を含み、かつ/またはそれらに対するアクセスを有する。たとえば、オンデバイス音声認識は、クライアントデバイスにおいてローカルに記憶されたエンドツーエンド音声認識機械学習モデルを使用して、(マイクロフォンによって検出された)オーディオデータを処理するオンデバイス音声認識モジュールを使用して実施され得る。オンデバイス音声認識は、オーディオデータ内に存在する口頭の発話(もしあれば)に対して認識されたテキストを生成する。また、たとえば、オンデバイス自然言語理解(NLU)は、NLUデータを生成するために、オンデバイス音声認識を使用して生成された、認識されたテキスト、および任意選択で、コンテキストデータを処理するオンデバイスNLUモジュールを使用して実施され得る。NLUデータは、口頭の発話に対応する意図、および任意選択で、その意図に対するパラメータ(たとえば、スロット値)を含み得る。

**【 0 0 1 9 】**

オンデバイス履行は、口頭の発話の意図(および任意選択で、その意図に対するパラメータ)を分析するために起こすアクションを決定するために、(オンデバイス音声認識からの)認識されたテキストおよび/または(オンデバイスNLUからの)NLUデータ、および任意選択で、他のローカルデータを利用するオンデバイス履行モジュールを使用して実施され得る。これは、口頭の発話に対するローカルおよび/またはリモート応答(たとえば、回答)、口頭の発話に基づいて実施するためのローカルにインストールされたアプリケーションとの対話、口頭の発話に基づいて(直接的に、または対応するリモートシステムを介して)モノのインターネット(IoT)デバイスに送信するためのコマンド、および/または口頭の発話に基づいて実施するための他の分析アクションを決定することを含み得る。オンデバイス履行は、次いで、口頭の発話を分析するために決定されたアクションのローカルおよび/またはリモート実施/実行を開始し得る。本明細書で説明するように、様々な実装形態では、オンデバイス履行は、様々なユーザ入力に応答して、ローカルに記憶されたプロアクティブキャッシュを利用する。たとえば、オンデバイス履行は、認識されたテキストおよび/またはNLUデータがプロアクティブキャッシュエントリのアシスタント要求パラメータにマッチすると決定することに基づいて、口頭の発話に応答して、ローカルに記憶されたプロアクティブキャッシュのプロアクティブキャッシュエントリのアクションコンテンツを利用し得る。

**【 0 0 2 0 】**

様々な実装形態では、リモート音声処理、リモートNLU、および/またはリモート履行は、少なくとも選択的に利用され得る。たとえば、認識されたテキストは、リモートNLUおよび/またはリモート履行のためにリモート自動アシスタント構成要素に少なくとも選択的に送信され得る。たとえば、認識されたテキストは、任意選択で、オンデバイス実施と

10

20

30

40

50

並行して、またはオンデバイスNLUおよび/またはオンデバイス履行の失敗に応じて、リモート実施のために送信され得る。しかしながら、オンデバイス音声処理、オンデバイスNLU、オンデバイス履行、および/またはオンデバイス実行は、少なくとも口頭の発話を分析するときにこれらが提供するレイテンシ低減により(口頭の発話を分析するためにクライアント-サーバラウンドトリップが必要とされないことにより)優先され得る。さらに、オンデバイス機能性は、ネットワーク接続性がないかまたは限定されている状況において利用可能である唯一の機能性であり得る。

#### 【0021】

様々な実装形態では、オンデバイス音声認識および/または他のオンデバイスプロセスは、明示的なアシスタント起動キューの発生の検出に応じて、かつ/または暗示的な起動キューの何らかの発生に応じて、アクティブ化される。明示的な起動キューは、別個に検出されるとき、少なくともオンデバイス音声認識を常にアクティブ化させることになるキューである。明示的な起動キューのいくつかの非限定的な例は、少なくともしきい値程度の信頼性を有する口頭のホットワード、明示的なアシスタントインターフェース要素(たとえば、ハードウェアボタンまたはタッチスクリーンディスプレイ上のグラフィカルボタン)の作動、少なくともしきい値強度の「フォンスクイズ(phone squeeze)」(たとえば、モバイルフォンのベゼル内のセンサによって検出されるような)および/または他の明示的な起動キューを検出することを含む。しかしながら、オンデバイス音声認識は、一定のコンテンツ内の発生(たとえば、他の暗示的なキューに続く、またはそれと組み合わせられた発生)など、それらのキューの何らかの発生に応じてのみアクティブ化されることになるという点で、他のキューは暗示的である。たとえば、オンデバイス音声認識は、任意選択で、独立したボイスアクティビティの検出に応じて、アクティブ化され得ないが、クライアントデバイスにおけるユーザの存在の検出および/またはしきい値距離内のクライアントデバイスにおけるユーザの存在の検出とともにボイスアクティビティを検出することに応じて、アクティブ化され得る。また、たとえば、ユーザがクライアントデバイスをピックアップしたこと、および/またはクライアントデバイスを現在保持していることを示すジャイロおよび/または加速度計など、非マイクロフォンセンサからのセンサデータは、任意選択で、独立して、オンデバイス音声認識をアクティブ化し得ない。しかしながら、オンデバイス音声認識は、ボイスアクティビティおよび/またはホットワードフリーオーディオデータ内の指向性音声(directed speech)(本明細書でより詳細に説明する)の検出とともにそのような表示に応じて、アクティブ化され得る。ホットワードフリーオーディオデータは、明示的なアシスタント起動キューである「ホットワード」を含む何らかの口頭の発話に欠けるオーディオデータである。さらに別の例として、しきい値強度未満の「フォンスクイズ」は、任意選択で、独立して、オンデバイス音声認識をアクティブ化するには不十分であり得る。しかしながら、オンデバイス音声認識は、ボイスアクティビティおよび/またはホットワードフリーオーディオデータ内の指向性音声の検出とともに、そのような低強度の「フォンスクイズ」に応じて、アクティブ化され得る。

#### 【0022】

本明細書で開示するいくつかの実装形態は、中央処理装置(CPU)、グラフィックス処理ユニット(GPU)、および/またはテンソル処理ユニット(TPU)など、1つまたは複数のプロセッサを含む、1つまたは複数のコンピューティングデバイスを含む。プロセッサのうちの1つまたは複数は、関連するメモリ内に記憶された命令を実施するように動作可能であり、これらの命令は、本明細書で説明する方法のうちのいずれかを実行させるように構成される。コンピューティングデバイスは、たとえば、マイクロフォン、ディスプレイ、および/または他のセンサ構成要素を備えたクライアントアシスタントデバイスを含み得る。いくつかの実装形態はまた、本明細書で説明する方法のうちのいずれかを実施するために1つまたは複数のプロセッサによって実行可能なコンピュータ命令を記憶した、1つまたは複数の非一時的コンピュータ可読記憶媒体を含む。

#### 【図面の簡単な説明】

#### 【0023】

10

20

30

40

50

【図1】本明細書で開示する実装形態が実装され得る例示的な環境のブロック図である。

【図2】様々な実装形態による、図1の様々な構成要素がどのように対話し得るかを説明する例示的なプロセスフローを示す図である。

【図3】プロアクティブキャッシュエントリのいくつかの例を示す図である。

【図4】本明細書で開示する実装形態による、プロアクティブキャッシュエントリをプリフェッチして記憶する例示的な方法を示すフローチャートである。

【図5A】プロアクティブキャッシュエントリを生成し、プロアクティブキャッシュエントリのローカルサブセットを所与のクライアントデバイスにプロビジョニングし、かつ/またはプロアクティブキャッシュエントリのリモートサブセットを決定する例示的な方法を示すフローチャートである。

10

【図5B】図5Aのブロック510のいくつかの実装形態を示すフローチャートである。

【図5C】図5Aのブロック510のいくつかの追加または代替の実装形態を示すフローチャートである。

【図6】コンピューティングデバイスの例示的なアーキテクチャを示す図である。

【発明を実施するための形態】

【0024】

最初に図1を参照すると、自動アシスタントクライアント170を少なくとも選択的に実行するクライアントデバイス160が示される。「アシスタントデバイス」という用語は、本明細書で、自動アシスタントクライアント170を少なくとも選択的に実行するクライアントデバイス160を指すために使用される。自動アシスタントクライアント170は、図1の例において、音声キャプチャエンジン171、視覚キャプチャエンジン172、オンデバイス音声認識エンジン173、オンデバイスNLUエンジン174、オンデバイス履行エンジン175、オンデバイス実行エンジン176、およびプリフェッチエンジン177を含む。

20

【0025】

1つまたは複数のリモート自動アシスタント構成要素180は、任意選択で、概して190において示される、1つまたは複数のローカルおよび/または広域ネットワーク(たとえば、インターネット)を介してクライアントデバイス160に通信可能に結合されている、1つまたは複数のコンピューティングシステム上で実装され得る。リモート自動アシスタント構成要素180は、たとえば、高性能サーバのクラスタを介して実装され得る。

【0026】

30

様々な実装形態では、自動アシスタントクライアント170のインスタンスは、1つまたは複数のクラウドベースの自動アシスタント構成要素180とのその対話により、ユーザの観点から、ユーザがそれを用いて人間対コンピュータ対話(たとえば、口頭の対話、ジェスチャベースの対話、および/またはタッチベースの対話)に関与し得る自動アシスタント195の論理インスタンスに見えるものを形成し得る。

【0027】

クライアントデバイス160は、たとえば、デスクトップコンピューティングデバイス、ラップトップコンピューティングデバイス、タブレットコンピューティングデバイス、モバイルフォンコンピューティングデバイス、ユーザの車両のコンピューティングデバイス(たとえば、車両内通信システム、車両内エンターテインメントシステム、車両内ナビゲーションシステム)、スタンドアロン対話型スピーカ、スマートテレビジョン(または、自動アシスタント機能を備えたネットワーク接続されたドングルを装備した標準テレビジョン)などのスマートアプライアンス、および/またはコンピューティングデバイスを含む、ユーザのウェアラブル装置(たとえば、コンピューティングデバイスを有するユーザのウォッチ、コンピューティングデバイスを有するユーザの眼鏡、仮想または拡張現実コンピューティングデバイス)であってよい。追加および/または代替のクライアントコンピューティングデバイスが提供され得る。

40

【0028】

クライアントデバイス160は、任意選択で、1つまたは複数の視野を有する、1つまたは複数の視覚構成要素163を装備し得る。視覚構成要素163は、モノグラフィックカメラ、

50

ステレオカメラ、LIDAR構成要素(または、他のレーザベースの構成要素)、レーダ構成要素など、様々な形態をとってよい。1つまたは複数の視覚構成要素163は、クライアントデバイス160が配備される環境の視覚フレーム(たとえば、画像フレーム、レーザベースのフレーム)をキャプチャするために、たとえば、視覚キャプチャエンジン172によって使用され得る。いくつかの実装形態では、そのような視覚フレームは、ユーザがクライアントデバイス160の近くに存在するかどうか、および/またはクライアントデバイスに対するユーザ(たとえば、ユーザの顔)の距離を決定するために利用され得る。そのような決定は、オンデバイス音声認識エンジン173、オンデバイスNLUエンジン174、オンデバイス履行エンジン175、および/またはオンデバイス実行エンジン176をアクティブ化するかどうかを決定する際に、自動アシスタントクライアント170によって利用され得る。視覚エンジンは、追加または代替として、様々なユーザタッチフリージェスチャ(たとえば、「サムズアップ」、「ハンドウェーブ」など)をローカルに検出するために利用され得る。任意選択で、検出されたジェスチャは、本明細書で説明するプロアクティブキャッシュエントリの「アシスタント要求パラメータ」であり得、そのジェスチャに基づいて起こすべき対応するアクションは、プロアクティブキャッシュエントリの「アクションコンテンツ」であり得る。

10

**【0029】**

クライアントデバイス160はまた、1つまたは複数のマイクロフォン165を装備し得る。音声キャプチャエンジン171は、マイクロフォン165を介してキャプチャされたユーザの音声および/または他のオーディオデータをキャプチャするように構成され得る。本明細書で説明するように、そのようなオーディオデータは、オンデバイス音声認識エンジン173によって利用され得る。

20

**【0030】**

クライアントデバイス160はまた、1つまたは複数の存在センサ167および/または1つまたは複数のディスプレイ169(たとえば、タッチセンシティブディスプレイ)を含み得る。ディスプレイ169は、任意選択で、オンデバイス音声認識エンジン173からのストリーミングテキスト転写をレンダリングするために利用され得る。ディスプレイ169はさらに、それを通して応答の視覚部分が自動アシスタントクライアント170からレンダリングされるユーザインターフェース出力構成要素のうちの一つであってよい。存在センサ167は、たとえば、PIRおよび/または他のパッシブ存在センサを含み得る。様々な実装形態では、自動アシスタントクライアント170の1つまたは複数の構成要素および/または機能は、存在センサ167からの出力に基づいて、人間の存在の検出に応じて開始され得る。自動アシスタントクライアント170の構成要素および/または機能の開始が1人または複数のユーザの存在の第1の検出を条件とする実装形態では、電力リソースが節約され得る。

30

**【0031】**

自動アシスタントクライアント170は、明示的な起動キューの発生を検出、および/または暗示的な起動キューの発生を検出に応じて、オンデバイス音声認識エンジン173をアクティブ化する。アクティブ化されるとき、オンデバイス音声認識エンジン173は、オーディオデータによってキャプチャされた口頭の発話(もしあれば)内で認識されたテキストを決定するために、オンデバイス音声認識モデル(簡単のために、図1に図示せず)を使用して、マイクロフォン165によってキャプチャされたオーディオデータを処理する。オンデバイス音声認識モデルは、任意選択で、エンドツーエンドモデルであってよく、任意選択で、追加の認識されたテキスト仮説を生成し、様々な考慮事項を使用して最善の仮説を選定することを探求する1つまたは複数の技法によって補足され得る。処理されるオーディオデータは、オンデバイス音声認識がアクティブ化された後にキャプチャされたオーディオデータ、および任意選択で、ローカルにバッファリングされた(たとえば、オンデバイス音声認識エンジン173のアクティブ化に先立ってバッファリングされた)最近のオーディオデータを含み得る。オンデバイス音声認識エンジン173によって処理されるオーディオデータは、生のオーディオデータ、および/またはその表現を含み得る。オーディオデータは、新しいオーディオデータが検出されるにつれて、ストリーミング様式でオンデバイス音声

40

50

認識エンジン173に提供され得る。

【0032】

いくつかの実装形態では、オンデバイス音声認識エンジン173がアクティブ化されるとき、そのようなアクティブ化が発生したことをユーザに知らせるために、かつ/または認識が発生するにつれて認識されたテキストのストリームをレンダリングするために、人間が知覚可能なキューがレンダリングされる。視覚的レンダリングはまた、タッチスクリーンディスプレイにおけるタッチ入力を介して選択されるとき、オンデバイス音声認識エンジン173を停止する選択可能「消去」要素を含んでもよい。本明細書で使用されるように、音声認識エンジン173または他の構成要素をアクティブ化することは、その構成要素がアクティブ化に先立って前に実施していたものを超える処理を少なくとも実施させることを意味する。それは、完全休眠状態から構成要素をアクティブ化することを意味し得る。

10

【0033】

オンデバイスNLUエンジン174、オンデバイス履行エンジン175、および/またはオンデバイス実行エンジン176は、任意選択で、明示的および/または暗黙的起動キューの発生の検出に応じて、アクティブ化され得る。代替として、これらのエンジンのうちの1つまたは複数は、認識されたテキストがアシスタント要求である可能性があることを示す、オンデバイス音声認識エンジン173からの認識されたテキストの初期解析に基づいてのみアクティブ化され得る。

【0034】

オンデバイスNLUエンジン174がアクティブ化されるとき、オンデバイスNLUエンジン174は、NLUデータを生成するために、オンデバイス音声認識エンジン173によって生成された、認識されたテキストに対してオンデバイス自然言語理解を実施する。NLUエンジン174は、任意選択で、NLUデータを生成する際、1つまたは複数のオンデバイスNLUモデル(簡単のために、図1には図示せず)を利用し得る。NLUデータは、たとえば、口頭の発話に対応する意図、および任意選択で、その意図に対するパラメータ(たとえば、スロット値)を含み得る。

20

【0035】

さらに、オンデバイス履行エンジン145がアクティブ化されるとき、オンデバイス履行エンジン145は、(オンデバイス音声認識エンジン173からの)認識されたテキストを使用して、かつ/または(オンデバイスNLUエンジン174からの)NLUデータを使用して、履行データを生成する。オンデバイス履行エンジン145は、認識されたテキストおよび/またはNLUデータがプロアクティブキャッシュのプロアクティブキャッシュエントリのアシスタント要求パラメータにマッチするかどうかを決定する際、クライアントデバイス160においてローカルに記憶されたプロアクティブキャッシュ178にアクセスし得る。マッチする場合、オンデバイス履行エンジン145は、生成された履行データのすべてまたは部分として、マッチするプロアクティブキャッシュエントリのアクションコンテンツを利用し得る。アクションコンテンツは、その要求および/またはそのような要求の履行に関連するデータに基づいて起こすことになるアクションを含み得る。マッチしない場合、オンデバイス履行エンジン145は、任意選択で、履行データの生成を試行する際に他のオンデバイス履行モデル(もしあれば)を利用することができるか、または(たとえば、自動アシスタントクライアント170がオンラインであり、かつ履行データを生成するために、認識されたテキストおよび/または他のデータをリモート自動アシスタント構成要素180に提供するとき)リモート自動アシスタント構成要素180からリモート履行データを待機し得る。オンデバイス音声認識とローカルプロアクティブキャッシュの組合せは、サーバとの間でデータを送信する必要を低減し、したがって、帯域幅およびネットワークリソースの使用を低減し得る。さらに、ネットワーク接続が不十分であるかまたはないエリアですら、低レイテンシ応答がユーザに提供され得る。

30

40

【0036】

認識されたテキストまたはNLUデータがプロアクティブキャッシュのプロアクティブキャッシュエントリのアシスタント要求パラメータにマッチするかどうかを決定する際、オ

50

ンデバイス履行エンジン145は、完全マッチおよび/またはソフトマッチを利用し得る。たとえば、認識されたテキストがアシスタント要求パラメータにマッチするかどうかを決定する際、オンデバイス履行エンジン145は、アシスタント要求パラメータのテキストに対する認識されたテキストの完全なマッチを要求し得るか、または最小差異(たとえば、一定の停止語の包含または排除)のみを可能にし得る。また、たとえば、NLUデータがアシスタント要求パラメータにマッチするかどうかを決定する際、オンデバイス履行エンジン145は、アシスタント要求パラメータのNLUデータに対するNLUデータの完全なマッチを必要とし得る。オンデバイス音声認識によって決定された、認識されたテキストに関して、本明細書において様々な例が説明されるが、タイプされたテキスト(たとえば、仮想キーボードを使用してタイプされた)および/またはタイプされたテキストに基づくNLUデータも同様の方法で提供され応答され得ることを理解されたい。

10

**【0037】**

履行データがオンデバイス履行エンジン175によって生成されるとき、履行データは、次いで、履行データに基づいたオンデバイス実行(たとえば、プロアクティブキャッシュエントリのアクションコンテンツに基づくアクションのオンデバイス実施)のためにオンデバイス実行エンジン176に提供され得る。履行データのアクションコンテンツに基づくオンデバイス実行は、たとえば、アクションコンテンツ内に含まれたディープリンクを実行すること、アクションコンテンツ内に含まれたテキスト、グラフィックス、および/またはオーディオ(または、オンデバイス音声合成器を使用して応答から変換されたオーディオデータ)をレンダリングすること、および/または周辺デバイスを制御するために、アクションコンテンツ内に含まれたコマンドを(たとえば、WiFiおよび/またはBluetoothを介して)周辺デバイスに送信することを含み得る。

20

**【0038】**

プリフェッチエンジン177は、プロアクティブキャッシュ178内に含めるためにプロアクティブキャッシュエントリをプロアクティブキャッシュシステム120からプリフェッチする。プリフェッチされ、プロアクティブキャッシュ178内に記憶される、1つまたは複数のプロアクティブキャッシュエントリは、自動アシスタントクライアント170に前に提供された任意のユーザ入力に準拠しないアシスタント要求パラメータを有する、かつ/または自動アシスタントクライアント170によって一度も利用されていないアクションコンテンツを有する、プロアクティブキャッシュエントリであり得る。したがって、そのようなキャッシュエントリは、自動アシスタントクライアント170によって一度も受信されていない様々なユーザ入力に対するアシスタント応答を提供するための時間量を低減し得る。

30

**【0039】**

いくつかの実装形態および/または状況では、プロアクティブキャッシュシステム120は、任意選択で、プロアクティブキャッシュエントリをプリフェッチエンジン177にプッシュし得る。しかしながら、他の実装形態および/または状況では、プロアクティブキャッシュシステム120は、プリフェッチエンジン177からのプリフェッチ要求に応じて、プロアクティブキャッシュエントリを送信する。それらの実装形態のうちいくつかでは、プリフェッチエンジン177は、1つまたは複数の条件が満たされているとの決定に応じて、プリフェッチ要求を送信する。条件は、たとえば、一定のネットワーク条件が存在する(たとえば、Wi-Fiネットワークに対する接続および/または一定の帯域幅条件を有するWi-Fiネットワークに対する接続)、クライアントデバイス160が充電しているかつ/または少なくともしきい値バッテリー充電状態を有する、(たとえば、オンデバイス加速度計および/またはジャイロスコープデータに基づいて)クライアントデバイスがユーザによってアクティブに利用されていない、クライアントデバイスの現在のプロセッサ使用および/または現在のメモリ使用が一定のしきい値を超えていない、などのうちの1つまたは複数を含み得る。したがって、プロアクティブキャッシュエントリは、一定の理想的な条件が存在する期間中取り出されることが可能であるが、一定の理想的な条件が存在しない場合を含めて、後で何らかの条件下で利用され得る。

40

**【0040】**

50



プリフェッチエンジン177からのプリフェッチ要求は、クライアントデバイス160のかつ/またはクライアントデバイスに関連するアカウントの識別子を含んでよく、かつ/またはプロアクティブキャッシュ178内にすでに記憶されているプロアクティブキャッシュエントリの別の指示を含んでもよい。プロアクティブキャッシュシステム120は、プロアクティブキャッシュ178内にすでに記憶されているプロアクティブキャッシュエントリの指示を利用して、プリフェッチ要求に応じて、プロアクティブキャッシュ内にまだ記憶されていないプロアクティブキャッシュエントリのみを提供し得る。これは、プロアクティブキャッシュ178に完全に取って代わるために既存のキャッシュエントリとともに新しい/更新されたプロアクティブキャッシュエントリを送信する代わりに、プロアクティブキャッシュ178に追加するために新しい/更新されたプロアクティブキャッシュエントリを送信することのみによってネットワークリソースを節約し得る。一例として、プロアクティブキャッシュシステム120は、クライアントデバイス160の識別子に対して、プロアクティブキャッシュ178内のアクティブなプロアクティブキャッシュエントリのリステイングを保持し得る。プロアクティブキャッシュシステム120は、識別子を含むプリフェッチ要求に応じて、どの新しい/更新されたプロアクティブキャッシュエントリを提供すべきかを決定する際にそのようなリステイングを利用し得る。別の例として、プリフェッチ要求は、プロアクティブキャッシュ178内に記憶されたプロアクティブキャッシュエントリのセットにマップするトークンまたは他の識別子を含んでよく、プロアクティブキャッシュシステム120は、どの新しい/更新されたプロアクティブキャッシュエントリが、マッピングされたセット内にはないか、かつそのトークンを含むプリフェッチ要求に応じて提供されるべきかを決定する際にそのようなトークンを利用し得る。

#### 【0041】

プリフェッチエンジン177は、プロアクティブキャッシュ178からプロアクティブキャッシュエントリを選択的に除去することも可能である。たとえば、プロアクティブキャッシュエントリは、そのメタデータの部分として有効期間(TTL:time to live)値を含み得る。TTL値は、その後で、プロアクティブキャッシュエントリが古いと見なされ得、結果として、オンデバイス履行エンジン175によって利用されず、かつ/またはプリフェッチエンジン177によってプロアクティブキャッシュ178から除去される持続時間、またはしきい値時間期間を定義し得る。たとえば、プロアクティブキャッシュエントリのTTL値が、それが7日にわたって存続すべきであり、プロアクティブキャッシュエントリのタイムスタンプが、それが8日前に受信されたことを示す場合、プリフェッチエンジン177は、プロアクティブキャッシュ178からプロアクティブキャッシュエントリを除去し得る。これは、クライアントデバイス160の制限されたストレージリソースを自由にし、クライアントデバイス160の制限されたストレージリソースによりそのサイズ全体が制約され得る、他の時間的にプロアクティブなキャッシュエントリに対する空間をプロアクティブキャッシュ178内に作り出すことができる。

#### 【0042】

いくつかの実装形態では、プリフェッチエンジン177は、追加または代替として、プロアクティブキャッシュ178からプロアクティブキャッシュエントリを除去し、除去されたプロアクティブキャッシュエントリのうちの1つまたは複数は、それらのTTL値に基づいて、古いとして示されていないにもかかわらず、新しいプロアクティブキャッシュエントリに対する空間を作り出すことができる。たとえば、プロアクティブキャッシュ178は、たとえば、ユーザセットであり得かつ/またはクライアントデバイス160の記憶容量に基づいて決定され得る、最大サイズを有し得る。プリフェッチ要求からの新しいプロアクティブキャッシュエントリが最大サイズを超える場合、プリフェッチエンジン177は、新しいプロアクティブキャッシュエントリに対する空間を作り出すために、1つまたは複数の既存のキャッシュエントリを除去し得る。いくつかの実装形態では、プロアクティブキャッシュエントリは、キャッシュエントリのタイムスタンプに基づいて除去されてよく、たとえば、最も古いプロアクティブキャッシュエントリがプリフェッチエンジン177によって最初に除去され得る。いくつかの実装形態では、既存のプロアクティブキャッシュエントリ

10

20

30

40

50

に関するメタデータは、スコア、ランキング、または他の優先度データ(本明細書でより一般的にランキング基準とも呼ばれる)を含んでよく、最低優先度を有するものはプリフェッチエンジン177によって除去され得る。追加または代替として、プリフェッチエンジン177は、任意選択で、一度も利用されておらず、少なくともしきい値持続時間にわたってプロアクティブキャッシュ内に存在していたプロアクティブキャッシュエントリに対してバイアスするためにそれらのタイムスタンプをやはり考慮して、一度も利用されていないプロアクティブキャッシュエントリの除去に対してバイアスし得る。追加または代替として、プロアクティブキャッシュシステム120は、任意選択で、プリフェッチ要求に応じて、何の既存のプロアクティブキャッシュエントリが除去されるべきかの指示を提供し得る。

#### 【0043】

プロアクティブキャッシュシステム120は、プロアクティブキャッシュエントリを生成し、要求しているクライアントデバイスに対して選択されたプロアクティブキャッシュエントリを用いて、クライアントデバイス160および他のクライアントデバイスからのプリフェッチ要求に応じる。プロアクティブキャッシュシステム120はさらに、リモート自動アシスタント構成要素180によって利用され、各々が1つまたは複数のクライアントデバイスに固有であり得るリモートプロアクティブキャッシュ184を生成し得る。

#### 【0044】

プロアクティブキャッシュシステム120は、キャッシュエントリ生成エンジン130、キャッシュアセンブリエンジン140、およびエンティティイベントエンジン150を含み得る。概して、キャッシュエントリ生成エンジン130は、本明細書で説明する様々な技法を利用して、(図1および図2で、キャッシュ候補134と呼ばれる)大量の候補プロアクティブキャッシュエントリを生成する。キャッシュアセンブリエンジン140は、複数のクライアントデバイスの各々に対して、対応するクライアントデバイスのローカルプロアクティブキャッシュ内に記憶するために対応するクライアントデバイスに提供されるべきキャッシュ候補134の対応するサブセットを決定する。キャッシュアセンブリエンジン140はまた、任意選択で、各々がクライアントデバイスのサブセットに関連付けられ、各々がキャッシュ候補の対応するサブセットを含み、各々が対応するクライアントデバイスに対する履行においてリモート自動アシスタント構成要素180によって利用される、リモートプロアクティブキャッシュ184を生成し得る。エンティティイベントエンジン150は、任意選択で、1つまたは複数のリモートシステム151との対話を通して、様々なエンティティに関するイベントの発生を決定し得る。いくつかの実装形態では、エンティティイベントエンジン150は、キャッシュエントリ生成エンジン130にそのエンティティに対する1つまたは複数の対応するキャッシュ候補134を生成させるために、それらのイベントに関する情報をキャッシュエントリ生成エンジン130に提供し得る。いくつかの実装形態では、エンティティイベントエンジン150は、追加または代替として、それらのイベントキャッシュアセンブリエンジン140に関する情報を提供することができ、イベントキャッシュアセンブリエンジン140は、対応するクライアントデバイスにかつ/またはリモートプロアクティブキャッシュ184内に含めるために様々なキャッシュ候補を提供するかどうかを決定する際にその情報を使用し得る。

#### 【0045】

図1では、キャッシュエントリ生成エンジン130は、要求パラメータモジュール131、アクションコンテンツモジュール132、およびメタデータモジュール133を含む。要求パラメータモジュール131は、各プロアクティブキャッシュエントリに対するアシスタント要求パラメータを生成する。プロアクティブキャッシュエントリのアシスタント要求パラメータは、所与のアクションを実施するための1つまたは複数のアシスタント要求を表す。アシスタント要求は、所与のアクションの実施を要求する、タイプされたまたは口頭の発話であり得る。たとえば、複数のアシスタント要求は各々、以下のタイプされたまたは口頭の発話など、当日に局地予報をレンダリングするための要求であり得る:「今日の予報」、「今日の局地予報」、「今日の天気は何か」、および「天気はどうか」。要求パラメータモジュール131は、同じ所与のアクションを実施するために、それらの複数のアシス

10

20

30

40

50

タント要求をキャプチャする、テキスト表現および/またはNLU表現を生成することを求める。たとえば、前の発話の各々のテキストがアシスタント要求パラメータ、および/または「{intent=weather; location=local; date=today}」の構造化表現など、発話のすべてに共通するNLU表現として含まれ得る。

【0046】

アクションコンテンツモジュール132は、各プロアクティブキャッシュエントリに対するアクションコンテンツを生成する。アクションコンテンツは、プロアクティブキャッシュエントリによって異なり得る。プロアクティブキャッシュエントリに対するアクションコンテンツは、たとえば、実行されることになるディープリンク、レンダリングされることになるテキスト、グラフィックス、および/またはオーディオ、および/または周辺デバイスに送信されることになるコマンドを含み得る。

10

【0047】

当日の局地予報の例を続けると、アクションコンテンツモジュール132は、複数のキャッシュエントリの各々に対して異なるアクションコンテンツを生成し得、各々に対するアクションコンテンツは、異なる地理的領域に対して調整されている。たとえば、第1のプロアクティブキャッシュエントリは、第1の都市に対してであってよく、アシスタント要求パラメータ、およびレンダリングされることになるテキスト、グラフィックス、および/またはオーディオを含み、第1の都市に関する日々の予報を指定する、第1のアクションコンテンツを含んでよい。第2のプロアクティブキャッシュエントリは、第2の都市に対してであってよく、同じ(または、同様の)アシスタント要求パラメータを含むが、レンダリングされることになるテキスト、グラフィックス、および/またはオーディオを含み、第2の都市に関する日々の異種の予報を指定する、第2のアクションコンテンツを含んでよい。キャッシュアセンブリエンジン140に関して説明したように、これらの2つの異種のキャッシュエントリは、クライアントデバイスの属性に基づいて、異なるクライアントデバイスおよび/またはリモートプロアクティブキャッシュ184に提供され得る。たとえば、第1の都市の第1のクライアントデバイスには、第1のプロアクティブキャッシュエントリが提供され得るが、第2のプロアクティブキャッシュエントリは提供され得ない。

20

【0048】

いくつかの実装形態では、アクションコンテンツモジュール132は、複数のキャッシュエントリの各々に対して異なるアクションコンテンツを生成し得、各々に対するアクションコンテンツは、1つまたは複数の異なるアプリケーションに対して調整されている。たとえば、第1のプロアクティブキャッシュエントリは、「Hypothetical Artistを再生する」、「あるHypothetical Artistを聴く」、および/または「{intent=listen to music; artist=Hypothetical Artist}」のアシスタント要求パラメータを含み得る。第2のプロアクティブキャッシュエントリは、同じ(または、同様の)アシスタント要求パラメータを含み得る。それでもなお、アクションコンテンツモジュール132は、第1のプロアクティブキャッシュエントリに対して、実行されたとき、Hypothetical Artistの音楽がストリーミングを始める状態で第1の音楽アプリケーションを開かせるディープリンクを含む第1のアクションコンテンツを生成し得る。アクションコンテンツモジュール132は、第2のプロアクティブキャッシュエントリに対して、実行されたとき、Hypothetical Artistの音楽がストリーミングを始める状態で第2の音楽アプリケーションを開かせる異なるディープリンクを含む第2のアクションコンテンツを生成し得る。キャッシュアセンブリエンジン140に関して説明したように、これらの2つの異種のキャッシュエントリは、クライアントデバイスの属性に基づいて、異なるクライアントデバイスおよび/またはリモートプロアクティブキャッシュ184に提供され得る。すなわち、クライアントデバイスが、どのアプリケーションをインストールしたかかつ/または音楽ストリーミングのための好ましいアプリケーションとして示したかに基づく。たとえば、唯一の音楽ストリーミングアプリケーションとして第1のアプリケーションを有する第1のクライアントデバイスには、第1のプロアクティブキャッシュエントリが提供され得るが、第2のプロアクティブキャッシュエントリは提供され得ない。

30

40

50

## 【 0 0 4 9 】

いくつかのアシスタント要求パラメータの場合、単一のプロアクティブキャッシュエントリのみが存在し得る。たとえば、Cavalier King Charles Spaniel(犬の種類)の画像を求めるアシスタント要求に関するアシスタント要求パラメータの場合、Cavalier King Charles Spanielの画像を有するアクションコンテンツを含む単一のプロアクティブキャッシュエントリが提供され得る。

## 【 0 0 5 0 】

メタデータモジュール133は、任意選択で、キャッシュエントリに関するメタデータを生成する。メタデータのうちのいくつかは、任意選択で、クライアントデバイスに送信される必要なしに、プロアクティブキャッシュシステム120によって利用され得る。たとえば、メタデータモジュール133は、プロアクティブキャッシュエントリに関連する1つまたは複数のエンティティを示すプロアクティブキャッシュエントリ、そのプロアクティブキャッシュエントリに対するアクションコンテンツの言語、および/またはそのプロアクティブキャッシュエントリに関する他のデータに関するメタデータを生成し得る。キャッシュアセンブリエンジン140は、どのクライアントデバイスにプロアクティブキャッシュエントリが提供されるべきかを決定する際、そのようなメタデータを利用し得る。たとえば、アクションコンテンツとして、第1の都市に関する局地天気予報を含むプロアクティブキャッシュエントリの場合、メタデータモジュール133は、第1の都市を示すメタデータを生成し得る。キャッシュアセンブリエンジン140は、現在のまたは好ましいロケーションとして第1の都市を有するクライアントデバイスに対してのみローカルまたはリモートプロアクティブキャッシュ内に含めるためにプロアクティブキャッシュエントリを選択する際にそのようなメタデータを利用し得る。また、たとえば、アクションコンテンツとして、女優に関するグラフィックスおよび/またはテキストを含むプロアクティブキャッシュエントリの場合、メタデータモジュール133は、その女優を示すメタデータを生成し得る。キャッシュアセンブリエンジン140は、その女優に対応する属性を有するクライアントデバイスに対してのみローカルまたはリモートプロアクティブキャッシュ内に含めるためにプロアクティブキャッシュエントリを選択する際にそのようなメタデータを利用し得る。たとえば、有名人に関するコンテンツを前に閲覧したことに基づいて、その女優を示す属性を有する、かつ/またはその女優を含むストーリーミング映画またはテレビジョン番組に基づいて、その女優が主演した映画/番組を示す属性を有する、クライアントデバイス。

## 【 0 0 5 1 】

メタデータモジュール133はまた、プロアクティブキャッシュエントリとともにクライアントデバイスに送信される、かつ/またはリモートプロアクティブキャッシュ内で保持する際に利用されるメタデータを生成し得る。たとえば、メタデータモジュール133は、それがいつ生成されたかかつ/または最後に検証されたかを示すプロアクティブキャッシュエントリ(たとえば、アクションコンテンツの精度の検証)に対してタイムスタンプを生成し得る。メタデータモジュール133はまた、プロアクティブキャッシュエントリに対するTTL値を生成し得る。所与のプロアクティブキャッシュエントリに対するTTL値は、アシスタント要求パラメータおよび/またはアクションコンテンツの特性など、様々な考慮事項に基づいて生成され得る。たとえば、天気に関するアクションコンテンツなど、いくつかのアクションコンテンツは動的であり、そのようなアクションコンテンツを有するプロアクティブキャッシュエントリは、比較的短いTTL(たとえば、6時間、12時間)を有し得る。他方で、いくつかのアクションコンテンツは静的であり、そのようなアクションコンテンツを有するプロアクティブキャッシュエントリは、比較的長いTTL(たとえば、7日、14日、30日)を有し得る。別の例として、静的コンテンツを含むが、エンティティイベントエンジン150によるイベント検出に基づいて提供および/または生成されるプロアクティブキャッシュエントリは、静的コンテンツを含むが、エンティティイベントエンジン150によるイベント検出とは無関係に提供および/または生成されるプロアクティブキャッシュエントリよりも短いTTLを有し得る。

## 【 0 0 5 2 】

10

20

30

40

50

図1では、キャッシュアセンブリエンジン140は、ローカルモジュール141およびリモートモジュール142を含む。ローカルモジュール141は、キャッシュ候補134から、各クライアントデバイスに対して、クライアントデバイスに提供するためのキャッシュ候補134の対応するサブセットを選択する。ローカルモジュール141は、プロアクティブキャッシュエントリの属性に対するクライアントデバイスの属性の比較、および/またはプロアクティブキャッシュエントリに対するランキング基準など、様々な考慮事項に基づいて、所与のクライアントデバイスに提供するためのサブセットを決定し得る。

#### 【0053】

たとえば、所与のクライアントデバイスに対するサブセットを選択する際、ローカルモジュール141は、そのメタデータが以下を示す任意のプロアクティブキャッシュエントリをフィルタで除去し得る:そのアクションコンテンツが所与のクライアントデバイス上にインストールされていない対応するアプリケーションのみに適用可能である(たとえば、そのアプリケーションのみに関するディープリンク)、そのアクションコンテンツが所与のクライアントデバイスに関連しない地理的領域のみに適用可能である、そのアクションコンテンツが所与のクライアントデバイスに対する1次(また任意選択で、2次)言語として設定されていない言語のみである。残りのプロアクティブキャッシュエントリは、所与のクライアントデバイスの属性に対するそれらの属性の比較、残りのプロアクティブキャッシュエントリに対するランキング基準(やはり属性と見なされ得る)、および/または他の考慮事項に基づいて、選択され得る。

#### 【0054】

たとえば、ローカルモジュール141は、所与のクライアントデバイスによって過去に対話されたと決定される1つまたは複数のエンティティへの対応に失敗する代替エンティティを示すメタデータを有するプロアクティブキャッシュエントリと比較して、所与のクライアントデバイスによって過去に対話されたと決定される1つまたは複数のエンティティに対応する1つまたは複数のエンティティを示すメタデータを有するプロアクティブキャッシュエントリを選択するより高い可能性があり得る。また、たとえば、所与のプロアクティブキャッシュエントリに対するランキング基準は、それらの対応するアシスタント要求がアシスタントインターフェースを介してどの程度の頻度で(全体的に、または所与のクライアントデバイスに対して)提出されるか、かつ/またはそれらの対応するアクションコンテンツがどの程度の頻度で(全体的に、または所与のクライアントデバイスに対して)レンダリングされるかを示し得る。アクションコンテンツをレンダリングすることは、アクションコンテンツの任意の関連するテキスト/グラフィカル/可聴コンテンツをそのアクションを実施するためにクライアントデバイスにおいてレンダリングさせることを含み得る。ディープリンクを実行することは、関連するアクション(たとえば、アプリケーションを特定の状態に開くこと)を自動的に実施することを含み得るか、またはクライアントデバイスを関連するアクションを実施するために準備状態にさせることを含み得、ここで、実施は、たとえば、ユーザインターフェース入力に応じてよい。ランキング基準は、エンティティイベントエンジン150による、所与のプロアクティブキャッシュエントリに関する最近のイベント検出に基づいてもよい。たとえば、所与のプロアクティブキャッシュが、新しい周辺デバイス(たとえば、新しいスマートサーモスタット)をインストールするためのアシスタント要求パラメータおよびインストール指示を含むアクションコンテンツを含むと仮定する。エンティティイベントエンジン150が新しい周辺デバイスに関する要求(アシスタント要求、検索エンジン要求、または他の要求)のかなりの増大および/または新しい周辺デバイスに関するコンテンツ(たとえば、ウェブページ、ソーシャルメディアコメント)のかなりの増大を決定する場合、対応するランキング基準は、より高いランキングを示し、プロアクティブキャッシュエントリが選択される可能性をより高くすることができる。

#### 【0055】

所与のクライアントデバイスに対するプロアクティブキャッシュエントリを決定する際、ローカルモジュール141は、所与のクライアントデバイスに対してプロアクティブキャッシュに割り振られているストレージスペースを考慮することもできる。さらに、所与の

10

20

30

40

50

クライアントデバイスが既存のプロアクティブキャッシュエントリをすでに含むときに、所与のクライアントデバイスに対して新しいプロアクティブキャッシュエントリを提供するかどうかを決定するとき、既存のプロアクティブキャッシュエントリが考慮され得る。たとえば、既存のプロアクティブキャッシュエントリに対するランキング基準が考慮され得、かつ/または所与のクライアントデバイスの既存のプロアクティブキャッシュの占有されていないストレージスペース(もしあれば)が考慮され得る。

**【0056】**

リモートモジュール142は、任意選択で、1つまたは複数のリモートプロアクティブキャッシュ184を生成し保持する。リモートプロアクティブキャッシュ184は各々、クライアントデバイスのサブセットに対する。リモートプロアクティブキャッシュ184のサブセットは、単一のクライアントデバイスであり得るか、または共通する1つまたは複数の(たとえば、すべての)属性を共有するクライアントデバイスのコレクションであり得る。たとえば、リモートプロアクティブキャッシュに対するクライアントデバイスのコレクションは、同じ地理的領域内にある、同じアプリケーションをインストールしている、かつ/またはその過去の対話が少なくともしきい値量の同じエンティティに共通の関心を示すクライアントデバイスであり得る。クライアントデバイスに対するリモートプロアクティブキャッシュ184の場合、リモートモジュール142は、フィルタで除去されなかった、かつローカルプロアクティブキャッシュ内の記憶のためにまだ提供されていなかった、1つまたは複数のキャッシュ候補134を選択し得る。たとえば、クライアントデバイス160に対して、プロアクティブキャッシュ178は500MB制限を有すると仮定する。ローカルモジュール141は、プロアクティブキャッシュ178に対して、500MBに相当するプロアクティブキャッシュエントリをすでに選択し提供しているとさらに仮定する。リモートモジュール142は、その場合、クライアントデバイス160に対してリモートプロアクティブキャッシュ184内に含めるためのローカルキャッシュモジュール141と同じ考慮事項を利用して、追加のプロアクティブキャッシュエントリを選択し得る。たとえば、クライアントデバイス160に対するリモートプロアクティブキャッシュ184は、2GB限度を有することがあり、リモートモジュール142は、プロアクティブキャッシュエントリおよびクライアントデバイス160の属性を比較し、かつ/またはプロアクティブキャッシュエントリに対するランキング基準を考慮することによって、2GB相当の残りのプロアクティブキャッシュエントリを選択し得る。

**【0057】**

エンティティイベントエンジン150は、様々なエンティティに関するイベント発生を監視する際、1つまたは複数のリモートシステム151と対話する。エンティティに対するイベントを決定するいくつかの例は、エンティティに関する要求の増大を決定すること、エンティティに関するインターネットコンテンツの増大を決定すること、および/またはエンティティに関する要求の増大を予測すること、である。たとえば、エンティティイベントエンジン150は、アシスタント要求(および/または旧来の検索要求)の量が特定のルータに対してスパイクしたかどうか、かつ/またはその特定のルータに対してインターネットコンテンツにスパイクが存在するかどうかを決定し得る。それに応じて、エンティティイベントエンジン150は、スパイクに関する基準をキャッシュアセンブリエンジン140に提供することができ、キャッシュアセンブリエンジン140は、ローカルプロアクティブキャッシングのために、かつ/またはリモートプロアクティブキャッシングのために、その特定のルータに関するプロアクティブキャッシュエントリを様々なクライアントデバイスに提供する(スパイクの前よりも)より高い可能性があり得る。様々なクライアントデバイスにプロアクティブキャッシュエントリを提供することは、それらの様々なクライアントデバイスがそのルータに関して対応する属性(たとえば、その特定のルータまたはルータ一般に対する過去の検索)を有すると決定することにさらに基づき得る。

**【0058】**

エンティティイベントエンジン150は、スパイクの指示をキャッシュエントリ生成エンジン130に提供することもできる。応答して、キャッシュエントリ生成エンジン130は、

任意選択で、特定のルータに対して1つまたは複数のプロアクティブキャッシュエントリを生成し得る。キャッシュエントリ生成エンジン130は、任意選択で、特定のルータに関する現在のキャッシュ候補134が存在しない、かつ/または特定のルータに関するしきい値未満のキャッシュ候補134を決定することに基づいて、その特定のルータに対して1つまたは複数のプロアクティブキャッシュエントリを生成し得る。たとえば、キャッシュ生成エンジン130は、特定のルータのクラス(たとえば、ルータの一般的なクラス)を決定し、そのクラスのエンティティに対する頻繁なクエリに対するテンプレートを決定することができる。たとえば、(他の特定のルータエイリアスに関する関連クエリに基づく)「[ルータエイリアス]に対する最大帯域幅は何か」または(他の特定のルータエイリアスに関する関連クエリに基づく)「[ルータエイリアス]に対するデフォルトIPアドレスは何か」のテンプレート。「ルータエイリアス」を特定のパラメータに対するエイリアスに置換することに基づいて、プロアクティブキャッシュエントリに対するアシスタント要求パラメータが次いで生成され得る。さらに、「ルータエイリアス」を特定のルータに対するエイリアスと置換するクエリに対する上位検索結果からのスニペットに基づいて、かつ/または他の技法を使用して、プロアクティブキャッシュエントリに対するアクションコンテンツが生成され得る。たとえば、「特定のルータに対するデフォルトIPアドレスは何か」のアシスタント要求に対して、「192.168.1.1」のアクションコンテンツが生成され得る。

#### 【0059】

いくつかの実装形態では、エンティティイベントエンジン150は、既存のキャッシュエントリのアクションコンテンツを示すイベントは古いと決定し、キャッシュエントリ生成エンジン130に、更新されたアクションコンテンツを反映するための新しいキャッシュエントリを生成させる、かつ/または古いコンテンツを有する既存のキャッシュエントリを除去させる指示をキャッシュエントリ生成エンジン130に提供し得る。本明細書で使用するように、更新されたアクションコンテンツを反映するために新しいプロアクティブキャッシュエントリを生成することは、プロアクティブキャッシュエントリのアシスタント要求パラメータを保持しながら、新しいアクションコンテンツ(および任意選択で、更新されたメタデータ)を反映するために既存のプロアクティブキャッシュエントリを更新することを含み得る。それはまた、既存のプロアクティブキャッシュエントリを完全に除去して、同じアシスタント要求パラメータを用いるが、更新されたアクションコンテンツ(および任意選択で、更新されたメタデータ)で、新しいプロアクティブキャッシュエントリを生成することを含み得る。一例として、エンティティイベントエンジン150は、地理的領域に対する天気予報が少なくともしきい値量だけ変更されたと決定し、結果として、新しい対応するプロアクティブキャッシュエントリを生成させることができる。

#### 【0060】

いくつかの実装形態では、リモート自動アシスタント構成要素180は、音声認識を実施するリモートARSエンジン181、自然言語理解を実施するリモートNLUエンジン182、および/または任意選択で、本明細書で説明するようリモートプロアクティブキャッシュ184を利用して、履行データを生成するリモート履行エンジン183を含み得る。任意選択で、ローカルにまたはリモートで決定された履行データに基づいて、リモート実行を実施するリモート実行モジュールが含まれてもよい。追加および/または代替として、リモートエンジンが含まれ得る。本明細書で説明するように、様々な実装形態では、オンデバイス音声処理、オンデバイスNLU、オンデバイス履行、および/またはオンデバイス実行は、(口頭の発話を分析するためにクライアント-サーバラウンドトリップが必要とされないため)口頭のおよび/またはタイプされた発話を分析するときの、少なくともこれらが提供するレイテンシおよび/またはネットワーク使用低減により、優先され得る。しかしながら、1つまたは複数のクラウドベースの自動アシスタント構成要素180は、少なくとも選択的に利用され得る。たとえば、そのような構成要素は、オンデバイス構成要素と並行して利用され、ローカル構成要素が失敗するとき利用されるそのような構成要素から出力され得る。たとえば、オンデバイス履行エンジン175は、一定の状況で(たとえば、サイズ制約されているプロアクティブキャッシュ178がマッチするプロアクティブキャッシュエントリを含

10

20

30

40

50

めることに失敗するとき)失敗することがあり、リモート履行エンジン183は、そのような状況で、よりロバスタなりモートプロアクティブキャッシュ184(または、リモートプロアクティブキャッシュがマッチを有さないとき、追加のリソース)を利用して、履行データを生成し得る。リモート履行エンジン184は、オンデバイス履行エンジン175と並行して動作することができ、その結果は、オンデバイス履行が失敗したときに利用されるか、またはオンデバイス履行の失敗の決定に応じて起動され得る。

#### 【0061】

様々な実装形態では、NLUエンジン(オンデバイスおよび/またはリモート)は、認識されたテキストの1つまたは複数の注釈および自然言語入力の用語のうちの一つまたは複数(たとえば、すべて)を含む注釈付き出力を生成し得る。いくつかの実装形態では、NLUエンジンは、自然言語入力における様々なタイプの文法情報を識別して注釈付けするように構成される。たとえば、NLUエンジンは、個々の語を形態素に分け、かつ/または形態素を、たとえば、それらのクラスで注釈付けすることができる形態素モジュール(morphological module)を含み得る。NLUエンジンはまた、それらの文法的な役割で用語に注釈付けするように構成された音声タグの一部を含み得る。また、たとえば、いくつかの実装形態では、NLUエンジンは、追加および/または代替として、自然言語入力における用語同士の間の構文関係を決定するように構成された依存性パーサ(dependency parser)を含み得る。

#### 【0062】

いくつかの実装形態では、NLUエンジンは、追加および/または代替として、人々(たとえば、文学上の人物、有名人、公人などを含む)、組織、ロケーション(実際のおよび架空の)などの参照など、1つまたは複数のセグメント内のエンティティ参照に注釈付けするように構成されたエンティティタグを含み得る。いくつかの実装形態では、NLUエンジンは、追加および/または代替として、1つまたは複数のコンテキストキューに基づいて、同じエンティティの参照をグループ化、または「クラスタ化」するように構成された共参照レゾルバ(図示せず)を含み得る。いくつかの実装形態では、NLUエンジンの1つまたは複数の構成要素は、NLUエンジンの1つまたは複数の構成要素からの注釈に依拠し得る。

#### 【0063】

NLUエンジンはまた、自動アシスタント195との対話に関与したユーザの意図を決定するように構成された意図マッチャを含み得る。意図マッチャは、ユーザの意図を決定するために様々な技法を使用し得る。いくつかの実装形態では、意図マッチャは、たとえば、文法と応答意図との間の複数のマッピングを含む、1つまたは複数のローカルおよび/またはリモートデータ構造に対するアクセスを有し得る。たとえば、マッピング内に含まれる文法は、経時的に選択および/または学習され得、ユーザの共通の意図を表し得る。文法に加えてまたはその代わりに、いくつかの実装形態では、意図マッチャは、単独で、または1つまたは複数の文法と組み合わせて、1つまたは複数のトレーニングされた機械学習モデルを採用し得る。これらのトレーニングされた機械学習モデルは、たとえば、口頭の発話から認識されたテキストを低減された次元空間内に埋め込み、次いで、どの他の埋込(および、したがって、意図)が最も近接するかを、たとえば、ユークリッド距離、コサイン類似度などの技法を使用して決定することによって、意図を識別するようにトレーニングされ得る。いくつかの文法は、スロット値で充填することができるスロット(たとえば、アーティスト)を有する。スロット値は、様々な方法で決定され得る。ユーザはしばしば、スロット値をプロアクティブに提供することになる。たとえば、「トッピング ピザを私に注文してください」という文法の場合、ユーザは、「ソーセージピザを私に注文してください」という句を話す可能性があり得、その場合、スロット トッピング は自動的に充填される。たとえば、ユーザロケーション、現在レンダリングされているコンテンツ、ユーザ選好、および/または他のキューに基づいて、他のスロット値を推論することができる。ローカルに実装され得る、本明細書で説明するような意図マネージャの使用は、クライアントデバイスに対するユーザインターフェース入力(キャッシュエントリの1つまたは複数のアシスタント要求パラメータに対するマッチ(任意選択で、完全なマッチ)でない

10

20

30

40

50



場合ですら、プロアクティブキャッシュエントリが取り出されることを可能にし得る。これは、デバイスの効用を改善し得る。

【0064】

次に図2を参照すると、様々な実装形態に従って図1の様々な構成要素がどのように対話し得るかを説明する例示的なプロセスフローが示される。

【0065】

図2では、プリフェッチエンジン177は、要求221をプロアクティブキャッシュシステム120に送る。プロアクティブキャッシュシステム120は、プロアクティブキャッシュエントリ222でその要求に回答する。本明細書で説明するように、プロアクティブキャッシュエントリ222は、キャッシュ候補のサブセットであり得、クライアントデバイス160の属性、キャッシュエントリ222の属性、キャッシュエントリ222に対するランキング基準に基づいて、かつ/またはプロアクティブキャッシュ178内にすでにあるプロアクティブキャッシュエントリに基づいて、クライアントデバイス160に対して選択され得る。プリフェッチエンジン177は、任意選択で、1つまたは複数の既存のプロアクティブキャッシュエントリを除去して、キャッシュエントリ222に対する空間を作って、キャッシュエントリ222をプロアクティブキャッシュ178内に記憶する。

【0066】

キャッシュエントリ222をプロアクティブキャッシュ178内に記憶した後(たとえば、数分または数時間後)の時点で、クライアントデバイス160(図1)のマイクロフォン165(図1)を介して、オーディオデータ223が検出される。検出されたオーディオデータ223は、クライアントデバイスに対するユーザインターフェース入力の一例である。オンデバイス音声認識モジュールは、オーディオデータ223を処理して、認識されたテキスト171Aを生成する。

【0067】

認識されたテキスト171Aは、任意選択で、オンデバイス履行エンジン175に、かつ/またはリモート履行エンジン183に提供され得る。認識されたテキスト171Aがオンデバイス履行エンジン175に提供され、オンデバイス履行エンジン175が、そのテキストがプロアクティブキャッシュ178のプロアクティブキャッシュエントリのアシスタント要求パラメータにマッチすると決定するとき、オンデバイス履行エンジン175は、マッチするプロアクティブキャッシュエントリの少なくともアクションコンテンツを含む履行データ175Aを生成し得る。

【0068】

認識されたテキスト171Aを考慮することに加えて、またはその代わりに、オンデバイス履行エンジン175は、認識されたテキスト171Aの処理に基づいて(かつ、任意選択で、コンテキストデータに基づいて)、オンデバイスNLUエンジン174によって生成されたNLUデータ174Aを考慮し得る。NLUデータ174Aがオンデバイス履行エンジン175に提供され、オンデバイス履行エンジン175が、そのテキストがプロアクティブキャッシュ178のプロアクティブキャッシュエントリのアシスタント要求パラメータにマッチすると決定するとき、オンデバイス履行エンジン175は、マッチするプロアクティブキャッシュエントリの少なくともアクションコンテンツを含む履行データ175Aを生成し得る。履行データは、NLUデータの意図が前記アシスタント要求パラメータにマッチする場合、生成されてもよい。

【0069】

オンデバイス実行エンジン176は、マッチするプロアクティブキャッシュエントリの実行アクションコンテンツを含む(または、それに制限される)オンデバイス履行データ175Aを処理し、対応するアクションを実施し得る。これは、アクションコンテンツに基づいて、可聴、視覚、および/または触覚応答を生成すること、アクションコンテンツのディープリンクを実行すること、および/またはアクションコンテンツ内に含まれたコマンドを(たとえば、BluetoothまたはWi-Fiを介して)送信すること、を含み得る。

【0070】

10

20

30

40

50

いくつかの実装形態では、認識されたテキスト171Aは、リモート履行エンジン183に提供される。認識されたテキスト171Aは、オンデバイス履行エンジン175に対するプロビジョニングと並行して、または任意選択で、オンデバイス履行エンジン175によるプロアクティブキャッシュ178内に何のマッチングエントリも存在しないとの(認識されたテキスト171Aおよび/またはNLUデータ174Aに基づく)決定のみに応じて、リモート履行エンジン183に提供され得る。リモート履行エンジン183は、クライアントデバイス160に割り当てられたリモートプロアクティブキャッシュ184Aにアクセスし、そのリモートプロアクティブキャッシュが、認識されたテキスト171Aおよび/または認識されたテキスト171Aに関してリモートで決定されたNLUデータにマッチするアシスタント要求パラメータを備えるプロアクティブキャッシュエントリを含むかどうかを決定し得る。含む場合、リモート履行エンジン183は、任意選択で、マッチするリモートプロアクティブキャッシュエントリからのアクションコンテンツを含むリモート履行データ183Aをオンデバイス実行エンジン176に提供し得る。任意選択で、リモート履行エンジン183は、(オンデバイス履行エンジン175による)ローカル履行が失敗したという指示に応じて、かつ/またはクライアントデバイス160からの(ローカル履行が成功するときに提供され得る)「停止」コマンドの受信の欠如に応じたのみ、リモート履行データ183Aを提供する。リモート履行エンジン183は、任意選択で、履行データ183A(たとえば、アクションコンテンツ)を生成するための他の技法を「オンザフライ」で利用してもよい。これは、機械リモートプロアクティブキャッシュエントリが存在するかどうかを決定するためにリモートプロアクティブキャッシュ184Aにアクセスすることと並行して行われ得る、かつ/またはマッチするリモートプロアクティブキャッシュエントリが存在しないとの決定に応じた実施され得る。リモートプロアクティブキャッシュ184A内のアクションコンテンツはすでに事前に生成されているため、リモート履行データは、リモート履行エンジン183によって、そのコンテンツがオンザフライで生成された場合よりも少ないリソースを利用して、より急速に取得され得る。したがって、様々な実装形態は、マッチングリモートプロアクティブキャッシュエントリが識別される場合、任意選択で、履行データをオンザフライで生成する試みなしに、かつ/またはマッチが決定された場合、オンザフライ生成を停止して、リモートプロアクティブキャッシュ184Aのみに基づいて、リモート履行データを提供し得る。

#### 【0071】

図3は、プロアクティブキャッシュエントリ310、320、および330のいくつかの非限定的な例を示す。そのようなプロアクティブキャッシュエントリは、大量の追加エントリとともに、ローカルプロアクティブキャッシュ178(図1)内に、またはリモートプロアクティブキャッシュ184(図1)のうちの1つの中に、記憶され得る。

#### 【0072】

プロアクティブキャッシュエントリ310は、明日の局地天気予報を取得する所与のアクションを実施するための様々なアシスタント要求を表す要求パラメータ310Aを含む。要求パラメータ310Aは、「明日の天気」、「天気明日」のテキスト表現、ならびに「天気」の意図ならびに「日」スロットに対する「明日」および「ロケーション」スロットに対する「局地」のスロット値を指定する構造化NLUデータ表現を含む。プロアクティブキャッシュエントリ310のアクションコンテンツ310Bは、明日の局地天気を記述するテキストならびに明日の局地天気を伝えるグラフィックを含む。テキストとグラフィックは両方とも、ユーザ入力(たとえば、口頭の発話またはタイプされた発話)が要求パラメータ310Aにマッチするとの決定に応じたレンダリングされ得る。任意選択で、テキストに基づく合成音声もそれに応じてレンダリングされ得る。プロアクティブキャッシュエントリ310のメタデータ310Cは、12時間のTTLおよびタイムスタンプを含む。本明細書で説明するように、プロアクティブキャッシュエントリ310は、TTLが満了したと決定されると、プロアクティブキャッシュから除去され得る(または、少なくとももはや利用され得ない)。本明細書でやはり説明するように、プロアクティブキャッシュエントリ310は、所与の地理的エリア内の所与のクライアントデバイスに対して提供され得、一方、同じ要求パラメータを有するが、異なるアクションコンテンツを有する、他のプロアクティブキャッシュエ

10

20

30

40

50

ントリは、他の地理的エリア内の他のクライアントデバイスに対して提供され得る。

【0073】

プロアクティブキャッシュエントリ320は、対応するアプリケーションのサーモスタットスケジュール調整状態にアクセスする所与のアクションを実施するための様々なアシスタント要求を表す要求パラメータ320Aを含む。要求パラメータ320Aは、「サーモスタットスケジュールを調整する」、「サーモスタットに対するスケジュールを変更する」のテキスト表現、ならびに「サーモスタット」の意図ならびに「設定」スロットに対する「スケジュールを変更/調整する」のスロット値を指定する構造化NLUデータ表現を含む。プロアクティブキャッシュエントリ320のアクションコンテンツ320Bは、特定のアプリケーションに対するディープリンクを含む。ディープリンクは、ユーザ入力(たとえば、口頭の発話またはタイプされた発話)が要求パラメータ320Aにマッチするとの決定に応じて実行され得る。ディープリンクを実行することは、サーモスタットスケジュール設定が調整され得る状態で特定のアプリケーションを開かせる(すなわち、アプリケーションは、次のユーザ入力アクションを実施させ得る状態にある)。プロアクティブキャッシュエントリ320のメタデータ320Cは、30日のTTLおよびタイムスタンプを含む。本明細書で説明するように、プロアクティブキャッシュエントリ320は、TTLが満了したと決定されると、プロアクティブキャッシュから除去され得る(または、少なくとももはや利用され得ない)。本明細書でやはり説明するように、プロアクティブキャッシュエントリ320は、クライアントデバイスが、1次サーモスタットアプリケーションとしてインストールおよび/または示された特定のアプリケーション(アクションコンテンツ320Bのディープリンクに対応する)を有すると決定することに基づいて、所与のクライアントデバイスに対して提供され得、一方、同じ要求パラメータを有するが、異なるアクションコンテンツ(たとえば、異なるディープリンク)を有する、他のプロアクティブキャッシュエントリは、特定のアプリケーションをインストールしていない他のクライアントデバイスに対して提供され得る。

【0074】

プロアクティブキャッシュエントリ330は、「John Doe」(仮説的人物)に対して推定された純資産を取得する所与のアクションを実施するための様々なアシスタント要求を表す要求パラメータ330Aを含む。要求パラメータ330Aは、「John Doeの純資産」および「John Doeの財産はどの程度」のテキスト表現を含む。プロアクティブキャッシュエントリ330のアクションコンテンツ330Bは、John Doeの純資産を記述するテキストを含む。テキストは、ユーザ入力(たとえば、口頭の発話またはタイプされた発話)が要求パラメータ330Aにマッチするとの決定に応じてレンダリングされ得る。任意選択で、テキストに基づく合成音声もそれに応じてレンダリングされ得る。プロアクティブキャッシュエントリ330のメタデータ330Cは、7日のTTLおよびタイムスタンプを含む。本明細書で説明するように、プロアクティブキャッシュエントリ330は、TTLが満了したと決定されると、プロアクティブキャッシュから除去され得る(または、少なくとももはや利用され得ない)。本明細書でやはり説明するように、プロアクティブキャッシュエントリ330は、所与のクライアントデバイスの属性がJohn Doeに関する決定することに基づいて、所与のクライアントデバイスに対して提供され得る。これは、たとえば、John Doeに対する過去のユーザ検索、John Doeに関連するインターネットコンテンツへの訪問、および/または(たとえば、知識グラフまたは他のデータ構造に基づいて決定されるような)John Doeとの強い関係を有する他のエンティティに対する検索に基づいてよい。様々な実装形態では、プロアクティブキャッシュエントリ330は、John Doeに関する要求および/またはインターネットコンテンツの増大など、John Doeに関するイベントの決定に少なくとも部分的に基づいて、生成および/または提供され得る。たとえば、John Doeの純資産に関するプロアクティブキャッシュエントリは、John Doeが有名人であり(クラス)、有名人に対する頻繁なクエリがテンプレート「[有名人エイリアス]の純資産は何か」を有すると決定することに基づいて生成され得る。また、たとえば、プロアクティブキャッシュエントリは、John Doeに関する要求および/またはインターネットコンテンツの増大の影響を受けるランキング基準に基づいて、ローカルまたはリモートプロアクティブキャッシュエントリ内

10

20

30

40

50

に記憶するために提供され得る。

【0075】

図4は、本明細書で開示する実装形態に従ってプロアクティブキャッシュエントリをプリフェッチして記憶する例示的な方法400を示すフローチャートを示す。便宜上、方法400の動作は、それらの動作を実施するシステムを参照しながら説明される。このシステムは、クライアントデバイスの1つまたは複数の構成要素(たとえば、図1のプリフェッチエンジン177)など、様々なコンピュータシステムの様々な構成要素を含み得る。さらに、方法400の動作は特定の順序で示されるが、これは限定されることを意味しない。1つまたは複数の動作は、並べ替えられてよく、省かれてよく、または追加されてもよい。

【0076】

ブロック410において、システムは、1つまたは複数のプリフェッチ条件が発生したかどうかを決定する。発生していない場合、システムは、プリフェッチ条件が発生したかどうかを決定し続ける。発生した場合、システムはブロック420に進む。プリフェッチ条件は、たとえば、クライアントデバイスに対して一定のネットワーク条件が存在すること、クライアントデバイスが充電していることおよび/または少なくともしきい値バッテリー充電状態を有すること、クライアントデバイスがユーザによってアクティブに利用されていないこと、クライアントデバイスの現在のプロセッサ使用率および/または現在のメモリ使用率が一定のしきい値を超えていないこと、および/または最近のプリフェッチ要求から一定の時間量(たとえば、少なくとも1時間)が経過したこと、のうちの1つまたは複数を含み得る。

【0077】

ブロック420において、システムは、プリフェッチ要求を送信する。プリフェッチ要求は、任意選択で、クライアントデバイスの識別子、および/またはクライアントデバイスにおいてすでにローカルに記憶されているプロアクティブキャッシュエントリを示すトークンまたは他の指示を含み得る。

【0078】

ブロック430において、システムは、ブロック420の要求に応じて、プロアクティブアシスタントキャッシュエントリを受信する。

【0079】

ブロック440において、システムは、受信されたプロアクティブアシスタントキャッシュエントリをローカルプロアクティブキャッシュ内に記憶する。ブロック440は、任意選択で、ブロック440Aを含んでよく、ここで、システムは、1つまたは複数の既存のプロアクティブキャッシュエントリをローカルプロアクティブキャッシュから除去して、受信されたプロアクティブキャッシュエントリに対する空間を作る。ブロック440の後、システムは、任意選択で、しきい値時間量が経過した後、ブロック410に再度進むことができる。

【0080】

図5Aは、プロアクティブキャッシュエントリを生成し、プロアクティブキャッシュエントリのローカルサブセットを所与のクライアントデバイスにプロビジョニングし、かつ/またはプロアクティブキャッシュエントリのリモートサブセットを決定する例示的な方法500を示すフローチャートを示す。図5Bは、図5Aのブロック510のいくつかの実装形態を示すフローチャートを示す。図5Cは、図5Aのブロック510のいくつかの追加または代替の実装形態を示すフローチャートを示す。便宜上、方法500の動作は、それらの動作を実施するシステムを参照しながら説明される。このシステムは、リモートサーバの1つまたは複数の構成要素(たとえば、図1のプロアクティブキャッシュシステム)など、様々なコンピュータシステムの様々な構成要素を含み得る。さらに、方法500の動作は特定の順序で示されるが、これは限定されることを意味しない。1つまたは複数の動作は、並べ替えられてよく、省かれてよく、または追加されてもよい。

【0081】

最初に図5Aから開始して、ブロック510において、システムは、プロアクティブアシス

10

20

30

40

50

タントキャッシュエントリを生成する。各プロアクティブアシスタントキャッシュエントリを生成する際、ブロック510は、サブブロック512、514、および/または516を含み得る。ブロック512において、システムは、プロアクティブアシスタントキャッシュエントリに対するアシスタント要求パラメータを生成する。ブロック514において、システムは、プロアクティブアシスタントキャッシュエントリに対するアクションコンテンツを生成する。ブロック516において、システムは、プロアクティブアシスタントキャッシュエントリに対するメタデータを生成する。ブロック510の実装形態の追加の説明は、下記で、図5Bおよび図5Cに関して、ならびに本明細書の他の場所で提供される。

【0082】

ブロック520において、システムは、クライアントデバイスを選択する。

10

【0083】

ブロック530において、システムは、選択されたクライアントデバイスに対するプロアクティブアシスタントキャッシュエントリのローカルサブセットを決定する。システムは、クライアントデバイスの属性およびプロアクティブアシスタントキャッシュエントリの属性に基づいて、ローカルサブセットを決定する。たとえば、プロアクティブアシスタントキャッシュエントリのうちのいくつかは、クライアントデバイスの属性に対応する属性を有することに基づいて選択され得る。そのような属性は、アプリケーション(たとえば、クライアントデバイス上にインストールされ、プロアクティブアシスタントキャッシュエントリのディープリンクに対応する)、地理的ロケーション(たとえば、クライアントデバイスの、かつプロアクティブアシスタントキャッシュエントリに対応する)、プロアクティブアシスタントキャッシュエントリに対応するエンティティと比較して(たとえば、コンテンツの検索または訪問を通して)クライアントデバイスを介して対話したエンティティ、および/または他の属性を含み得る。また、たとえば、プロアクティブキャッシュエントリのうちのいくつかは、追加または代替として、プロアクティブキャッシュエントリに対するランキング基準に基づいて選択され得る。ローカルサブセット内に含まれたプロアクティブアシスタントキャッシュエントリの量は、クライアントデバイスのプロアクティブアシスタントキャッシュのサイズの影響を受ける可能性がある。

20

【0084】

ブロック540において、システムは、任意選択で、選択されたクライアントデバイスに対するプロアクティブキャッシュエントリのリモートサブセットを決定する。リモートサブセットは、ブロック530のローカルサブセット内に含まれていないプロアクティブキャッシュエントリを含み得る(または、それに制限され得る)。いくつかの実装形態では、ブロック540は、ブロック540のリモートサブセットを決定する際、クライアントデバイスの属性およびプロアクティブアシスタントキャッシュエントリの属性を考慮し得る。リモートサブセット内に含まれたプロアクティブアシスタントキャッシュエントリの量も、リモートプロアクティブアシスタントキャッシュのサイズの影響を受ける可能性がある。リモートプロアクティブアシスタントキャッシュは、クライアントデバイスに固有であり得るか、またはそのクライアントデバイスおよび他の同様のクライアントデバイスを含むクライアントデバイスの制限付きグループに固有であり得る。リモートプロアクティブアシスタントキャッシュは、本明細書で説明するように、クライアントデバイスから生じる要求に対する応答をプロビジョニングする際のレイテンシを低減する際、リモート自動アシスタント構成要素によって利用され得る。

30

40

【0085】

システムは、次いで、ブロック520に戻り、別のクライアントデバイスを選択し、他のクライアントデバイスに対してブロック530および540を実施し得る。ブロック520、530、および540の複数の反復は、各々、異なるクライアントデバイスに対して、並行して実施され得ることを理解されたい。ブロック520、530、および540は、最新のローカルおよび/またはリモートプロアクティブアシスタントキャッシュを維持し、ブロック510の複数の反復を通して生成され得る、新しく生成されたプロアクティブアシスタントキャッシュエントリを明らかにするために、規則的または不規則的な間隔で様々なクライアント

50

デバイスに対して繰り返されてよいことをさらに理解されたい。

【0086】

ブロック550において、システムは、任意選択で、所与のクライアントデバイスからプリフェッチ要求を受信する。たとえば、所与のクライアントデバイスは、方法400に関して説明したように、プリフェッチ要求を送信し得る。

【0087】

ブロック560において、システムは、ブロック530の反復において所与のクライアントデバイスに対して決定された、ローカルサブセットの1つまたは複数のプロアクティブアシスタントキャッシュエントリを所与のクライアントデバイスに提供する。ローカルサブセットのどのエントリが提供されるかは、どれが(もしあれば)所与のクライアントデバイスのローカルプロアクティブキャッシュ内にすでに記憶されているかを決定することに基づき得る。たとえば、任意選択で、ローカルプロアクティブキャッシュ内にまだ存在しないエントリのみがブロック560において提供され得る。提供されたプロアクティブアシスタントキャッシュエントリは、自動アシスタントアプリケーションが所与のクライアントデバイスにおいて提供される将来のユーザインターフェース入力をローカルに履行する際に利用するために、所与のクライアントデバイスの自動アシスタントアプリケーションによって、ローカルプロアクティブキャッシュ内に記憶され得る。

【0088】

ブロック550が実施されるとき、ブロック560は、プリフェッチ要求に応じて実施され得る。ブロック550が実施されないとき、ブロック560は、所与のクライアントデバイスからの明示的な要求とは無関係に、プロアクティブアシスタントキャッシュエントリをプロアクティブにプッシュすることを含み得る。ブロック560(および任意選択で、ブロック550)は、大量のクライアントデバイスの各々に対して実施されることになり、クライアントデバイスの各々に対して複数の時点で実施されることになることを理解されたい。方法500を通して、異種のプロアクティブキャッシュエントリが異なるクライアントデバイスに(かつ/または対応するリモートプロアクティブキャッシュ内に記憶するために)提供されることになり、経時的に更新されることになることをさらに理解されたい。

【0089】

図5Bを参照すると、図5Aのブロック510のいくつかの実装形態を示すフローチャート510Bが提供される。ブロック511Bにおいて、システムは、エンティティに対するイベントが存在するかどうかを決定する。存在しない場合、システムは、ブロック511Bにおいてイベントを監視し続ける。存在する場合、システムは、任意選択で、ブロック512Bに進み、エンティティに対するアシスタント要求パラメータを生成し、ブロック514Bに進み、アシスタント要求パラメータに対するアクションコンテンツを生成し、ブロック516Bに進み、プロアクティブキャッシュエントリに関するメタデータを生成することによって、イベントに基づいてプロアクティブキャッシュエントリを生成する。プロアクティブキャッシュエントリは、アシスタント要求パラメータ、アクションコンテンツ、およびメタデータを含むことになる。ブロック517Bにおいて、システムは、エンティティに対するさらなるプロアクティブアシスタントキャッシュエントリを生成するかどうかを決定する。生成する場合、システムは、任意選択で、ブロック512Bの別の反復を実施し、ブロック514Bおよびブロック516Bの別の反復を実施することによって、エンティティに対する別のプロアクティブアシスタントキャッシュエントリを生成する。生成しない場合、システムは、ブロック511Bに戻り、そのエンティティおよび/または別のエンティティに対する別のイベントを監視する。

【0090】

一例として、ブロック511Bにおいて、イベントは、地理的エリアに対する天気予報に対する変更であってよい。この例を続けると、プロアクティブキャッシュエントリが生成され得、プロアクティブキャッシュエントリは、ブロック514Bにおいて、新しいアクションコンテンツ(新しい天気予報を記述する)を反映し、ブロック516Bにおいて新しいメタデータ(たとえば、タイムスタンプ)を反映するが、地理的エリアに対する天気予報に対す

10

20

30

40

50

る前のエントリと同じアシスタント要求パラメータを維持する。

【0091】

別の例として、ブロック511Bにおいて、イベントは特定のミュージカルアーティストに関する音楽ストリーミング要求の増大であり得る。この例を続けると、特定のミュージカルアーティストからの音楽をストリーミングするためのアシスタント要求に関する、512Bにおいて生成されるアシスタント要求パラメータ、第1のアプリケーションに関する特定のアーティストをストリーミングするためのディープリンクを含む、ブロック514Bにおいて生成されるアクションコンテンツ、およびブロック516Bにおいて生成されるメタデータ(たとえば、TTL値)を含むプロアクティブキャッシュエントリが生成され得る。さらに、ブロック517Bにおいて、同じアシスタント要求パラメータおよび/またはメタデータを10 含むが、異なるディープリンクをアクションコンテンツ内に含む、別のプロアクティブキャッシュエントリを生成することが決定され得る。異なるディープリンクは、ブロック514Bの別の反復において生成され、第2のアプリケーションに対して特定のミュージカルアーティストをストリーミングするためのディープリンクを含む。

【0092】

さらに別の例として、ブロック511Bにおいて、イベントは特定の都市に関する要求および/またはコンテンツの増大であり得る。この例を続けると、他の都市に対する頻繁なアシスタント要求(たとえば、特定の都市の人口を要求する)に基づく、512Bにおいて生成されるアシスタント要求パラメータ、視覚および/またはテキスト応答(たとえば、人口の視覚および/またはテキスト表現)を含む、ブロック514Bにおいて生成されるアクションコ20 ンテンツ、およびブロック516Bにおいて生成されるメタデータ(たとえば、TTL値)を含むプロアクティブキャッシュエントリが生成され得る。さらに、ブロック517Bにおいて、他の都市に対する他の頻繁なアシスタント要求に基づく追加のプロアクティブキャッシュエントリを生成することが決定され得る。たとえば、他の都市に対する他の頻繁なアシスタント要求(たとえば、特定の都市に関する気候情報を要求する)に基づく、512Bにおいて生成されるアシスタント要求パラメータ、視覚および/またはテキスト応答(たとえば、気候情報の視覚および/またはテキスト表現)を含む、ブロック514Bにおいて生成されるア30 クションコンテンツ、およびブロック516Bにおいて生成されるメタデータ(たとえば、TTL値)を含む追加のプロアクティブキャッシュエントリが生成され得る。

【0093】

図5Cを参照すると、図5Aのブロック510のいくつかの追加または代替の実装形態を示すフローチャート510Cが提供される。たとえば、ブロック510のいくつかの反復はフローチャート510Bに基づいて実施され得、他の反復はフローチャート510Cに基づいて実施され得る。ブロック512Cにおいて、システムは、アシスタント要求パラメータを生成する。実施例として、「あるBluegrassを再生する」、「あるBluegrassを演奏する」、および/または「{Intent=stream music; genre=bluegrass}」のパラメータなど、Bluegrass音楽をストリーミングするためのアシスタント要求パラメータが生成され得る。

【0094】

ブロック514Cにおいて、システムは、サブブロック514C1において、要求パラメータに対するN個のアプリケーションを決定することによって、アクションコンテンツを生成し、ここで、Nは、1よりも大きい整数である。たとえば、システムは、ブルーグラス音楽をストリーミングするための15個のアプリケーションを決定し得る。さらに、サブブロック514C2において、システムは、N個のアプリケーションの各々に対するアクションコンテンツを決定する。たとえば、システムは、N個のアプリケーションの各々に対するアクションコンテンツとして、実行されたとき、対応するアプリケーションにブルーグラス音楽をストリーミングさせる、対応するディープリンクを決定する。

【0095】

ブロック516Cにおいて、システムはメタデータを生成する。

【0096】

ブロック518Cにおいて、システムは、N個のプロアクティブアシスタントキャッシュエ50

ントリを生成する。N個の生成されたプロアクティブアシスタントキャッシュエントリの各々は、512Cの同じアシスタント要求パラメータ、および任意選択で、516Cの同じメタデータを有するが、異なるアクションコンテンツを含む(すなわち、各々は、N個のアプリケーションのうちの単一のアプリケーションに対する対応するディープリンクのみを有するアクションコンテンツを含み得る)。

【0097】

図6は、本明細書で説明する技法の1つまたは複数の態様を実施するために任意選択で利用され得る例示的なコンピューティングデバイス610のブロック図である。いくつかの実装形態では、クライアントデバイス、クラウドベースの自動アシスタント構成要素、および/または他の構成要素のうちの1つまたは複数は、例示的なコンピューティングデバイス610の1つまたは複数の構成要素を備え得る。

10

【0098】

コンピューティングデバイス610は、一般に、バスサブシステム612を介していくつかの周辺デバイスと通信する、少なくとも1つのプロセッサ614を含む。これらの周辺デバイスは、たとえば、メモリサブシステム625およびファイルストレージサブシステム626を含むストレージサブシステム624、ユーザインターフェース出力デバイス620、ユーザインターフェース入力デバイス622、およびネットワークインターフェースサブシステム616を含み得る。入力および出力デバイスは、コンピューティングデバイス610とのユーザ対話を可能にする。ネットワークインターフェースサブシステム616は、外部ネットワークにインターフェースを提供し、他のコンピューティングデバイス内の対応するインターフェースデバイスに結合される。

20

【0099】

ユーザインターフェース入力デバイス622は、キーボード、マウス、トラックボール、タッチパッド、またはグラフィックタブレットなどのポインティングデバイス、スキャナ、ディスプレイに組み込まれたタッチスクリーン、音声認識システム、マイクロフォンなどの音声入力デバイス、および/または他のタイプの入力デバイスを含み得る。概して、「入力デバイス」という用語の使用は、情報をコンピューティングデバイス610内にまたは通信ネットワーク上に入力するすべての考えられるタイプのデバイスおよび方法を含むことを意図する。

【0100】

ユーザインターフェース出力デバイス620は、ディスプレイサブシステム、プリンタ、ファックス機、またはオーディオ出力デバイスなどの非視覚的ディスプレイを含み得る。ディスプレイサブシステムは、陰極線管(CRT)、液晶ディスプレイ(LCD)などのフラットパネルデバイス、投影デバイス、または視覚画像を作成するための何らかの他の機構を含み得る。ディスプレイサブシステムはまた、オーディオ出力デバイスを介してなど、非視覚的ディスプレイを提供し得る。概して、「出力デバイス」という用語の使用は、情報をコンピューティングデバイス610からユーザもしくは別の機械またはコンピューティングデバイスに出力するすべての考えられるタイプのデバイスおよび方法を含むことを意図する。

30

【0101】

ストレージサブシステム624は、本明細書で説明するモジュールのうちのいくつかまたはすべての機能性を提供するプログラミングおよびデータ構成を記憶する。たとえば、ストレージサブシステム624は、本明細書で説明する方法の選択された態様を実施するための、ならびに本明細書で示す様々な構成要素を実装するための論理を含み得る。

40

【0102】

これらのソフトウェアモジュールは、概して、単独で、または他のプロセッサと組み合わせて、プロセッサ614によって実行される。ストレージサブシステム624内で使用されるメモリ625は、プログラム実行中に命令およびデータを記憶するためのメインランダムアクセスメモリ(RAM)630、および固定命令が記憶される読取り専用メモリ(ROM)632を含めて、いくつかのメモリを含み得る。ファイルストレージサブシステム626は、プログ

50



ラムおよびデータファイルに対する永続記憶装置を提供することができ、ハードディスクドライブ、関連するリムーバブルメディアとともにフロッピーディスクドライブ、CD-ROMドライブ、光ドライブ、またはリムーバブルメディアカートリッジを含み得る。いくつかの実装形態の機能性を実装するモジュールは、ファイルストレージサブシステム626によって、ストレージサブシステム624内に、またはプロセッサ614によってアクセス可能な他の機械の中に、記憶され得る。

#### 【0103】

バスサブシステム612は、意図されるように、コンピューティングデバイス610の様々な構成要素およびサブシステムに互いと通信させるための機構を提供する。バスサブシステム612は単一のバスとして概略的に示されているが、バスサブシステムの代替実装形態は、複数のバスを使用し得る。

10

#### 【0104】

コンピューティングデバイス610は、ワークステーション、サーバ、コンピューティングクラスタ、ブレードサーバ、サーバファーム、または任意の他のデータ処理システムまたはコンピューティングデバイスを含めて、様々なタイプのものであり得る。コンピュータおよびネットワークの絶えず変化する性質により、図6に示したコンピューティングデバイス610の記述は、いくつかの実装形態を示すための特定の例であることのみを意図する。図6に示したコンピューティングデバイスよりも多数または少数の構成要素を有するコンピューティングデバイス610の多くの他の構成が可能である。

#### 【0105】

本明細書で説明するシステムがユーザについての個人情報収集するもしくは場合によっては監視する、または個人および/または監視される情報を利用し得る状況において、ユーザは、プログラムまたは機能がユーザ情報(たとえば、ユーザのソーシャルネットワーク、社会的なアクションもしくはアクティビティ、職業、ユーザの選好、またはユーザの現在の地理的ロケーション)を収集するかどうかを制御するための、または、ユーザにより関連があり得るコンテンツサーバからコンテンツを受信するかどうか、および/もしくはどのように受信するかを制御するための、機会を与えられ得る。また、いくつかのデータは、個人を識別できる情報が削除されるように、記憶または使用される前に1つまたは複数の方法で扱われ得る。たとえば、ユーザのアイデンティティは、ユーザについて個人を識別できる情報が決定され得ないように、またはユーザの特定の地理的ロケーションが決定され得ないように地理的ロケーション情報が取得される場合にユーザの地理的ロケーションが(都市レベル、ZIPコードレベル、または州のレベルなどに)一般化され得るように扱われ得る。したがって、ユーザは、情報がどのようにユーザについて収集されかつ/または使用されるかを制御することができる。

20

30

#### 【0106】

いくつかの実装形態では、1つまたは複数のプロセッサによって実施される方法が提供され、この方法は、所与のアクションを実施するための1つまたは複数のアシスタント要求を表すアシスタント要求パラメータを決定するステップを含む。アシスタント要求パラメータは、アシスタント要求の1つまたは複数のテキスト表現および/またはアシスタント要求の1つまたは複数の意味表現を定義する。この方法は、所与のアクションが第1のアプリケーションを使用して実施され得、かつ第2のアプリケーションを使用しても実施され得ると決定するステップをさらに含む。この方法は、第1のアプリケーションに対する第1のアクションコンテンツを生成するステップと、第2のアプリケーションに対する第2のアクションコンテンツを生成するステップとをさらに含む。第1のアクションコンテンツは、第1のアプリケーションに対する第1のディープリンクを含む。第1のディープリンクは、第1のアプリケーションをインストールしたクライアントデバイスのアシスタントクライアントアプリケーションによってローカルに実行可能であり、第1のディープリンクのローカルな実行は、所与のアクションを実施するための第1の状態第1のアプリケーションを開かせる。第2のアクションコンテンツは、第2のアプリケーションに対する第2のディープリンクを含む。第1のディープリンクとは異なる第2のディープリンクは、第2のア

40

50

アプリケーションをインストールしたクライアントデバイスのアシスタントクライアントアプリケーションによってローカルに実行可能であり、第2のディープリンクのローカルな実行は、所与のアクションを実施するための第2の状態第2のアプリケーションを開かせる。この方法は、アシスタント要求パラメータおよび第1のアクションコンテンツを含む第1のプロアクティブアシスタントキャッシュエントリを生成するステップと、アシスタント要求パラメータおよび第2のアクションコンテンツを含む第2のプロアクティブアシスタントキャッシュエントリを生成するステップとをさらに含む。この方法は、所与のクライアントデバイスに対するプロアクティブキャッシュエントリを生成するステップをさらに含む。プロアクティブキャッシュエントリを生成するステップは、所与のクライアントデバイスが、第1のアプリケーションをインストールしているが、第2のアプリケーションをインストールしていないことに基づいて、第1のプロアクティブキャッシュエントリを含めるが、第2のプロアクティブキャッシュエントリは含めないステップを含む。この方法は、任意選択で、所与のクライアントデバイスによって送信されたプロアクティブキャッシュ要求の受信に応じて、プロアクティブキャッシュエントリを所与のクライアントデバイスに送信するステップをさらに含む。所与のクライアントデバイスの自動アシスタントアプリケーションは、所与のクライアントデバイスにおいて提供される将来のユーザーインターフェース入力をローカルに履行する際に自動アシスタントアプリケーションによる利用のために、プロアクティブキャッシュエントリをローカルプロアクティブキャッシュ内に記憶する。

10

【0107】

20

本明細書で開示する技術のこれらの実装形態および他の実装形態は、以下の特徴のうちの1つまたは複数を含むことができる。

【0108】

いくつかの実装形態では、この方法は、所与のアクションに対する要求の増大を決定するステップと、所与のアクションに対する要求の増大の決定に応じて、第1のプロアクティブキャッシュエントリを生成し、第2のプロアクティブキャッシュエントリを生成するステップをさらに含む。

【0109】

いくつかの実装形態では、この方法は、所与のアクションに対する要求の増大を決定するステップをさらに含み、第1のプロアクティブキャッシュエントリをプロアクティブキャッシュエントリ内に含めるステップは、所与のアクションに対する要求の増大にさらに基づく。それらの実装形態のうちのいくつかでは、所与のアクションに対する要求の増大にさらに基づいて、第1のプロアクティブキャッシュエントリをプロアクティブキャッシュエントリ内に含めるステップは、所与のアクションが対象とするエンティティが、それに対してプロアクティブキャッシュエントリが生成される所与のクライアントデバイスに対する1つまたは複数の属性に対応すると決定するステップを含む。それらの実装形態のいくつかのバージョンでは、第1のアプリケーションは、音楽ストリーミングアプリケーションであり、所与のアクションは、エンティティの音楽をストリーミングすることであり、所与のクライアントデバイスに対する属性は、そのエンティティに対する所与のクライアントデバイスの関連付けを含む。これらの要求は、たとえば、自動アシスタント要求および/または追加要求を含み得る。自動アシスタント要求は、対応する自動アシスタントアプリケーションに対し、各々、対応するユーザーインターフェース入力に応じて生成される。追加要求は、自動アシスタント要求に対する追加であり、自動アシスタントアプリケーションに対する追加である1つまたは複数の追加のアプリケーションから生じる。

30

40

【0110】

いくつかの実装形態では、この方法は、所与のアクションに対する要求の増大を予測するステップと、所与のアクションに対する要求の増大の決定に応じて、第1のプロアクティブキャッシュエントリを生成し、第2のプロアクティブキャッシュエントリを生成するステップをさらに含む。いくつかの追加または代替実装形態では、この方法は、所与のアクションに対する要求の増大を予測するステップをさらに含み、第1のプロアクティブキ

50

キャッシュエントリをプロアクティブキャッシュエントリ内に含めるステップは、所与のアクションに対する要求の予測された増大にさらに基づく。所与のアクションに対する要求の増大を予測するステップは、所与のアクションが対象とするエンティティに関するインターネットコンテンツの増大を決定するステップおよび/または所与のアクションが対象とするエンティティに関連する将来のイベントを決定するステップを含み得る。

**【0111】**

いくつかの実装形態では、この方法は、第1のプロアクティブアシスタントキャッシュエントリに対する有効期間値を生成するステップと、有効期間値を第1のプロアクティブアシスタントキャッシュエントリ内に含めるステップとをさらに含む。有効期間値は、有効期間値によって定義される持続時間の満了に応じて、所与のクライアントデバイスにローカルプロアクティブキャッシュから第1のプロアクティブアシスタントキャッシュエントリを除去させる。それらの実装形態のうちいくつかでは、この方法は、所与のクライアントデバイスのアシスタントクライアントアプリケーションが、第1のプロアクティブアシスタントキャッシュエントリに対するタイムスタンプに対して有効期間値を比較することに基づいて、プロアクティブキャッシュエントリから第1のプロアクティブアシスタントキャッシュエントリを除去するステップをさらに含む。

10

**【0112】**

いくつかの実装形態では、この方法は、所与のクライアントデバイスのアシスタントクライアントアプリケーションが、プロアクティブキャッシュ要求の送信に応じて、プロアクティブキャッシュエントリを受信するステップと、プロアクティブキャッシュエントリを所与のクライアントデバイスのローカルプロアクティブキャッシュ内に記憶するステップとをさらに含む。それらの実装形態のうちいくつかでは、この方法は、所与のクライアントデバイスのアシスタントクライアントアプリケーションが、プロアクティブキャッシュエントリをローカルプロアクティブキャッシュ内に記憶するステップの後、オンデバイス音声認識を使用して、クライアントデバイスの1つまたは複数のマイクロフォンによってユーザインターフェース入力として検出されたオーディオデータ内でキャプチャされた口頭の発話に基づいて、認識されたテキストを生成するステップと、プロアクティブキャッシュにアクセスすることに基づいて、第1のプロアクティブアシスタントキャッシュエントリのアシスタント要求パラメータが、認識されたテキストおよび/または認識されたテキストに基づいて生成された自然言語理解データにマッチすると決定するステップと、マッチの決定に応じて、第1のディープリンクをローカルに実行して、所与のアクションを実施するための第1の状態第1のアプリケーションを開かせるステップとをさらに含む。

20

30

**【0113】**

いくつかの実装形態では、この方法は、所与のクライアントデバイスのアシスタントクライアントアプリケーションが、所与のクライアントデバイスのネットワークステータスおよび/または所与のクライアントデバイスの計算負荷ステータスが1つまたは複数の条件を満たすと決定するステップと、ネットワークステータスおよび/または計算負荷ステータスが1つまたは複数の条件を満たすとの決定に応じて、プロアクティブキャッシュ要求を送信するステップと、プロアクティブキャッシュ要求の送信に応じて、プロアクティブキャッシュエントリを受信するステップと、プロアクティブキャッシュエントリを所与のクライアントデバイスのローカルプロアクティブキャッシュ内に記憶するステップとをさらに含む。

40

**【0114】**

いくつかの実装形態では、1つまたは複数のプロセッサによって実施される方法が提供され、この方法は、特定のエンティティに関連するイベントの発生を決定するステップを含む。この方法は、特定のエンティティに関連するイベントの発生の決定に応じて、特定のエンティティに対する1つまたは複数のプロアクティブアシスタントキャッシュエントリを生成するステップをさらに含む。プロアクティブアシスタントキャッシュエントリの各々は、それぞれのアシスタント要求パラメータおよびそれぞれのアシスタントアクショ

50

ンコンテンツを定義する。それぞれのアシスタント要求パラメータは、各々、特定のエンティティに関する1つまたは複数のそれぞれのアシスタント要求を表し、アシスタント要求の1つまたは複数のテキスト表現および/またはアシスタント要求の1つまたは複数の意味表現を定義する。それぞれのアシスタントアクションコンテンツは、各々、アシスタントクライアントアプリケーションによって、特定のエンティティに関して、かつ1つまたは複数のそれぞれのアシスタント要求に応じて、アシスタントアクションをローカルに実施させるために、クライアントデバイスのアシスタントクライアントアプリケーションによってローカルに解釈可能である。この方法は、サブセットのクライアントデバイスが各々、特定のエンティティに対応する1つまたは複数の対応する属性を有すると決定することに基づいて、クライアントデバイスのサブセットを選択するステップをさらに含む。この方法は、プロアクティブアシスタントキャッシュエントリを選択されたサブセット内にはない他のクライアントデバイスに送信せずに、特定のエンティティに対するプロアクティブアシスタントキャッシュエントリをサブセットの複数のクライアントデバイスに送信するステップをさらに含む。プロアクティブアシスタントキャッシュエントリを送信するステップは、クライアントデバイスの各々の対応する自動アシスタントアプリケーションに、所与のクライアントデバイスにおいて提供される将来の口頭の発話をローカルに履行する際に自動アシスタントアプリケーションによる利用のために、プロアクティブアシスタントキャッシュエントリをローカルプロアクティブキャッシュ内にローカルにキャッシュさせる。

10

**【0115】**

20

本明細書で開示する技術のこれらの実装形態および他の実装形態は、以下の特徴のうちの1つまたは複数を含むことができる。

**【0116】**

いくつかの実装形態では、特定のエンティティに関連するイベントの発生を決定するステップは、特定のエンティティに関する要求の増大および/または特定のエンティティに関するインターネットコンテンツの増大を決定するステップを含む。

**【0117】**

いくつかの実装形態では、特定のエンティティに対する1つまたは複数のプロアクティブアシスタントキャッシュエントリのうちの所与の1つを生成するステップは、特定のエンティティの1つまたは複数の属性に基づいて、特定のエンティティに関する1つまたは複数のそれぞれのアシスタント要求を生成するステップと、1つまたは複数のそれぞれのアシスタント要求に基づいて、それぞれのアシスタント要求パラメータを生成するステップとを含む。

30

**【0118】**

いくつかの実装形態では、特定のエンティティは、特定の人物または特定の組織である。

**【0119】**

いくつかの実装形態では、特定のエンティティの1つまたは複数の属性に基づいて、1つまたは複数のそれぞれのアシスタント要求を生成するステップは、エンティティのクラスを決定するステップと、クラスのエンティティに対する最も頻繁なクエリに対するテンプレートを決定するステップと、テンプレートおよびエンティティのエイリアスを使用して、それぞれのアシスタント要求のうちの少なくとも1つを生成するステップとを含む。

40

**【0120】**

いくつかの実装形態では、特定のエンティティに関連するイベントは、特定のエンティティに関する変更された属性、または特定のエンティティに関する新しい属性である。それらの実装形態のうちのいくつかでは、特定のエンティティに対する1つまたは複数のプロアクティブアシスタントキャッシュエントリを生成するステップは、プロアクティブアシスタントキャッシュエントリのうちの所与の1つに対するアクションのローカル実施中に属性をレンダリングさせるために、前に生成されたプロアクティブアシスタントキャッシュエントリに対して、それぞれのアクションコンテンツを改変することによって、その所与の1つを生成するステップを含む。それらの実装形態のいくつかのバージョンでは、

50

特定のエンティティは、地理的エリアにおける天気であり、属性は、その地理的エリアにおける高温度、その地理的エリアにおける低温度、および/またはその地理的エリアにおける降水確率である。それらの実装形態のいくつかの他のバージョンでは、特定のエンティティはイベントであり、属性は、イベントの開始時間、イベントの終了時間、および/またはイベントのロケーションである。

**【 0 1 2 1 】**

いくつかの実装形態では、この方法は、サブセットのクライアントデバイスと重複していない、クライアントデバイスの第2のサブセットを選択するステップをさらに含む。この方法は、プロアクティブアシスタントキャッシュエントリを第2のサブセットのクライアントデバイスからのアシスタント要求に応じて利用される1つまたは複数のリモートプロアクティブキャッシュ内に記憶するステップをさらに含む。それらの実装形態のうちのいくつかでは、1つまたは複数のリモートプロアクティブキャッシュは、第2のサブセットのクライアントデバイスの各々に対するリモートプロアクティブキャッシュのうちの対応する1つを含む。

10

**【 0 1 2 2 】**

いくつかの実装形態では、この方法は、サブセットのクライアントデバイスの所与のクライアントデバイスのアシスタントクライアントアプリケーションがプロアクティブキャッシュ要求を送信するステップと、プロアクティブキャッシュ要求の受信に応じて、プロアクティブアシスタントキャッシュエントリを所与のクライアントデバイスに送信するステップとをさらに含む。

20

**【 0 1 2 3 】**

いくつかの実装形態では、この方法は、サブセットのクライアントデバイスの所与のクライアントデバイスのアシスタントクライアントアプリケーションが、プロアクティブキャッシュエントリを受信するステップと、プロアクティブキャッシュエントリを所与のクライアントデバイスの所与のローカルプロアクティブキャッシュ内に記憶するステップとをさらに含む。それらの実装形態のうちのいくつかでは、この方法は、プロアクティブキャッシュエントリをローカルプロアクティブキャッシュ内に記憶するステップの後に、所与のクライアントデバイスのアシスタントクライアントアプリケーションが、オンデバイス音声認識を使用して、クライアントデバイスの1つまたは複数のマイクロフォンによって検出されたオーディオデータ内でキャプチャされた口頭の発話に基づいて、認識されたテキストを生成するステップと、プロアクティブキャッシュにアクセスすることに基づいて、プロアクティブアシスタントキャッシュエントリのうちの所与の1つのそれぞれのアシスタント要求パラメータが、認識されたテキストおよび/または認識されたテキストに基づいて生成された自然言語理解データにマッチすると決定するステップと、マッチの決定に応じて、プロアクティブアシスタントキャッシュエントリのうちの所与の1つのそれぞれのアシスタントアクションコンテンツをローカルに解釈するステップとをさらに含む。それらの実装形態のいくつかのバージョンでは、プロアクティブアシスタントキャッシュエントリのうちの所与の1つのそれぞれのアシスタントアクションコンテンツをローカルに解釈するステップは、アシスタントアクションコンテンツのテキストおよび/またはグラフィカルコンテンツをクライアントデバイスにおいてレンダリングさせるステップを含む。

30

40

**【 0 1 2 4 】**

いくつかの実装形態では、1つまたは複数のプロセッサによって実装される方法が提供され、この方法は、特定のエンティティに関連するイベントの発生を決定するステップを含む。この方法は、特定のエンティティに関連するイベントの発生を決定することに基づいて、サブセットのクライアントデバイスが各々特定のエンティティに対応する1つまたは複数の対応する属性を有すると決定することに基づいて、クライアントデバイスのサブセットを選択するステップをさらに含む。この方法は、プロアクティブアシスタントキャッシュエントリを選択されたサブセット内にはない他のクライアントデバイスに送信せずに、特定のエンティティに対する1つまたは複数のプロアクティブアシスタントキャッシュエントリをサブセットの複数のクライアントデバイスに送信するステップをさらに含む。

50

特定のエンティティに対するプロアクティブアシスタントキャッシュエントリは各々、特定のエンティティに関する1つまたは複数のそれぞれのアシスタント要求を表すそれぞれのアシスタント要求パラメータ、およびアシスタントクライアントアプリケーションによって、特定のエンティティに関して、かつ1つまたは複数のそれぞれのアシスタント要求に応じて、アシスタントアクションをローカルに実施させるために、クライアントデバイスのアシスタントクライアントアプリケーションによってローカルに解釈可能であるそれぞれのアシスタントアクションコンテンツを定義する。プロアクティブアシスタントキャッシュエントリを送信するステップは、クライアントデバイスの各々の対応する自動アシスタントアプリケーションに、所与のクライアントデバイスにおいて提供される将来の口頭の発話をローカルに履行する際に自動アシスタントアプリケーションによる利用のために、プロアクティブアシスタントキャッシュエントリをローカルプロアクティブキャッシュ内にローカルにキャッシュさせる。

10

## 【0125】

本明細書で開示する技術のこれらの実装形態および他の実装形態は、以下の特徴のうちの1つまたは複数を含むことができる。

## 【0126】

いくつかの実装形態では、特定のエンティティに関連するイベントの発生を決定するステップは、特定のエンティティに関するインターネットコンテンツの増大を決定するステップ、特定のエンティティに対する要求の増大を決定するステップ、1つまたは複数の非自動アシスタントアプリケーションからのものである要求を含めるステップ、および/または1つまたは複数のリモートサーバとの対話を通して、特定のエンティティの属性の変更を決定するステップを含む。

20

## 【符号の説明】

## 【0127】

- 120 プロアクティブキャッシュシステム
- 130 キャッシュエントリ生成エンジン
- 131 要求パラメータモジュール
- 132 アクションコンテンツモジュール
- 133 メタデータモジュール
- 134 キャッシュ候補
- 140 キャッシュアセンブリエンジン、イベントキャッシュアセンブリエンジン
- 141 ローカルモジュール
- 142 リモートモジュール
- 145 オンデバイス履行エンジン
- 150 エンティティイベントエンジン
- 151 リモートシステム
- 160 クライアントデバイス
- 163 視覚構成要素
- 165 マイクロフォン
- 167 存在センサ
- 169 ディスプレイ
- 170 自動アシスタントクライアント
- 171 音声キャプチャエンジン
- 171A 認識されたテキスト
- 172 視覚キャプチャエンジン
- 173 オンデバイス音声認識エンジン
- 174 オンデバイスNLUエンジン
- 174A NLUデータ
- 175 オンデバイス履行エンジン
- 175A 履行データ

30

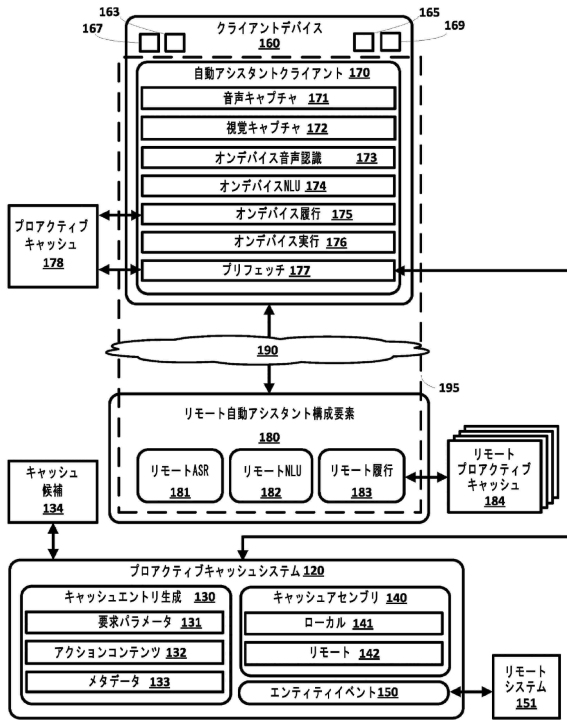
40

50

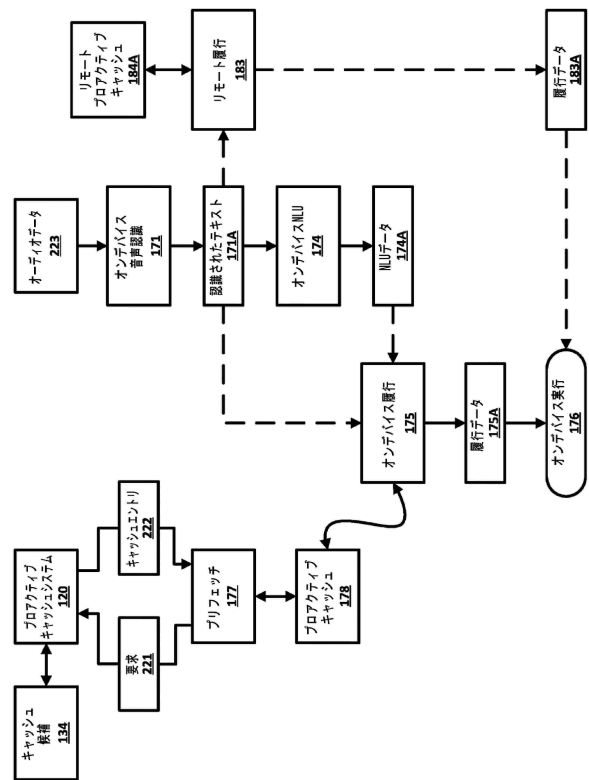
176	オンデバイス実行エンジン	
177	プリフェッチエンジン	
178	プロアクティブキャッシュ	
180	リモート自動アシスタント構成要素	
181	リモートARSエンジン	
182	リモートNLUエンジン	
183	リモート履行エンジン	
183A	リモート履行データ、履行データ	
184	リモートプロアクティブキャッシュ	
184A	リモートプロアクティブキャッシュ	10
195	自動アシスタント	
221	要求	
222	プロアクティブキャッシュエントリ、キャッシュエントリ	
223	オーディオデータ	
310	プロアクティブキャッシュエントリ	
310A	要求パラメータ	
310B	アクションコンテンツ	
310C	メタデータ	
320	プロアクティブキャッシュエントリ	
320A	要求パラメータ	20
320B	アクションコンテンツ	
320C	メタデータ	
330	プロアクティブキャッシュエントリ	
330A	要求パラメータ	
330B	アクションコンテンツ	
330C	メタデータ	
400	方法	
500	方法	
610	コンピューティングデバイス	
612	バスサブシステム	30
614	プロセッサ	
616	ネットワークインターフェースサブシステム	
620	ユーザインターフェース出力デバイス	
622	ユーザインターフェース入力デバイス	
624	ストレージサブシステム	
625	メモリサブシステム、メモリ	
626	ファイルストレージサブシステム	

【図面】

【図1】



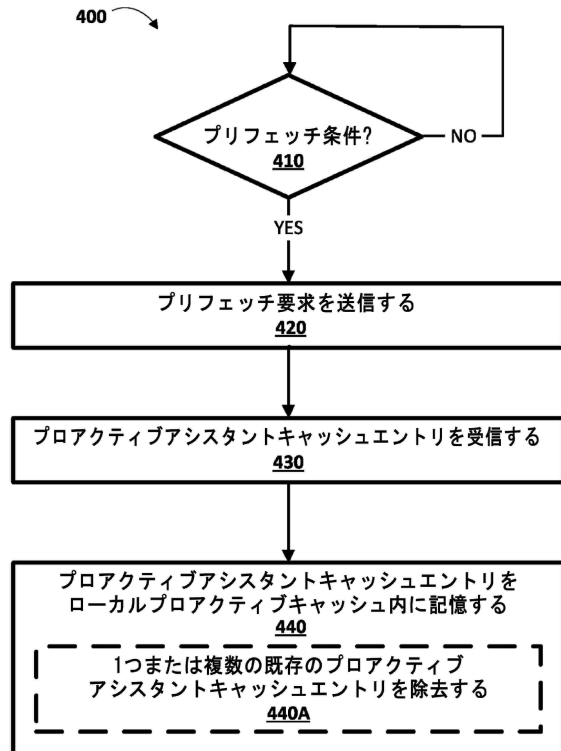
【図2】



【図3】

キャッシュエントリ	要求パラメータ	アクションコンテンツ	メタデータ
1 310	<ul style="list-style-type: none"> <li>明日の天気</li> <li>天気明日</li> <li>Intent=weather;</li> <li>Day=tomorrow;</li> <li>Location=local</li> </ul>	<ul style="list-style-type: none"> <li>最高気温68、最低気温50の晴天</li> </ul>	<ul style="list-style-type: none"> <li>TTL=12時間</li> <li>タイムスタンプ=5/4/19、18:50</li> </ul>
2 320	<ul style="list-style-type: none"> <li>サーモスタットスケジュールを調整する</li> <li>サーモスタットに対するスケジュールを変更する</li> <li>Intent=thermostat;</li> <li>Setting=change/adjust schedule</li> </ul>	<ul style="list-style-type: none"> <li>アプリケーション#へのディープリンク</li> </ul>	<ul style="list-style-type: none"> <li>TTL=30日</li> <li>タイムスタンプ=5/3/19、18:12</li> </ul>
3 330	<ul style="list-style-type: none"> <li>John Doeの総資産</li> <li>John Doeの財産はどの程度</li> </ul>	<ul style="list-style-type: none"> <li>-\$50,000,000</li> </ul>	<ul style="list-style-type: none"> <li>TTL=7日</li> <li>タイムスタンプ=5/2/19、13:45</li> </ul>

【図4】



10

20

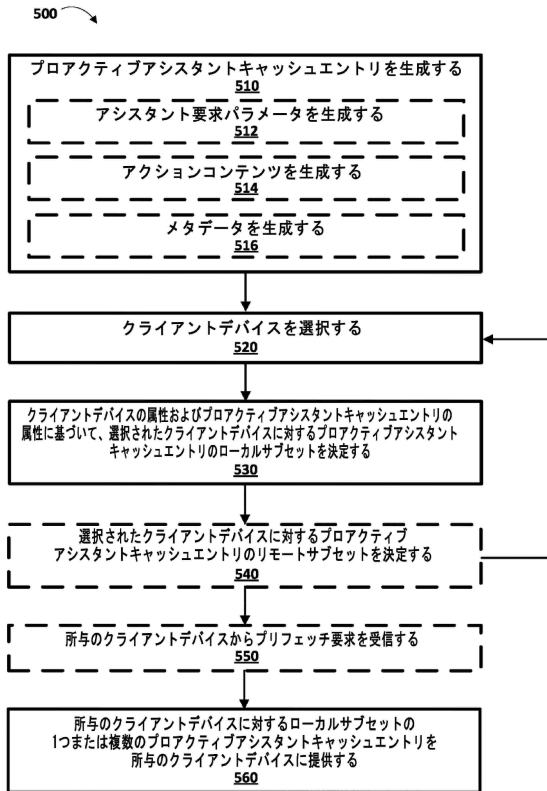
30

40

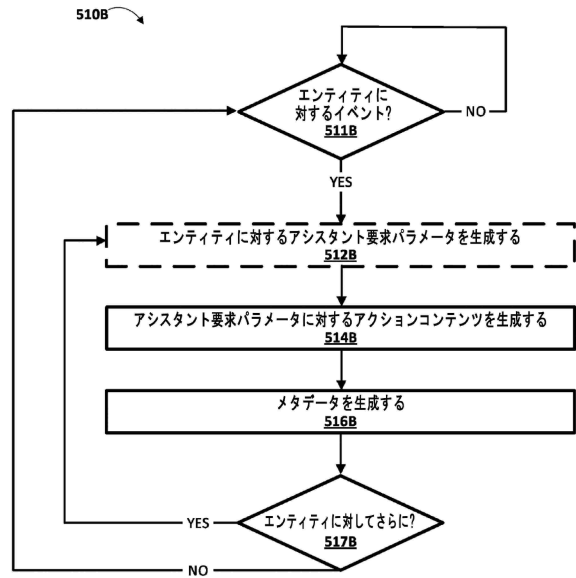
50



【図 5 A】



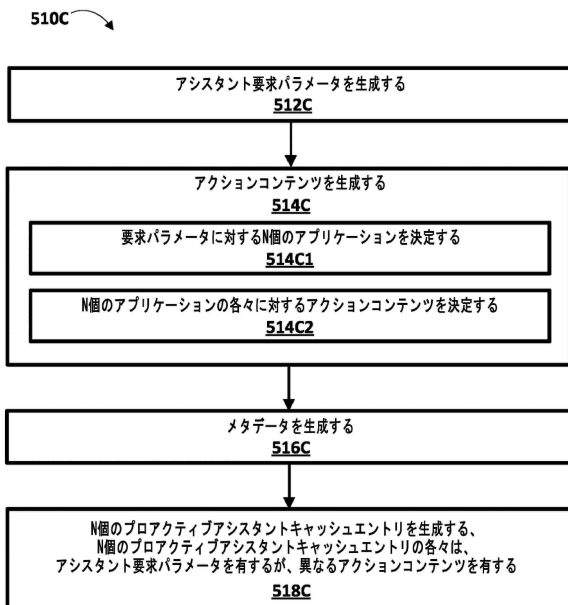
【図 5 B】



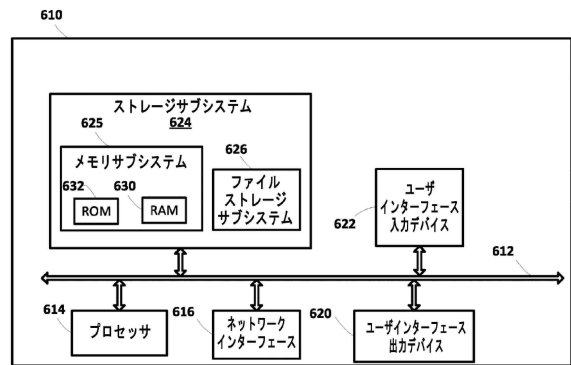
10

20

【図 5 C】



【図 6】



30

40

50

## フロントページの続き

- (72)発明者 ダニエル・コッティング  
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク  
ウェイ・1600
- (72)発明者 ザヒド・サバー  
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク  
ウェイ・1600
- (72)発明者 ラン・フオ  
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク  
ウェイ・1600
- (72)発明者 ブライアン・クリストファー・ホーリング  
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク  
ウェイ・1600
- (72)発明者 ベシャド・ベザディ  
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク  
ウェイ・1600
- (72)発明者 ルーカス・ミレルマン  
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク  
ウェイ・1600
- (72)発明者 マイケル・ゴリコフ  
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク  
ウェイ・1600
- (72)発明者 デニス・ブラコフ  
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク  
ウェイ・1600
- (72)発明者 スティーヴ・チェン  
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク  
ウェイ・1600
- (72)発明者 ボーダン・ヴラシュク  
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク  
ウェイ・1600
- (72)発明者 セルゲイ・ナザロフ  
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク  
ウェイ・1600
- (72)発明者 マリオ・ベツラー  
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク  
ウェイ・1600
- (72)発明者 ルヴ・コタリ  
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク  
ウェイ・1600
- 審査官 木村 雅也
- (56)参考文献 特表2004-531829(JP,A)  
特表2017-527844(JP,A)  
米国特許出願公開第2002/0198991(US,A1)  
韓国登録特許第10-0545442(KR,B1)
- (58)調査した分野 (Int.Cl., DB名)  
H04L 67/568  
G06F 16/90  
H04L 67/52  
H04L 67/561

G 1 0 L 1 5 / 1 0