

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2017-526041

(P2017-526041A)

(43) 公表日 平成29年9月7日(2017.9.7)

(51) Int.Cl.
G06F 13/00 (2006.01)F I
G O 6 F 13/00 5 4 0 Bテーマコード (参考)
5 B 0 8 4

審査請求 有 予備審査請求 未請求 (全 33 頁)

(21) 出願番号 特願2016-572796 (P2016-572796)
 (86) (22) 出願日 平成26年6月26日 (2014. 6. 26)
 (85) 翻訳文提出日 平成29年2月13日 (2017. 2. 13)
 (86) 国際出願番号 PCT/CN2014/080832
 (87) 国際公開番号 W02015/196414
 (87) 国際公開日 平成27年12月30日 (2015. 12. 30)

(71) 出願人 502208397
 グーグル インコーポレイテッド
 アメリカ合衆国 カリフォルニア州 94
 043 マウンテン ビュー アンフィシ
 アター パークウェイ 1600
 (74) 代理人 100108453
 弁理士 村山 靖彦
 (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】 バッチ最適化レンダリング及びフェッチアーキテクチャ

(57) 【要約】

実施形態は、バッチ最適化されたレンダリングおよびフェッチアーキテクチャを含む。アーキテクチャにより実行される方法例は、ウェブページをレンダリングするバッチプロセスから要求を受信し、ウェブページをレンダリングするための仮想時計とタスクリストを初期化することを含む。仮想時計は、埋め込みアイテムの要求が未処理でありタスクが実行準備ができているとき、停止する。方法は、仮想時計がタスクリスト内の停止タスクのランタイムと一致するとき、ウェブページのレンダリング結果を生成しバッチプロセスにレンダリング結果を提供することを含んでもよい。別の方法例は、ウェブページをレンダリングするためのバッチプロセスから要求を受信し、ウェブページの埋め込みアイテムを識別し、書き換え規則に基づいて以前にフェッチされた埋め込みアイテムのためのコンテンツと重複するコンテンツを有する埋め込みアイテムを識別することを含んでもよい。

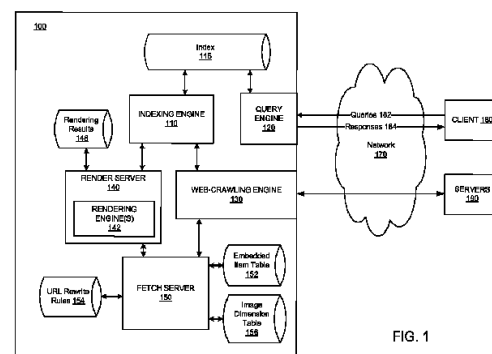


FIG. 1

【特許請求の範囲】**【請求項 1】**

少なくとも 1 つのプロセッサと、
メモリと、を含むコンピュータシステムであって、前記メモリが、
埋め込みアイテムのためのコンテンツのデータストアと、
前記少なくとも 1 つのプロセッサによって実行されたときに、前記システムに、
バッチプロセスから、ウェブページをレンダリングするための要求を受信すること
と、
前記ウェブページの埋め込みアイテムを識別することと、
書き換え規則に基づいて、前記埋め込みアイテムが以前にフェッチされた埋め込み
アイテムのためのコンテンツの重複であるコンテンツを有すると判定することと、
前記判定に応答して、前記データストアから前記以前にフェッチされた埋め込みア
アイテムのための前記コンテンツを提供することと、
前記以前にフェッチされた埋め込みアイテムのための前記コンテンツを使用して、
前記ウェブページについてのレンダリング結果を生成することと、
前記バッチプロセスに前記レンダリング結果を提供することと、
を行わせる命令と、を格納する、
コンピュータシステム。

10

【請求項 2】

前記埋め込みアイテムが以前にフェッチされた埋め込みアイテムのためのコンテンツの
重複であるコンテンツを有すると判断することの一部として、前記命令が、前記少なく
も 1 つのプロセッサによって実行されるときに、前記システムに、
前記書き換え規則のテンプレートに前記埋め込みアイテムを照合させることをもたら
す命令をさらに含み、前記書き換え規則がリダイレクト識別子も含み、
ここで、前記以前にフェッチされた埋め込みアイテムのための前記コンテンツを提供
することが、前記以前にフェッチされた埋め込みアイテムのための前記コンテンツを示す
ための前記リダイレクト識別子を使用することを含む、
請求項 1 に記載のシステム。

20

【請求項 3】

前記テンプレートが、クエリ文字列なしで URL を含む、請求項 2 に記載のシステム。

30

【請求項 4】

前記埋め込みアイテムが、第 1 の埋め込みアイテムであり、かつ前記命令が、前記少な
くとも 1 つのプロセッサによって実行されるとき、前記システムに、
前記ウェブページの第 2 の埋め込みアイテムを識別し、
前記第 2 の埋め込みアイテムがブラックリストに載せられているか否かを判断し、
前記第 2 の埋め込みアイテムが前記第 2 の埋め込みアイテムのためのフェッチコンテ
ンツなしでブラックリストに載せられているとき、エラーを返し、
前記第 2 の埋め込みアイテムのための前記コンテンツなしで前記レンダリング結果を
生成することをもたらす命令をさらに含み、
請求項 1 に記載のシステム。

40

【請求項 5】

前記命令が、前記少なくとも 1 つのプロセッサによって実行されるとき、前記システム
に、
前記レンダリング結果を生成するとき、仮想時計を使用することをもたらす命令をさ
らに含み、前記仮想時計が、リアルタイムとは無関係に進む、
請求項 1 に記載のシステム。

【請求項 6】

前記命令が、前記少なくとも 1 つのプロセッサによって実行されるとき、前記システム
に、
前記レンダリング結果を生成するとき、仮想時計を使用することをもたらす命令をさ

50

らに含み、

前記仮想時計が、前記以前にフェッチされた埋め込みアイテムの前記提供されたコンテンツを待つ間に進行しない、

請求項 1 に記載のシステム。

【請求項 7】

前記埋め込みアイテムが、第 1 の埋め込みアイテムであり、かつ前記命令が、前記少なくとも 1 つのプロセッサによって実行されるとき、前記システムに、

前記ウェブページの第 2 の埋め込みアイテムを識別し、

前記第 2 の埋め込みアイテムが画像を含むことを判断し、

サイズテーブルを使用して、前記第 2 の埋め込みアイテムのためのサイズを指定するモック画像を生成し、

前記レンダリング結果を生成するときに前記モック画像を使用することをもたらす命令をさらに含む、

請求項 1 に記載のシステム。

【請求項 8】

ウェブページをレンダリングするための要求をバッチプロセスから受信するステップと

、
少なくとも 1 つのプロセッサを使用して、仮想時計と前記ウェブページをレンダリングするためのタスクリストを初期化するステップであって、前記仮想時計が、埋め込みアイテムのための要求が未処理であるときかつタスクが準備完了であるとき、依然として静止している、初期化するステップと、

前記仮想時計が前記タスクリストの停止タスクのためのランタイムに一致するとき、前記少なくとも 1 つのプロセッサを使用して、前記ウェブページのためのレンダリング結果を生成するステップと、

前記バッチプロセスのための前記レンダリング結果を提供するステップと、を含む、方法。

【請求項 9】

前記タスクリストを初期化するステップが、ランタイムセットを備えた前記停止タスクを前記仮想時計に追加された所定の時間に追加するステップを含む、請求項 8 に記載の方法。

【請求項 10】

前記バッチプロセスが、インデックスエンジンを含み、かつ前記方法が、インデックスのトークンをランク付けするために前記レンダリング結果を使用するステップをさらに含む、請求項 8 に記載の方法。

【請求項 11】

埋め込みアイテムに対する要求が未解決でない、かつ前記仮想時計よりも大きいランタイムを有するただ一つのタスクが前記タスクリストにあるとき、前記仮想時計を前記タスクリスト内のタスクのランタイムに進めるステップをさらに含む、請求項 8 に記載の方法。

【請求項 12】

前記ウェブページの埋め込み画像を識別するステップと、

前記埋め込み画像のためのコンテンツを要求するステップと、

前記要求に回答して、前記埋め込み画像のためにサイズを指定するが、空のコンテンツを有するモック画像を受信するステップと、

前記レンダリング結果を生成するときに前記モック画像を使用するステップと、をさらに含む、

請求項 8 に記載の方法。

【請求項 13】

前記バッチプロセスが、インデックスエンジンであり、かつ前記方法が、前記レンダリング結果の情報に基づいて、前記ウェブページのためのランクを降格させるステップをさ

10

20

30

40

50

らに含む、請求項 8 に記載の方法。

【請求項 14】

前記バッチプロセスが、インデックスエンジンであり、かつ前記方法が、動的に生成されたコンテンツにインデックスを付けるために前記レンダリング結果を使用するステップをさらに含む、請求項 8 に記載の方法。

【請求項 15】

ウェブページ内の埋め込みアイテムのユニフォームリソースロケータ (URL) のためのバッチレンダリングプロセスから要求を受信するステップと、

少なくとも 1 つのプロセッサを使用して、書き換えられた URL を判断する書き換え規則を適用するステップと、

前記少なくとも 1 つのプロセッサを使用して、前記書き換えられた URL のためのコンテンツがデータストアに存在するか否かを判断するステップと、

前記コンテンツが存在するとき、前記コンテンツを前記バッチレンダリングプロセスに提供するステップと、

前記コンテンツが存在しないとき、

前記コンテンツのフェッチを初期化し、ここで前記バッチレンダリングプロセスが前記フェッチの間にタイムアウトなしで待機するように構成され、

ウェブ巡回エンジンから前記コンテンツを受信するステップと、

前記バッチレンダリングプロセスのための前記コンテンツを提供するステップと、

前記データストアの前記コンテンツを保存するステップと、を含む、

方法。

【請求項 16】

前記バッチレンダリングプロセスによって、前記ウェブページのレンダリング結果を生成する前記コンテンツを使用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 17】

前記レンダリング結果が、レイアウト情報および動的に生成されたコンテンツを含む、請求項 16 に記載の方法。

【請求項 18】

書き換え規則を適用するステップが、前記 URL を、リダイレクト URL と関連しているテンプレートに照合させるステップを含み、

前記 URL が前記テンプレートに一致するとき、前記リダイレクト URL が、前記書き換えられた URL であると判断され、

前記 URL がテンプレートに一致しないとき、前記 URL が、前記書き換えられた URL である、

請求項 15 に記載の方法。

【請求項 19】

変更率または前記データストアに保存された前記埋め込みアイテムのタイプに基づいて、前記書き換えられた URL のための前記コンテンツが古いと判断するステップと、

前記書き換えられた URL のための前記コンテンツが古いと判断する前記ステップに回答して、

前記ウェブ巡回エンジンから更新されたコンテンツを受信するステップと、

前記更新されたコンテンツを備えた前記データストアを更新するステップと、

書き換えられた URL のための前記コンテンツとして、前記更新されたコンテンツを提供するステップと、をさらに含む、

請求項 15 に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、バッチ最適化レンダリング及びフェッチアーキテクチャに関する。

【背景技術】

【 0 0 0 2 】

ワールドワイドウェブは、情報が豊富である。今日では、1兆以上のユニークなウェブページがあると推定されている。これらのページの多くは、例えばNew York Timesのホームページなどに動的に作成され、レンダリングされたウェブページのコンテンツや外観に効果を与えることができる、画像やビデオのような埋め込みコンテンツへのリンクを有する。例えば、ブラウザがJavaScript（登録商標、以下同じ）コードのようなスクリプトを実行するとき、これはブラウザがウェブページのレンダリングを完了した後に、ウェブページがユーザに対してどのように見えるかに影響を与え、ページのコンテンツ及び/又は外観を変更することができる。別の例として、いくつかのウェブページは、テキストの見た目を変える方法をブラウザに指示するスタイルシートを使用する。典型的なウェブページは、そのような追加の埋め込みアイテムのハンドラを有することができ、そのいくつかはブラウザレンダリングエンジンのために特別に設計されるか、あるいはブラウザレンダリングエンジン向けられている。レンダリングプロセスによって生成された追加の情報は、インターネット検索エンジンのような、ダウンストリームシステムに役立つことができる。単一のユーザのウェブブラウザのために、リアルタイムで単一のウェブページをレンダリングすることは比較的容易であるが、ましてやワールドワイドウェブ上の全てのページ（1兆ページ）のように多数のページをレンダリングすること、あるいはワールドワイドウェブ（100億ページ）上の上位1%だけをリアルタイムでレンダリングすることは難しい。

10

【 発明の概要 】

20

【 課題を解決するための手段 】

【 0 0 0 3 】

実施形態は、ウェブページインデックスシステムのような、ダウンストリームユーザのためのウェブページのバッチレンダリングのために最適化されたレンダリングサーバおよびフェッチサーバを含む。ダウンストリームユーザは、1つまたは複数の埋め込みアイテムを有するウェブページ（例えば、そのURLを使用して）を識別するとき、ダウンストリームユーザはレンダリングサーバがレンダリング結果を生成するためにURLをレンダリングすることを要求してもよい。レンダリングサーバは、多数の（例えば、何万もの）レンダリングエンジンを含むことができる。各レンダリングエンジンは、多くのレンダリングエラーを排除するバッチレンダリングのために最適化されたブラウザカーネルをシミュレートする。レンダリングの間、レンダリングエンジンは埋め込みアイテムを発見すると、レンダリングエンジンはフェッチサーバから埋め込みアイテムを要求する。フェッチサーバは、各埋め込みアイテム（例えば、そのURL）のための識別子と、ウェブクローラ（web-crawler）によって検索されたアイテムのためのコンテンツとによって適合された埋め込みアイテムのデータストアを含む。埋め込みアイテムのためのデータストアを探索前に、フェッチサーバは書き換え規則を使用してURLを書き換えてもよい。URLのためのコンテンツが別の埋め込みアイテム（例えば、リダイレクトURLによって表される）と重複しているとき、書き換え規則は、URLをリダイレクトURLと置き換えてもよい。要求された埋め込みアイテムが重複している場合、フェッチサーバは、リダイレクトURLを使用するようにURLを書き換えてもよく、これは要求されたURLのためのコンテンツをフェッチする代わりに、リダイレクトURLのための既に検索されたコンテンツを使用できるようにする。このような重複排除方法は、フェッチサーバによって作られるクローラ要求の実際数を劇的に減少することができ、レンダリングエンジンの応答時間を向上することができる。書き換え規則は、URLがブラックリストに載せられていることを示してもよい。いくつかの実施形態において、フェッチサーバは、埋め込まれた画像の実際のコンテンツではなく、サイズを格納してもよい。レンダリングエンジンが画像を要求したとき、フェッチサーバが画像のサイズを有するモック画像を生成し、モック画像をレンダリングエンジンに返してもよい。レンダリングエンジンがウェブページのレンダリングを完了しているとき、レンダリング結果をウェブページの処理を向上するためにレンダリング結果の情報を使用することができる、インデックスエンジンのようなダウンス

30

40

50

トリームユーザに提供してもよい。

【0004】

1つの態様において、コンピュータシステムは、少なくとも1つのプロセッサと、メモリと、を含むコンピュータシステムであって、メモリが、埋め込みアイテムのためのコンテンツのデータストアと、少なくとも1つのプロセッサによって実行されたときにシステムに動作を行わせる命令を格納する。該動作は、バッチプロセスから、ウェブページをレンダリングするための要求を受信することと、ウェブページの埋め込みアイテムを識別することと、を含む。該動作は、書き換え規則に基づいて、前記埋め込みアイテムが以前にフェッチされた埋め込みアイテムのためのコンテンツの重複であるコンテンツを有すると判定することと、判定に応答して、データストアから前記以前にフェッチされた埋め込みアイテムのための前記コンテンツを提供することと、以前にフェッチされた埋め込みアイテムのための前記コンテンツを使用して、前記ウェブページについてのレンダリング結果を生成することと、およびバッチプロセスにレンダリング結果を提供することと、も含む。

10

【0005】

本明細書に記載される主題の1つ又は複数の実施形態は、以下の特徴の内の1つ又は複数を含むことができる。例えば、埋め込みアイテムが以前にフェッチされた埋め込みアイテムのためのコンテンツの重複であるコンテンツを有すると判断することは、書き換え規則のテンプレートに前記埋め込みアイテムを照合させることを含み、書き換え規則がリダイレクト識別子を含むことができる。このような実施形態において、以前にフェッチされた埋め込みアイテムのための前記コンテンツを提供することが、以前にフェッチされた埋め込みアイテムのための前記コンテンツ及び/又はクエリ文字列なしでURLを含んでもよいテンプレートを示すための前記リダイレクト識別子を使用することを含む。

20

【0006】

他の例として、埋め込みアイテムは、第1の埋め込みアイテムであり、かつ動作がウェブページの第2の埋め込みアイテムを識別することを含んでもよく、前記第2の埋め込みアイテムがブラックリストに載せられているか否かを判断し、第2の埋め込みアイテムが第2の埋め込みアイテムのためのフェッチコンテンツなしでブラックリストに載せられているとき、エラーを返し、および第2の埋め込みアイテムのためのコンテンツなしでレンダリング結果を生成することを含んでもよい。別の例として、レンダリング結果を生成するとき、動作が仮想時計の使用を含んでもよく、仮想時計はリアルタイムとは無関係に進む。別の例において、レンダリング結果を生成するとき、動作が仮想時計の使用を含んでもよく、仮想時計が以前にフェッチされた埋め込みアイテムの提供されたコンテンツを待つ間に進行しない。

30

【0007】

他の例として、埋め込みアイテムは第1の埋め込みアイテムであってもよく、動作がウェブページの第2の埋め込みアイテムを識別することと、第2の埋め込みアイテムが画像を含むことを判断することと、サイズテーブルを使用して、第2の埋め込みアイテムのためのサイズを指定するモック画像を生成することと、レンダリング結果を生成するときにモック画像を使用することと、を含んでもよい。

40

【0008】

別の態様において、コンピュータ実装された方法は、ウェブページをレンダリングするための要求をバッチプロセスから受信するステップと、仮想時計とウェブページをレンダリングするためのタスクリストを初期化するステップであって、仮想時計が、埋め込みアイテムのための要求が未処理であるときかつタスクが準備完了であるとき、依然として停止している、初期化するステップと、を含む。該方法はまた、仮想時計がタスクリストの停止タスクのためのランタイムに一致するとき、ウェブページのためのレンダリング結果を生成するステップと、バッチプロセスのためのレンダリング結果を生成するステップと、を含む。

【0009】

50

本明細書に記載される主題の1つ又は複数の実施形態は、以下の特徴の内の1つ又は複数を含むことができる。例えば、タスクリストを初期化するステップが、ランタイムセットを備えた停止タスクを仮想時計に追加された所定の時間に追加するステップを含んでもよい。所定の時間は少なくとも5秒であってもよい。別の例として、該方法は、埋め込みアイテムに対する要求が未処理ではない、かつ仮想時計よりも大きいランタイムを有するただ一つのタスクがタスクリストにあるとき、仮想時計をタスクリスト内のタスクのランタイムに進めるステップをさらに含んでもよい。別の例として、該方法は、ウェブページの埋め込み画像を識別するステップと、埋め込み画像のためのコンテンツを要求するステップと、要求に応答して、埋め込み画像のためにサイズを指定するが、空のコンテンツを有するモック画像を受信するステップと、レンダリング結果を生成するときにモック画像を使用するステップと、を含んでもよい。別の例として、バッチプロセスは、インデックスエンジンであってもよく、かつ該方法は、レンダリング結果の情報に基づいて、ウェブページのためのランクを降格させるステップをさらに含み、及び/又は動的に生成されたコンテンツにインデックスを付けるためにレンダリング結果を使用するステップをさらに含む。

10

20

30

40

50

【0010】

別の態様において、方法は、ウェブページ内の埋め込みアイテムのユニフォームリソースロケータ(URL)のためのバッチレンダリングプロセスから要求を受信するステップと、書き換えられたURLを判断する書き換え規則を適用するステップを含む。該方法は、書き換えられたURLのためのコンテンツがデータストアに存在するか否かを判断するステップと、コンテンツが存在するとき、コンテンツを前記バッチレンダリングプロセスに提供するステップを含んでもよい。コンテンツが存在しないとき、該方法は、コンテンツのフェッチを初期化し、ここでバッチレンダリングプロセスがフェッチの間にタイムアウトなしで待機するように構成される、初期化するステップと、ウェブ巡回エンジンからコンテンツを受信するステップと、バッチレンダリングプロセスのためのコンテンツを提供するステップと、データストアのコンテンツを保存するステップを含んでもよい。コンテンツは、ウェブページのレンダリング結果をウェイ請求項するためにバッチレンダリングプロセスによって使用されてもよい。

【0011】

本明細書に記載される主題の1つ又は複数の実施形態は、以下の特徴の内の1つ又は複数を含むことができる。例えば、書き換え規則を適用するステップは、URLをリダイレクトURLと関連しているテンプレートに照合させるステップを含んでもよく、ここでURLがテンプレートに一致するとき、前記リダイレクトURLが、書き換えられたURLであると判断され、およびURLがテンプレートに一致しないとき、URLが書き換えられたURLであると判断される。別の例として、該方法は、変更率またはデータストアに保存された埋め込みアイテムのタイプに基づいて、書き換えられたURLのための前記コンテンツが古いと判断するステップと、書き換えられたURLのためのコンテンツが古いと判断するステップに応答して、ウェブ巡回エンジンから更新されたコンテンツを受信するステップと、更新されたコンテンツを備えたデータストアを更新するステップと、書き換えられたURLのためのコンテンツとして、更新されたコンテンツを提供するステップを含んでもよい。

【0012】

別の態様として、コンピュータシステムは、少なくとも1つのプロセッサと、画像識別によって格納されたサイズのテーブルおよび少なくとも1つのプロセッサによって実行されたときにシステムに動作を行わせる命令を保存するメモリと、含む。該動作は、ウェブページの埋め込みアイテムを識別することと、サイズのテーブルから埋め込み画像のためのサイズを判断し、サイズを使用してモック画像を生成することと、を含むことができる。該動作は、モック画像を使用してウェブページのためのレンダリング結果を生成することを含んでもよい。

【0013】

別の態様として、コンピュータ実装された方法は、ウェブページをレンダリングするための要求をバッチプロセスから受信するステップと、ウェブページの少なくとも1つの埋め込み画像を識別するステップと、を含む。該方法は、フェッチサーバからモック画像を受信するステップと、モック画像を使用してウェブページのためのレンダリング結果をレンダリングするステップも含み、モック画像は埋め込み画像のサイズと空のコンテンツを有する。いくつかの実施形態において、該方法は、レンダリング結果をウェブページを要求したバッチプロセスに提供してもよい。

【0014】

別の態様において、非一時的コンピュータ可読媒体は、コンピュータシステムに1つ又は複数の上述した方法を実行させる、回路基板に形成された少なくとも1つのプロセッサによって実行可能な命令を含んでもよい。

【0015】

本明細書に記載される主題の1つまたは複数の実施形態は、1つまたは複数の以下の利点を実現するように実装されることができる。一例として、バッチレンダリングエンジンが入力デバイス（例えば、キーボード、マウス）あるいは出力デバイス（例えば、ディスプレイ、タッチスクリーンなど）に結合されていないため、レンダリングエンジンは、例えばユーザフレンドリーなAPIではなくマシンフレンドリーなAPIを有する、実際のブラウザレンダラ（browser renderer）よりも簡素で洗練されたものにすることができる。また、レンダリングエンジンは、レンダリングされた最終ページを表示する、あるいはユーザと対話する必要がないので、レンダリングエンジンは、実際の時間ではなく終了したタスクに基づいて進む仮想時計を使用することができ、これはレンダリングプロセスの処理速度を上げ、および一般的なエラーを避けることができる。例えば、バッチ環境でのフェッチは、個人のウェブ環境よりもはるかに遅くなることがあり、多くのタイムアウトエラーを引き起こす可能性がある。仮想時計は、フェッチ待ち時間を隠し、タイムアウトエラーを回避する。仮想時計は、より確定的な結果も可能にする。例えば、日付/時刻コンポーネントを含むURLにおいて、日付/時刻コンポーネントを固定された時間に置き換えるのではなく、システムは仮想時計の値を使用してもよい。これは、ウェブページのすべての時間パラメータが同じ値を有するわけではないが、ウェブページがレンダリングされるたびに特定の時間パラメータが同じ値を有することを意味する。この柔軟性は、システムが時間を進めることを可能にし、これはレンダリングされた結果の正確性のためにいくつかのウェブページにおいて重要であるが、要求された一連のURLがレンダリング全体で同じままであることが常に保証される（クロール要求を少なくする）。システムは、例えばブラックリストに載せられたアイテムのような不要なアイテムをフェッチすることを回避してもよい。画像の実際のコンテンツではなく、画像のためのサイズを保存することは、フェッチサーバにおいて画像のためのストレージ要件を削減し、レンダリングエンジンに転送されるデータが少なく済み、およびレンダリングエンジンでレンダリング時間がさらに向上する。URLを書き換えることは、重複したコンテンツをフェッチすることを回避し、バッチレンダリングプロセスをさらに高速にする。

【図面の簡単な説明】

【0016】

【図1】開示された主題に従ったシステム例を図示する。

【図2】埋め込みアイテムを有するウェブページのブロック図である。

【図3】実施形態に従った、バッチレンダリングエンジンのブロック図である。

【図4】実施形態に従った、バッチレンダリングエンジンが、埋め込みオブジェクトを備えたウェブページをレンダリングすることができるプロセス例を図示する流れ図である。

【図5】実施形態に従った、バッチレンダリングエンジンが、仮想時計を進めるプロセス例を図示する流れ図である。

【図6】実施形態に従った、フェッチサーバが埋め込みアイテムのためのコンテンツをバッチレンダリングエンジンに提供するプロセス例を図示する流れ図である。

【図7】実施形態に従った、フェッチサーバがモック画像をバッチレンダリングエンジン

10

20

30

40

50

に提供するプロセス例を図示する流れ図である。

【図 8】開示した技術を実装するために使用されることができるコンピュータデバイスの例を示す。

【図 9】開示した技術を実装するために使用されることができる分散型コンピュータデバイスの例を示す。

【発明を実施するための形態】

【0017】

ウェブページを完全にレンダリングするために、ウェブページの全ての埋め込み外部リソースのコンテンツを最初を取得する必要がある。そのようなリソースは、外部画像、JavaScriptコード、およびスタイルシートを含むが、これらに限定はしない。しばしば、同じ外部リソースが多く異なるウェブページに埋め込まれている。単一のユーザのウェブブラウザがGoogle Analytics JavaScriptコードのような外部ウェブページリソースをリアルタイム（すなわち、リソースが埋め込まれたウェブページがレンダリングされたとき）で要求することは効率的であるが、バッチレンダリングエンジンがそうすることは実現可能でも効率的でもない。例えば、ウェブページインデックスプロセスのためのバッチレンダリングエンジンは、効率的かつ迅速に、一度に多数のウェブページを素早くレンダリングするように設計されている。しかし、埋め込み外部リソースをフェッチすることは、遅くなる可能性があり、そのようなリソースはバッチ処理の目的（例えば、人間のユーザが最終的にレンダリングされた製品を見ることなしに）のために重要ではないことがある。バッチ環境においてウェブページをレンダリングする処理時間を改善するために、レンダリングエンジンは、仮想時計を使用して動作し、重複および不要なフェッチを避けるためにフェッチサーバと連動し、および視覚の処理あるいはウェブページにおける他のユーザ指向の要素を最小限にしてもよい。

【0018】

図 1 は、実施形態例に従ったシステムのブロック図である。システム 100 は、要求プロセスのためにバッチモードでウェブページを効率よく迅速にレンダリングするために使用されてもよい。システム 100 に図示される要求プロセスは、インターネット検索エンジンのためのインデックスエンジンであるが、実装はレンダリングされたウェブページのダウストリームユーザとしてのインデックスエンジンに限定しない。例えば、要求プロセスは、ページを分析して遅さを障害追跡する分析エンジンであってもよく、あるいはGoogle Analyticsのようなツールが正しくセットアップされているかを判断する分析エンジンであってもよく、または広告システム、または複雑なウェブページとの、例えばフォームの記入や要素のクリックなどの自動対話に依存する別のシステムであってもよい。したがって、システム 100 は、インデキシングのためにバッチ生成されたレンダリング結果を使用するものとして説明されてもよく、システム 100 は、レンダリング結果で提供される情報が有用である他のバッチシステムのために使用されることができる。

【0019】

システム 100 は、コンピューティングデバイスであってもよく、あるいは多くの異なるデバイスの形を有するデバイスであってもよい。例えば、システム 100 は、単打サーバ、そのようなサーバのグループ、クライアントサーバシステム、あるいはラックサーバシステムであってもよい。さらに、システム 100 は、パーソナルコンピュータで実装されてもよい。システム 100 は、図 8 で図示するような、コンピュータデバイスの例 800、あるいは図 9 で図示するようなコンピュータデバイス 900 であってもよい。

【0020】

システム 100 は、ウェブ巡回エンジン 130、インデックスエンジン 110 のような要求プロセス、レンダリングサーバ 140、およびフェッチサーバ 150 を含む。ウェブ巡回エンジン 130、レンダリングサーバ 140、およびフェッチサーバ 150 は、ワールドワイドウェブ上に見られるウェブページのような多数のウェブページを効率的にレンダリングするためにともに動作してもよい。

【 0 0 2 1 】

インデックスエンジン 1 0 0 は、インデックス 1 1 5 を作成するために、1 つまたは複数の機械実行可能な命令またはソフトウェアの一部、ファームウェア、またはそれらの組合せを実行するように構成された 1 つ又は複数のプロセッサを含むことができる。たとえば、インデックスエンジン 1 1 0 は、ウェブ巡回エンジン 1 3 0 を介してサーバから情報を受信してもよい。インデックスエンジン 1 1 0 は、インデックス 1 1 5 を生成するために受信したウェブページのコンテンツを処理してもよい。サーバ 1 9 0 は、1 つまたは複数のウェブページまたは 1 つ又は複数のウェブページに埋め込まれたリソースをホストするインターネットを介してアクセス可能な任意のタイプのコンピューティングデバイスであってもよい。巡回エンジン 1 3 0 によってアクセスされるウェブページは、スタイルシート、Java Script、画像などの埋め込みアイテムを含んでもよく、そのうちのいくつかは、レンダリングされたウェブページのコンテンツおよびレイアウトを変更してもよい。インデックスエンジン 1 1 0 は、巡回エンジン 1 3 0 を介して提供されたものにインデックス付けができるが、インデックスエンジンは、レンダリングサーバ 1 4 0 に、レイアウト情報およびインデックスエンジン 1 1 0 にそれ以外は利用できない動的コンテンツを含むブラウザレンダリングされたウェブページのレンダリング結果を提供するように要求することができる。インデックスエンジン 1 1 0 は、インデックス 1 1 5 のドキュメントについて利用可能な情報を向上するためにレンダリング結果コンテンツを使用することができる。例えば、インデックスエンジン 1 1 0 は、ウェブページの画像のテキストの位置あるいはサイズに基づいて、ウェブページのテキスト要素のランクを変更してもよい。例として、一大ニュースを表すテキスト（例えば、スクロールなしで見ることができる）は、広告のテキストよりもより重要であると考えられ得る。別の例として、広告内のテキストは、ウェブページにとってあまり重要ではないと考えられ得る。さらに、いくつかのコンテンツが動的に生成されるとき、例えばウェブページがレンダリングされるまで利用可能でないとき、インデックスエンジン 1 1 0 は、レンダリング結果を動的に生成されたコンテンツをインデックス付けするために使用してもよい。簡潔にするために図 1 には示されていないが、いくつかの実施形態において、インデックスエンジン 1 1 0 は、1 つまたは複数の別個のコンピューティングデバイスに分散されてもよい。

10

20

【 0 0 2 2 】

インデックスエンジン 1 1 0 と同様に、クエリエンジン 1 2 0 は、例えば従来または別の情報検索技術を使用して、クエリ 1 8 2 の検索結果を識別するためにインデックス 1 1 5 を使用する 1 つ又は複数のサーバを含んでもよい。クエリエンジン 1 2 0 は、クライアント 1 8 0 のようなリクエスタからクエリ 1 8 2 を受信する 1 つまたは複数のサーバを含んでもよい。クエリエンジン 1 2 0 は、インデックス 1 1 5 を使用して、クエリに回答してドキュメントを識別してもよく、および検索結果 1 8 4 として応答ドキュメントからの情報をリクエスタに提供してもよい。いくつかの実施形態において、クエリエンジン 1 2 0 は、検索結果 1 8 4 の一部としてサムネイルを提供するために、レンダリング結果データストア 1 4 8 のレンダリング結果を使用してもよい。クエリエンジン 1 2 0 は、例えば、1 つまたは複数のランク付け信号を使用して、クエリに回答してドキュメントのためのスコアを計算するランク付けエンジンを含んでもよい。1 つまたは複数のランク付け信号は、ドキュメントに関連するレンダリング結果から取得されたコンテンツに基づくことができる。ランク付けエンジンは、スコアを使用してクエリに回答して見つかったドキュメントをランク付けしてもよい。

30

40

【 0 0 2 3 】

システムは、ウェブ巡回エンジン 1 3 0 を含んでもよい。ウェブ巡回エンジン 1 3 0 は、1 つまたは複数の機械実行可能命令あるいはソフトウェアの一部、ファームウェア、あるいはそれらの組合せを実行するように構成された 1 つ又は複数のプロセッサを含むことができる。ウェブ巡回エンジン 1 3 0 は、標準サーバ、そのようなサーバのグループ、クライアントサーバシステム、あるいはラックサーバシステムのようなコンピューティングデバイスであってもよい。いくつかの実施形態において、ウェブ巡回エンジン 1 3 0 は、

50

フェッチサーバ150またはインデックスエンジン110のようなシステム100の別のコンポーネントと、メモリまたはハードウェアプロセッサのようなコンポーネントを共有してもよい。ウェブ巡回エンジン130は、ワールドワイドウェブ上で見つけれられることができるウェブページをクロールしてもよい。ウェブ巡回エンジン130は、クロールされたウェブページを受信したとき、すなわち、クロールされたウェブページのためのコンテンツを受信すると、ウェブ巡回エンジン130は、インデックスエンジン110あるいはフェッチサーバ150であってもよいリクエストにコンテンツを提供してもよい。ウェブ巡回エンジン130は、データストア（図示せず）にコンテンツを格納し、およびリクエストにその位置を提供してもよい。本明細書で使用されるウェブページのコンテンツは、ウェブページレンダリングエンジンに提供されかつウェブブラウザに表示するためのウェブページをレンダリングするために使用されるHTMLコードを参照し、例えばスタイルシート、Java Script、別のウェブページ、あるいは画像ファイルなどのウェブページに埋め込まれた外部オブジェクトへの任意のリンクを含む。ウェブ巡回エンジン130は、これらの埋め込みアイテムのコンテンツをフェッチするためにフェッチサーバ150によって使用されてもよい。ウェブ巡回エンジン130は、フェッチサーバ150に埋め込みアイテムのコンテンツを提供してもよく、あるいは埋め込みアイテムテーブル152のようなデータストアにフェッチされたコンテンツを格納してもよい。ウェブ巡回エンジン130は、埋め込みアイテムがクロールされたときにリクエストに通知してもよい。

【0024】

前述したように、システム100は、フェッチサーバ150を含む。フェッチサーバ150は、1つまたは複数の機械実行可能な命令またはソフトウェアの一部、ファームウェア、あるいはそれらの組合せを実行するように構成された1つまたは複数のプロセッサを含むことができる。フェッチサーバ150は、標準サーバ、そのようなサーバのグループ、クライアントサーバシステム、あるいはラックサーバシステムのようなコンピューティングデバイスであってもよい。いくつかの実施形態において、フェッチサーバ150は、レンダリングサーバ140、ウェブ巡回エンジン130、またはインデックスエンジン110のようなシステム100の別のコンポーネントと、メモリまたはハードウェアプロセッサのようなコンポーネントを共有してもよい。フェッチサーバ150は、ウェブ巡回エンジン130が特定の埋め込みアイテムのコンテンツを、例えばそのURLによってフェッチすることを要求し、要求された埋め込みアイテムのフェッチされたコンテンツとクロール時間を受信するように構成される。フェッチサーバ150は、ウェブ巡回エンジン130から直接、またはウェブ巡回エンジンが更新する埋め込みアイテムテーブル152から、コンテンツとフェッチ時間を受信してもよい。フェッチサーバ150は、レンダリングエンジン142から埋め込みアイテムのための要求を受信してもよい。フェッチサーバ150は、要求しているレンダリングエンジン142に応答を提供してもよい。応答は、埋め込みアイテムテーブル152にフェッチされた、または格納されたコンテンツ、画像寸法テーブル156に基づくモック画像、あるいはエラー応答を含んでもよい。いくつかの実施形態において、フェッチサーバ150は、埋め込みアイテムのコンテンツおよびクロール時間を、埋め込みアイテムを要求したレンダリングサーバ140のレンダリングエンジン142に送信することによって、埋め込みアイテムのコンテンツを提供することができる。あるいは、フェッチサーバ150は、埋め込みアイテムテーブル152の指定の位置を介して埋め込みアイテムのコンテンツおよびクロール時間が利用可能であることをレンダリングエンジン142に通知することによってコンテンツを提供することができ、およびレンダリングエンジン142は、データストアからウェブページのコンテンツおよびクロール時間を取得することができる。

【0025】

フェッチサーバ150は、要求された埋め込みアイテム（例えば、要求されたURL）にURL書き換え規則154を適用してもよい。URL書き換え規則154は、URLが別のURLと同一のコンテンツに関連付けられているとき、URLを書き換えるための規則を含む。これは、ウェブサイトの所有者が、リソースがリクエストされる度にコンテンツをダウンロード

することをブラウザに望み、そのために動的に生成されたURLあるいはキャッシュバusting (cache-busting) URLを提供する場合によく発生する。このようなURLは、URLの一部としてタイムスタンプまたは埋め込まれたランダム文字列を有することが多く、例えば、キャッシュバusting URLを生成するJavaScriptによってウェブページがレンダリングされる度に、URLが一意になるようにする。しかしながら、ホスティングサーバから提供された動的に生成されたURLのためのコンテンツは、変更されない、あるいはバッチレンダリングの目的で意味のある方法で変更されない。フェッチサーバ150は、埋め込みアイテムに対する要求により効率的に応答するためにURL書き換え規則154を使用してよい。例えば、URL書き換え規則154は、パターンまたはテンプレートを含んでもよく、および規則のテンプレートに一致するURLは、同一のコンテンツ、例えば重複コンテンツを返す。いくつかの実施形態において、テンプレートは、オフライン、またはフェッチログを使用して様々なURLのコンテンツを比較し重複コンテンツを有するURLに共通のURLのパターンを識別するバッチプロセスによって、決定されてもよい。フェッチログは、例えばウェブ巡回エンジン130またはフェッチサーバ150によって維持されてもよい。テンプレートは、ユーザ入力であってもよい。要求された埋め込みアイテムがテンプレートのうちの一つに一致するURLを有する場合、URL書き換え規則154は、フェッチサーバ150に要求されたアイテムが重複であることを伝えてもよく、要求されたURLをリダイレクトURLで書き換えるように命令してもよい。これは、以前にフェッチされたコンテンツ、例えば埋め込みアイテムテーブル152のコンテンツを有するURLと関連する。以前にフェッチされた埋め込みアイテムのURLは、リダイレクトURLと見なされてもよい。これは、フェッチサーバが150が、必要のないフェッチを避けることを可能にし、要求しているバッチレンダリングエンジン142への応答を迅速化し、過度のフェッチ要求によって引き起こされるホスティングサーバへのストレスを取り除く。もちろん、要求されたURLがURL書き換え規則154のテンプレートに一致しない場合、URLを書き換えることは、要求されたURLは変更されない結果となる。

10

20

30

40

50

【0026】

URL書き換え規則154は、ブラックリストに載せられたURLのパターンあるいはテンプレートを含んでもよい。要求された埋め込みアイテムがブラックリストに載せられたURLパターンと一致する場合、システムは、URLのためのコンテンツをフェッチしようとするのではなく、所定のエラーを返してもよい。ブラックリストに載せられたURLは、コンテンツがバッチレンダリングの目的のために必要とされないことを決定したあとに、ユーザによってURL書き換え規則154にURLを入力されてもよい。この一例は、多くのウェブページに含まれるGoogle Analytics JavaScriptコードである。このJavaScriptコードは、レンダリングされたページのレイアウトに重要であるとは見なされなくともよく、バッチレンダリングエンジンの目的のために実行される必要はない。したがって、レンダリング効率のために、いくつかの埋め込みアイテムはURL書き換え規則154を使用してブラックリストに載せられてもよい。いくつかの実施形態において、ブラックリストに載せられたURLのためのエラーを返すのではなく、システムは、期限切れにならない、および埋め込みアイテムに適した所定のコンテンツを有する埋め込みアイテムテーブル152のエントリに、上述したリダイレクトURLを使用してURLを書き換えてもよい。いくつかの実施形態において、URL書き換え規則は、テンプレートに一致するときブラックリストに載せられたものとしてURLにフラグを立ててもよい。URL書き換え規則154は、ウェブ巡回エンジン130を介してフェッチされた埋め込みアイテムの数を劇的に減少することができ、フェッチサーバ150のリソースのための任意の要求への応答時間を改善し、特定のサーバ190のフェッチ量を最小限にする。フェッチ量を最小限にすることは、システムがフェッチ要求でサーバ190に負担をかけないことを保証する。いくつかの実施形態において、フェッチサーバ150及び/又はウェブ巡回エンジン130は、サーバ190に向けられたフェッチ要求の数を制限するように構成されてもよく、要求が制限を超える場合、システムは要求をキューに入れ始めてもよい。キューが大きすぎる場合、システムはフェッチ要求を失敗する可能性がある。従って、URL書き換え規則154はまた、フェッチ量を

最小化することができる。

【0027】

いくつかの実施形態において、フェッチサーバ150は画像寸法テーブル156を含んでもよい。画像寸法テーブル156は、画像URLを画像のための既知のサイズと関連付けるキー値記憶装置であってもよい。既知のサイズは、画像がフェッチされたときに決定されてもよい。要求された画像のサイズを使用して、フェッチサーバ150は要求された画像と同じサイズを有するが、空のコンテンツであるか、コンテンツとして単純なタイルを有するモック画像を生成してもよい。モック画像は、要求された画像と同じサイズを有するが同じ画像データではない、有効な画像である。フェッチサーバ150はバッチレンダリングエンジンのためのコンテンツをフェッチするので、実際の画像はレンダリング結果にとって重要ではなくてもよいが、画像のサイズはレンダリングされたページのレイアウトに影響を及ぼし得る。実際の画像ではなくモック画像を使用することは、ファイルサイズが非常に小さくなり（例えば、画像あたり数十バイト）、これはモック画像を転送する際のネットワーク帯域幅、およびバッチレンダリングエンジンのためのプロセッサおよびメモリリソースを節約する。いくつかの実施形態において、画像寸法テーブル156は、SSTableなどのキー値記憶装置であってもよいが、寸法テーブル156は、画像識別子によってサイズを格納する任意のデータ構造体であってもよい。

10

【0028】

システム100は埋め込みアイテムテーブル152を含んでもよい。埋め込みアイテムテーブル152は、URLによってキー付されてもよく、およびウェブ巡回エンジン130から返された埋め込みアイテムのためのフェッチされたコンテンツを格納してもよい。いくつかの実施形態において、埋め込みアイテムテーブル152は、クロール履歴も格納してもよい。例えば、いくつかの実施形態において、埋め込みアイテムテーブル152は、例えば7日間、2週間などの期間にわたってフェッチされたコンテンツを含んでもよい。埋め込みアイテムテーブル152は、クロール履歴に基づいて変更率を含んでもよい。いくつかの実施形態において、埋め込みアイテムテーブル152は、BigTable、関係型データベース、Hadoop Distributed Fileなどとして実装されてもよい。フェッチサーバ150は、以前にフェッチされた埋め込みアイテムのためのコンテンツを素早く返すために、埋め込みアイテムテーブル152を使用してもよい。フェッチサーバ150は、何千ものバッチレンダリングシステムに対する要求を処理することができるので、要求された埋め込みアイテムが前のフェッチ要求に応答して以前にフェッチされている可能性が高い。フェッチされたコンテンツが埋め込みアイテムテーブル152に位置するとき、ウェブ巡回エンジン130にコンテンツを提供するように求めるのではなく、フェッチサーバ150は埋め込みアイテムテーブル152のコンテンツを使用して要求に応答してもよい。これにより、フェッチされたコンテンツを格納するサーバ190の負担が軽減され、フェッチサーバ150が埋め込みアイテムのための要求により迅速に応答することを可能にする。フェッチサーバ150は、URL書き換え規則154を使用して、重複を排除するURLによってクロール要求をさらに減らすことができる。

20

30

【0029】

レンダリングプロセスの任意の段階で、1つ又は複数の要求された埋め込みアイテムのコンテンツが埋め込みアイテムテーブル152に格納されていない、あるいは古い場合、フェッチサーバ150は、要求された埋め込みアイテムのクロールをスケジュールするためにウェブ巡回エンジン130に指示してもよい。一旦ウェブ巡回エンジン130がレンダリングされた埋め込みアイテムをクロールすると、フェッチサーバ150に通知する。フェッチサーバ150は、次いでフェッチされたコンテンツアイテムをクロール時間とともに埋め込みアイテムテーブル152に格納してもよい。埋め込みアイテムが画像である場合、フェッチサーバ150は、代わりに、または追加で、フェッチされた画像のサイズをクロール時間とともに画像寸法テーブル156に格納してもよい。フェッチサーバ150は、次いで要求されたコンテンツを、すなわち画像サイズを有するモック画像が送信されてもよい画像ファイルの代わりに、要求しているレンダリングエンジン142に送り返

40

50

してもよい。

【0030】

システム100は、レンダリングサーバ140を含む。レンダリングサーバ140は、1つまたは複数の機械実行可能な命令またはソフトウェアの一部、ファームウェア、あるいはそれらの組合せを実行するように構成された1つまたは複数のプロセッサを含むことができる。レンダリングサーバ140は、標準サーバ、そのようなサーバのグループ、クライアントサーバシステム、あるいはラックサーバシステムのようなコンピューティングデバイスであってもよい。いくつかの実施形態において、レンダリングサーバ140は、レンダリングサーバ140、フェッチサーバ150、またはインデックスエンジン110のようなシステム100の別のコンポーネントと、メモリまたはハードウェアプロセッサのようなコンポーネントを共有してもよい。レンダリングサーバ140は、特定のウェブページをレンダリングするために、インデックスエンジン110または別の要求プロセスから要求を受信する。言い換えると、レンダリングサーバ140は、要求されたウェブページのURLを受信してもよい。レンダリングサーバ140は、1つ又は複数のレンダリングエンジン142を含んでもよい。いくつかの実施形態において、レンダリングサーバ140は、数万のレンダリングエンジン142を含み、およびウェブページをレンダリングするレンダリングエンジン142を選択するためにロード場合によってはランシングを実行してもよい。一旦レンダリングエンジン142が選択されると、レンダリングエンジン142は、レンダリング結果のためにウェブページをレンダリングしようとする。ウェブページは、通常は追加の埋め込みアイテムを含むので、埋め込むウェブページと呼ばれてもよい。

【0031】

各レンダリングエンジン142は、パーソナルウェブブラウザ用のレンダラをエミュレートするように構成されるが、バッチレンダリングのための最適化を伴う。従って、レンダリングエンジン142が埋め込むウェブページを受信したあと、レンダリングエンジン142がレンダリング結果を生成するために行う作業を表すタスクをタスクリストに投入し始めてもよい。多くのタスクがすぐに実行されるようにスケジュールされていてもよいが、いくつかのタスクは将来スケジュールされてもよい。レンダリングエンジン142のバッチ最適化の一つは、仮想時計を使用し、レンダリングが特定の時間に完了したことを示すタスクリストにタスクを追加することである。例えば、いくつかの実施形態において、タスクは、レンダリングが現在の時間+20秒で完了することを示してもよい。所定の時間は、多数のウェブページ設計者が完全に見えるようにウェブページを設計する時間に基づいてもよく、例えば、任意のアニメーションあるいはレイアウト変更が所定の時間内に終わるように設計される。ほとんどのユーザは、ページが読み込まれるまで非常に長く待つことを快く思わないので、所定の時間は5から20秒の間であってもよいが、状況によってはそれより長くなる可能性がある。レンダリングエンジン142は、仮想時計を使用するために20秒全部かかることはなく、および埋め込みアイテムがクロールされていない場合（例えば、フェッチサーバ150が埋め込みアイテムテーブル152のコンテンツを示すことができる）、しばしばフルレンダリングが数秒で発生する可能性がある。従って、最終的なレンダリング結果を生成するタスクは、現在の時間から20秒の開始時刻でタスクリストに追加されてもよい。現在の時間は、仮想時計の初期化時間に基づいており、これはゼロまたは実際の時計の現在の時間とすることができる。

【0032】

レンダリングプロセスの一部、例えば、レンダリングタスクの一つとして、レンダリングエンジン142は、埋め込むウェブページが、スタイルシート、画像ファイル、JavaScriptなどの任意の埋め込みアイテムを含むか否かを決定してもよい。これらの埋め込みアイテムは、プライマリ埋め込みオブジェクトと呼ばれる。ウェブページが任意の埋め込みオブジェクトを含まない場合、レンダリングエンジンは、ウェブページをレンダリング結果に即座に処理することができ、およびレンダリング結果をレンダリング結果データストア148に格納してもよい。しかし、ウェブページが埋め込みアイテムを含む場合、レ

レンダーリングエンジン 142 は、埋め込みアイテムをすべて抜き出し、埋め込みアイテムのコンテンツのために要求をフェッチサーバ 150 に送信してもよい。要求された埋め込みアイテムは、それぞれの URL によって各々表される。レンダーリングエンジン 142 は、しかしながら、フェッチされたリソースを待つ間、レンダーリングを停止するか、タイムアウトすることはない。むしろ、レンダーリングエンジン 142 は仮想時計を使用するので、以下でより詳細に説明するように、ウェブ巡回エンジン 130 を介してフェッチされるリソースを待つことは、時計を進めず、レンダーリングエンジンはタイムアウトしない。

【0033】

要求された埋め込みアイテムのためのコンテンツが受信されたとき、レンダーリングエンジン 142 は、コンテンツを処理するためにタスクリストにタスクを追加してもよい。コンテンツ処理の一部は、要求された埋め込みオブジェクト（すなわち、プライマリ埋め込みオブジェクト）自身が埋め込みオブジェクト（すなわち、セカンダリ埋め込みオブジェクト）を有するか否かを見つけることを含んでもよい。プライマリ埋め込みオブジェクトがセカンダリ埋め込みオブジェクトを含まない場合、レンダーリングエンジン 142 は、画像のプロパティを変更するレンダーリングタスク（例えば、JavaScript コードを実行する）に取り組むことを続けることができる。しかし、プライマリ埋め込みオブジェクトが 1 つ又は複数のセカンダリ埋め込みオブジェクトを含む場合、レンダーリングエンジン 142 は、フェッチサーバ 150 からセカンダリ埋め込みオブジェクトを要求する。埋め込みオブジェクトを発見し要求するこのプロセスは、レンダーリングエンジンが、レンダーリングされるウェブページに埋め込まれた全てのオブジェクト（例えば、プライマリ、セカンダリ、第 3 の、など）を発見し、要求し、および受信するまで、繰り返される。

【0034】

各埋め込みアイテム要求は、フェッチサーバ 150 が一旦その時間のコンテンツを返すと削除されるタスクリスト内のタスクであってもよい。コンテンツが返されるとき、レンダーリングエンジン 142 は、コンテンツを処理するためのタスクを追加してもよく、これは、画像の不透明度の変更、スクリプトの実行など、追加のタスクを順に追加してもよい。各タスクは、ランタイムに関連付けられてもよい。いくつかのタスクは、将来のランタイムを有してもよい。例えば、イメージをフェードイン（またはアウト）するために、ブラウザはタスクリストにいくつかのタスクを追加してもよく、時間のインターバルにわたって画像の不透明度をそれぞれ変更する。以下でより詳細に説明するように、レンダーリングエンジン 142 は、タスクが実行準備が整った時点を決断するために、タスクリストに関してリアルタイムではなく仮想時計を使用してもよい。

【0035】

レンダーリングエンジン 142 は、レンダーリングが完了するまで、例えば、レンダーリング結果が生成されるまで、タスクリストのタスクをレンダーリングするプロセスに取り組む。次いでレンダーリングエンジン 142 は、レンダーリング結果をレンダーリング結果データストア 148 に格納し、及び / 又はレンダーリング結果をレンダーリングプロセス（例えば、インデックスエンジン 110）に提供してもよい。インデックスエンジン 110 のようなレンダーリングプロセスは、次いでウェブページを処理する際にレンダーリング結果から抽出された情報を使用してもよい。例えば、要求プロセスは、JavaScript エラー、レイアウト情報、スタイル情報、広告スペース情報、フェッチされたリソースのリスト、性能統計等を使用してもよく、そのすべてはレンダーリング結果に含まれるが、要求プロセスにとってはそれ以外使用できない。

【0036】

システム 100 は、ネットワーク 170 を介してクライアント 180 およびサーバ 190 と通信してもよい。ネットワーク 170 は、例えばインターネット、または有線あるいは無線ローカルエリアネットワーク（LAN）であってもよいネットワーク 170、ワイドエリアネットワーク（WAN）、これらの組合せであってもよく、例えばゲートウェイデバイス、ブリッジ、スイッチ、あるいは同様のものを使用して実装されてもよい。ネットワーク 170 を介して、クエリエンジン、ウェブ巡回エンジン 130 及び / 又はフェッ

チサーバ 150 は、クライアント 180 及び / 又はサーバ 190 と通信し、かつデータを送受信してもよい。

【0037】

システム 100 は、簡潔にするために図示されていない他のコンポーネントを含んでもよい。例えば、インデックスエンジン 110、クエリエンジン 120、ウェブ巡回エンジン 130、レンダリングサーバ 140、およびフェッチサーバ 150 のうちの 1 つまたは複数は、1 つまたは複数のコンピューティングデバイスにわたって分散されてもよい。同様に、インデックス 115、レンダリング結果データストア 148、埋め込みアイテムテーブル 152、および画像寸法テーブル 156 は、複数のコンピューティングデバイスにわたって格納されてもよい。いくつかの実施形態において、システム 100 の様々なコンポーネントは、コンピューティングデバイスのハードウェアコンポーネントを共有してもよく、あるいは同じコンピューティングデバイスの論理パーティションであってもよい。

【0038】

図 2 は、埋め込みオブジェクトを有するウェブページのブロック図である。図に示すように、ウェブページ 200 は、複数の埋め込みアイテムを含むことができる。これらの埋め込みオブジェクトは、他のウェブページ 210、スタイルシート 220、画像ファイル 230、いわゆるキャッシュバusting URL 240、および JavaScript コード 250 を含むことができるが、これらに限定されない。追加された、埋め込みオブジェクトの異なるタイプはもちろん可能である。さらに、ウェブページ 200 に埋め込まれたそれぞれのオブジェクトは、他のオブジェクトを埋め込んでもよい。例えば、ウェブページ 200 に埋め込まれたウェブページ 210 は、他のウェブページ、画像ファイル、スタイルシートなどを埋め込んでもよい。同様に、ウェブページ 200 に埋め込まれたスタイルシート 220 は、背景画像ファイルのような他のオブジェクトを埋め込んでもよい。さらに、ウェブページ 210 またはスタイルシート 220 に埋め込まれたオブジェクトのそれぞれは、さらにオブジェクトを埋め込んでもよい。このようなウェブページを画像ファイルに完全にレンダリングするために、バッチレンダリングエンジンは、埋め込みオブジェクト 210 - 250 (プライマリ埋め込みオブジェクト) のそれぞれを要求し、埋め込みオブジェクト 210 - 250 に埋め込まれたすべてのオブジェクト (セカンダリ埋め込みオブジェクト)、および埋め込みオブジェクト 210 - 250 に埋め込まれたオブジェクトに埋め込まれたすべてのオブジェクト (第 3 埋め込みオブジェクト) およびその他を要求しなければならない。

【0039】

上述したように、個々のユーザのウェブブラウザはこれらの埋め込みオブジェクトの全てを効率的に要求し、および完全にレンダリングするためにそれらを使用し、およびリアルタイムでウェブページ 200 を表示することができるが、バッチレンダリングエンジンは、重複あるいは不要なコンテンツをフェッチしないように、埋め込みオブジェクトのコンテンツを待ってタイムアウトしないように、およびタスクの内部タイミングに関係なく、できるだけ早くレンダリングが完了するように、最適化されることができる。したがって、多数のクロールされたウェブページを効率的にレンダリングするために、図 1 に開示されたようなシステムが使用されることができる。

【0040】

図 3 は、一実施形態に従った、バッチレンダリングエンジン 142 のいくつかのコンポーネントのブロック図である。バッチレンダリングエンジンは、図 3 に示されていない追加のコンポーネントを含んでもよい。バッチレンダリングシステム 142 は、ページタスクリスト 305、仮想時計 310、およびレンダリング結果 315 を含む。仮想時計 310 は、ウェブページをロードするためにタイムラインを歪ませ、フェッチされたリソースを待つことによって発生する可能性がある多数のエラーを回避するために使用されてもよい。仮想時計 310 は、レンダリングプロセスの開始時間にゼロまたは現在のクロック時間に初期化されてもよく、およびレンダリングエンジンが埋め込みアイテムのフェッチを待っていないときおよびページタスクリスト 305 に現時点で実行の準備ができているタ

スクがあるときのみ進んでもよい。仮想クロックが進められたとき、レンダリングエンジン 142 はページタスクリスト 305 に基づいて仮想時計 310 を進める。言い換えると、レンダリングエンジン 142 は、仮想時計 310 を次に発生するタスクによって表される時間に進める。この意味において、埋め込みアイテムをフェッチして Java Script を実行することは仮想時間がかからず、ライブ（または個人）ブラウザによって発生したエラーのクラス全体を回避することができる。さらに、レンダリングプロセスは、タスクリストで指定された時間よりもずっと早くリアルタイムで終了してもよい。たとえば、“最終レンダリングを生成する”タスクは、20秒で発生するように設定されているが、仮想時計は通常は、ページタスクリスト 305 のタスクを実際に完了するのにかかる時間に応じて、実際の数秒で20秒進む。ページタスクリスト 305 の“最終レンダリングを生成する”タスクは、レンダリングが完了したとき、バッチレンダリングエンジン 142 に伝える停止タスクの一例である。

【0041】

レンダリングエンジン 142 は、埋め込むウェブページのレンダリング結果 315 をレンダリングしてもよい。レンダリング結果 315 は、様々なコンポーネントを含んでもよい。たとえば、レンダリング結果 315 は、レンダリングされたページの画像 316 を含むことができる。画像 316 は、ライブ（または個人）ウェブブラウザのユーザに表示される画像であってもよく、たとえばレンダリングされたページのサムネイルをユーザに表示するために使用されることができる。レンダリング結果 315 は、ドキュメント・オブジェクト・モデル（Document Object Model、DOM）ツリー 317 も含むことができる。DOM ツリー 317 は、ウェブページの HTML 構造を表す。たとえば、システムは、DOM ツリーを処理することによって、トークンまたはユーザに見えるドキュメントのテキストを決定してもよい。レンダリング結果 315 は、レイアウト 318 を含んでもよい。レイアウト 318 は、ウェブページの各要素に対するボックスを含み、ボックスは画像 316 の要素の座標を決定する。たとえば、レイアウトは、DOM ツリーの DOM ノードのボックス表現を含むことができる（すべての DOM 要素が対応するレンダリングボックスを有するわけではない）。ボックスは、レンダリングツリーと呼ばれるツリー構造で編成されることができる。したがって、たとえば、テーブルは、レイアウトのボックスによって表されてもよく、パラグラフはレイアウトの別のボックスによって表されてもよい。したがって、レイアウト 318 は、ウェブページ上のどこに要素が発生するか、ウェブページ上でどれくらいのス
 30
 ペースを要するのか、などの指示を提供する。したがって、レイアウト 318 は、ウェブページ上のどれだけが広告であるか、パラグラフがどれくらい目立つか（たとえば、アバブ・ザ・フォールド（above-the-fold）またはビロウ・ザ・フォールド（below-the-fold））、要素が見えるか否かなどの情報を提供する。したがって、レイアウト 318 は、ウェブページの要素についての幾何学的情報を提供する。レンダリング結果 315 は、エラー 320 を含んでもよい。エラー 320 は、たとえば Java Script などのスクリプトの実行結果として引き起こされるエラーを含む。レンダリング結果 315 は、レンダリングの間にフェッチされた埋め込みリソース 319 のリストを含んでもよく、レンダリングプロセスの一部として生成された別の要素を含むことができる。したがって、レンダリング結果 315 は、ホスティングサーバからのコンテンツのフェッチを介してだけ、要求プロセスに利用可能でない情報を提供する。たとえばインデックスエンジンは、インデックスの要素をランク付けし、目に見えない要素を断片の一部として提供し、および動的に生成されたコンテンツにインデックス付けするためにレンダリング結果情報を使用することができる。動的に生成されたコンテンツは、ウェブページをレンダリングしたあとに存在するコンテンツであり、クロールされたコンテンツには存在しない。

【0042】

図 4 は、実施形態に従った、バッチレンダリングエンジンが、埋め込みオブジェクトを備えたウェブページをレンダリングすることができるプロセス例 400 を図示する流れ図である。プロセス 400 は、図 1 のシステム 100 のようなシステムによって実行されてもよい。広告システム、またはインターネットインデックスシステムのようなシステムは

10

20

30

40

50

、ダウンストリームプロセスの要求でバッチモードのウェブページのレンダリング結果を生成するために、プロセス400を使用してもよい。いくつかの実施形態において、プロセス400は、レンダリングサーバのバッチレンダリングエンジンによって実行されてもよく、および要求プロセスからの要求に応答して開始されてもよい。

【0043】

プロセス400は、ウェブページをレンダリングする要求を受信することから開始してもよい(405)。いくつかの実施形態において、要求は、要求されたウェブページのURLおよび/またはフェッチされたコンテンツと、関連するメタデータ(たとえばクロール時間)とを含んでもよい。いくつかの実施形態において、ウェブページのコンテンツを受信するのではなく、バッチレンダリングエンジンは、ウェブページのコンテンツがデータベースで利用可能であるという通知を受信し、データベースからコンテンツおよび関連するメタデータ(たとえば、クロール時間)を取り出すことができる。フェッチされたコンテンツは、たとえば、要求されたプロセスがすでにコンテンツをフェッチしているという理由で提供されてもよい。バッチレンダリングエンジンは、仮想時計を初期化し、タスクリストに停止タスクを追加することによって開始してもよい(410)。たとえば、バッチレンダリングエンジンは、仮想時計をゼロに設定し、レンダリングエンジンが所定の時間で完了したとレンダリングエンジンに判断させる停止タスクをタスクリストに追加してもよい。この停止タスクに関連付けられたランタイムは、個々のユーザのマシンでほとんどのウェブページが読み込みを完了までの時間であってもよい。たとえば、時間は15または20秒であってもよい。レンダリング開始の一部として、バッチレンダリングエンジンは、ウェブページのコンテンツをフェッチする(コンテンツが提供されなかった場合)、およびウェブページのコンテンツを処理する、などの他のタスクをタスクリストに追加してもよい。これらのタスクは、仮想時間ゼロで追加されてもよく、したがって、たとえばすぐに開始することができる。

【0044】

次いでバッチレンダリングエンジンは、タスクリスト内のタスクに取り掛かることを開始してもよい(415)。たとえば、ウェブページのコンテンツの処理の一部として、バッチレンダリングエンジンは、1つまたは複数の埋め込みアイテムを識別してもよい(420)。次いでバッチレンダリングエンジンは、フェッチサーバから埋め込みアイテムのコンテンツを要求してもよい(425)。フェッチサーバは、図1のフェッチサーバ150であってもよい。いくつかの実施形態において、バッチレンダリングエンジンは、識別された埋め込みアイテムおよびフェッチサーバがそれぞれの埋め込みアイテムのためのコンテンツを返したか否かを追跡してもよい。いくつかの実施形態において、この埋め込みアイテムのリストは、ウェブページのレンダリング結果に含まれてもよい。バッチレンダリングエンジンが埋め込みアイテムを要求したあと、バッチレンダリングエンジンは、フェッチサーバがコンテンツを返すのを待つ間に実行する準備ができている別のタスク(415)上で作業を継続してもよい。現在の仮想時間に実行する準備ができていない場合、さもなければバッチレンダリングエンジンがフェッチサーバからの応答を待ってもよい。フェッチが未処理であるが、バッチレンダリングエンジンは仮想時計を進めないで、したがってバッチレンダリングエンジンはフェッチを待つのでタイムアウトしない。

【0045】

フェッチサーバからの応答を受信したとき、バッチレンダリングエンジンは、埋め込みアイテムのコンテンツを処理する(430)。たとえば、コンテンツの受信に応答して、バッチレンダリングエンジンは、埋め込みアイテムの受信したコンテンツを解析するなどのタスクをタスクリストに追加してもよい。これらのタスクは、現在の仮想時計の開始時間を与られてもよく、これはタスクが実行準備ができていないことを示す(たとえば、仮想時計上の現在の時間)。受信したコンテンツを解析し、最初に要求されたウェブページのためあるいは埋め込みアイテムのためであるかにかかわらず、バッチレンダリングプロセスが追加のタスクをタスクリストに追加してもよい。たとえば、埋め込みアイテムのコ

10

20

30

40

50

コンテンツを解析することは、追加の埋め込みアイテム（たとえば、セカンダリ埋め込みアイテム）を検出してもよく、これはバッチレンダリングエンジンに埋め込みアイテムを要求させ、および返されたときにコンテンツを解析させてもよい。コンテンツがたとえばJavaScriptなどのスクリプトを含む場合、スクリプトを実行することは、追加のタスクにレイアウトの生成やウェブページの1つまたは複数の要素の概観の変更などの追加のタスクを実行させてもよい。これらのタスクのいくつかは、将来開始することがスケジュールされてもよい。たとえば、画像の不透明度を一定の間隔で変更すると、画像がフェードインしているかのようにユーザに表示される。不透明度の各変化はタスクであり、スクリプトはタスクリストに、いくつかのそのようなタスクを追加させ、それぞれ現在の仮想時計のランタイムに指定された量を加えたものとなる。

10

【0046】

レンダリングプロセスの一部として、バッチレンダリングエンジンは、レンダリングが完了したか否かを判断してもよい（435）。この判断は、例えば、バッチレンダリングエンジンがタスクを完了する度に、あるいは所定の時間の間隔などで行われてもよい。仮想時計が停止タスクで指定された時間に達すると、レンダリングは終了してもよい。埋め込みアイテムのフェッチが未処理である間、仮想時計が進まないの、仮想時計が停止タスクで指定された時間に達したとき、バッチレンダリングエンジンは各フェッチリクエストの要求を受信することが保証される。したがって、バッチレンダリングエンジンは、リソースの待機をタイムアウトすることはない。

20

【0047】

レンダリングが完了していない場合（435、いいえ）、バッチレンダリングエンジンは、タスクリスト内のタスクに取り掛かることを継続し、1つまたは複数の埋め込みアイテムの要求に対する応答などを待ってもよい。レンダリング暗愚が完了した場合（435、はい）バッチレンダリングエンジンは、要求されたウェブページのためのレンダリング結果をファイナライズし（440）、レンダリング結果を要求プロセスに返してもよい。レンダリング結果の要素は、バッチレンダリングエンジンによって完了されたタスクの結果として以前に生成されたものでもよい。たとえば、スクリプトの実行中にフェッチされた埋め込みアイテムのリストおよび遭遇したエラーがレンダリングが完了する前に生成されてもよい。レイアウトを決定するなどの別の要素がレンダリングが完了したあとに発生してもよい。いくつかの実施形態において、レンダリングプロセスの一部として実行されるスクリプトが要素の位置を要求しない限り、バッチレンダリングエンジンは、レンダリングが終了するまでレイアウトを決定しない。レンダリングが完了する前にレイアウトが生成されたとしても、バッチレンダリングエンジンは、レンダリング結果のファイナライズの一部としてレイアウトを最終的に生成してもよい。したがって、レンダリング結果をファイナライズすることは、新しい要素を生成することおよびすでに生成された要素を集めることを含んでもよい。いくつかの実施形態において、バッチレンダリングエンジンは、レンダリング結果をメモリに格納し、レンダリング結果の位置を要求プロセスに提供してもよい。いくつかの実施形態において、システムは、いつ生成されたかを示すタイムスタンプとともにレンダリング結果を格納してもよく、レンダリング結果の2つ以上のバージョンを格納してもよい。バッチ処理の最適化とともにバッチモードでレンダリング結果を生成し、次いでプロセス400は終了する。

30

40

【0048】

図5は、一実施形態に従った、バッチレンダリングエンジンが仮想時計を進めるプロセス例500を示す流れ図である。プロセス500は、レンダリングが終了したか否かを判断する一部として実行されてもよいが（例えば、図4のステップ435）、他の時間に（例えば、定期的に）実行されてもよい。プロセス500は、バッチレンダリングエンジンが埋め込みアイテムの要求を待っているか否かを判断することから開始してもよい（505）。例えば、バッチレンダリングエンジンがフェッチサーバから埋め込みアイテムを要求し、フェッチサーバからまだ応答を受信していない場合、バッチレンダリングエンジンは待機中である。バッチレンダリングエンジンが待機中である場合（505、はい）、

50

仮想時計は進められず、バッチレンダリングエンジンは、存在するならば現在の仮想時間で実行する準備ができていないタスクに取り組んでもよく、あるいは待機してもよい(510)。このステップは、図4のステップ415の一部として実行されてもよい。バッチレンダリングエンジンがフェッチ要求を待っていない場合(505、いいえ)、バッチレンダリングエンジンは、タスクリスト内に実行準備ができていないタスクが存在するか否かを判断してもよい(515)。たとえば、タスクリスト内のタスクが仮想時計と等しいランタイムを有する場合、タスクは実行準備ができていない。タスクが実行準備ができていない場合(515、はい)、バッチレンダリングエンジンはタスクに取り組んでもよい(520)。タスクに取り組むことは、タスクリストに別のタスクを追加してもよく、そのうちのいくつかは実行準備ができていてもよく、そのうちの他のものは将来のランタイムを有していてもよい(例えば、現在の仮想時間プラス所定の時間)。このステップは、図4のステップ415の一部として実行されてもよい。実行準備が整っている保留中のタスクが存在しない場合(515、いいえ)、バッチレンダリングエンジンは仮想時計をタスクリストで指定された次のランタイムに進めてもよい。言い換えると、バッチレンダリングエンジンは、仮想時計を先に歪めてもよく、そうすることによってタスクリスト内のラインの次のタスクはすぐに実行できる。

10

20

30

【0049】

タスクリスト内のラインの次のタスクが停止タスクである場合(530、はい)、レンダリングは完了する。そうでなければ、バッチレンダリングエンジンは、保留中のタスクに取り組むことを継続してもよい(520)。プロセス500は、実行準備ができていない保留中のタスクがある間、あるいは埋め込みアイテムのフェッチを待機している間に、仮想時計がどのように進まないかを示す。したがって、これらのイベントに対して仮想時計が“停止して”おり、レンダリングエンジンが実際のクロックを使用するときに遭遇したエラーのクラスを避けることができる。さらに、プロセス500は、仮想時計がどのように前に歪められることができるかを示し、そうすることによっていくつかの例において、レンダリングプロセスは、タスクによって指示されたタイミング(例えば、画像のフェードインするまたはアニメーションの再生をする時間間隔を待機する)よりも実時間をかけないことができる。これは、本明細書でより詳細に説明するように、埋め込みアイテムがクロールなしで返されるときに特に当てはまる。もちろん、保留中のタスクのチェック(515)およびフェッチ要求(505)の順序を逆にする 것도でき、よび実施形態は、図5に示された順序に限定されないおkとが理解される。

【0050】

図6は、実施形態に従った、フェッチサーバが埋め込みアイテムのコンテンツをバッチレンダリングエンジンに提供するプロセス例を図示する流れ図である。プロセス600は、図1のシステム100のようなシステムによって実行されてもよい。システムは、複数のバッチレンダリングシステムから、埋め込みアイテムのためのフェッチリソースに対応するプロセス600を使用してもよい。いくつかの実施形態において、プロセス600は、フェッチサーバによって実行されてもよく、およびバッチレンダリングエンジンの1つからの要求に応答して開始されてもよい。

【0051】

プロセス600は、埋め込みアイテム(605)のためのURLをフェッチサーバが受信することから開始してもよい。URLは、バッチレンダリングエンジンによって提供されてもよく、およびバッチレンダリングエンジンによって要求された複数のURLの1つであってもよい。フェッチサーバは、要求された埋め込みアイテム(610)のURLに書き換え規則を適用してもよい。書き換え規則は、図1のURL書き換え規則154であってもよい。書き換え規則は、テンプレートおよびリダイレクトURLを含んでもよい。書き換え規則を適用することは、URLが書き換え規則の一つのためのパターンまたはテンプレートに一致するか否かを判断することを含んでもよい。例えば、テンプレートは、任意のクエリ文字列が削除されたURLであってもよく、およびシステムはテンプレートに一致するかを調べるために要求された埋め込みアイテムのURLからクエリ文字列を削

40

50

除してもよい。別の例として、テンプレートは、任意の文字または文字列がワイルドカード文字と一致することができる場所を示す、*および?のようなワイルドカード文字を含んでもよい。

【0052】

URLがパターンと一致した場合、書き換え規則は、リダイレクトURLを提供してもよく、およびフェッチサーバは、リダイレクトURLを備えた要求された埋め込みアイテムのURLを代用してもよい。書き換え規則を適用するための理由の一つは、フェッチサーバが、URLを識別し、および同じコンテンツを返すURLを識別し、不要なフェッチをスケジュールする必要がないようにするためにリダイレクトURLを使用できるようにするためである。一般的に埋め込まれたアイテムの特定のタイプは、動的に生成されたURLを有する。例えば、いくつかの埋め込みアイテムのURLは、乱数発生器によって生成される乱数、または日付と時刻関数によって返される現在の日付と時間に依存する。キャッシュバusting (cache-busting) トラッキングURLと呼ばれるこれらの埋め込みオブジェクトは、一般的に、広告費用または収益を測定する目的で、ウェブページの固有のヒット数またはビュー数を測定するために使用される。このような埋め込みオブジェクトのコンテンツは、通常は同一であるが、固有のURLは、レンダリングエンジンによって検出される度にそのオブジェクトに対して生成される。したがって、そのような埋め込みアイテムを含むウェブページのために、レンダリングエンジンは、ウェブページのレンダリングを試みる度にオブジェクトのための新しく異なるURLを見て、書き換え規則を適用することなく、フェッチサーバは同じコンテンツを何度もフェッチする。これを避けるために、書き換え規則は、フェッチサーバにこれらのURLを識別させ、フェッチ要求をリダイレクトURLの下に格納された以前に取得されたコンテンツにリダイレクトするためのテンプレートを適用してもよい。

10

20

【0053】

書き換え規則を適用するもう一つの理由は、ブラックリストに載っているURLを識別するためである。書き換え規則は、ブラックリストに載っているURLを識別する規則、あるいはブラックリストに載っているURLのためのパターンあるいはテンプレートを含んでもよい。例えば、書き換え規則は、テンプレート、および関連したリダイレクトURL、エラー、あるいはフラグを含んでもよい。要求された埋め込みアイテムのためのURLがブラックリストに載っているURLあるいはブラックリストに載っているURLのためのテンプレートと一致した場合、フェッチサーバは、URLをブラックリストに載っているものとして識別してもよい。いくつかの実施形態において、書き換え規則を適用することは、URLがリダイレクトURLに置き換えられる要因となってもよい。いくつかの実施形態において、書き換え規則を適用することは、URLがブラックリストに載っているものとしてフラグを立ててもよく、あるいはURLによって識別された埋め込みアイテムのための要求に対する応答として返すためにエラーを提供してもよい。

30

【0054】

URLがブラックリストに載っている場合(615、はい)、フェッチサーバは要求しているバッチレンダリングエンジン(620)にエラーを返してもよい。エラーは、リソースが見つからなかったことを示す標準的なエラー、あるいはリソースが必要ではない、またはスキップができることなどをレンダリングエンジンに知らせる特定のエラーであってもよい。エラーは、書き換え規則がリダイレクトURLを提供したとき、埋め込みアイテムテーブルから、書き換え規則、ハードコードなどのフラグに基づいて選択された一致させている書き換え規則によって、埋め込みアイテムテーブルから提供されてもよい。次いでこのURLのためのフェッチ要求が完了し、プロセス600は終了する。

40

【0055】

URLがブラックリストに載っていない場合(615、いいえ)、フェッチサーバは、埋め込みアイテムデータストア(625)内の書き換えられたURLを探してもよい。埋め込みアイテムデータストアは、図1の埋め込みアイテムテーブル152であってもよい。オリジナルURLが書き換え規則の識別されたパターンと一致した場合、書き換えられたURLは

50

、書き換え規則によって提供されたリダイレクトURLであってもよい。URLが書き換え規則の任意のテンプレートと一致しない場合、書き換えられたURLは、オリジナルURLであってもよい。URLが埋め込みアイテムデータストア（625、はい）内にある場合、フェッチサーバは、要求されたURLが画像用であるか否かを任意に決定してもよい（630）。これは任意であり、画像のためのテストをしない実施形態において、ステップ630は省略されてもよい。要求された埋め込みアイテムが画像であるか否かは、要求の情報、URL自体に基づいて、あるいは書き換えられたURLのための埋め込みアイテムデータストア内のフィールドに基づいて、決定されてもよい。埋め込みアイテムが画像である場合（630、はい）、図7のプロセス700に関してより詳細に説明されるように、システムは、画像のサイズについてサイズテーブル内を調べ、およびサイズを有するモック画像を返してもよい。いくつかの実施形態において、フェッチサーバは、埋め込みアイテムデータストアを前に、またはエントリが古いかどうかを判定した後に、書き換え規則を適用する前にステップ630を実行してもよいことが理解される。

10

20

30

40

50

【0056】

要求された埋め込みアイテムが画像でない場合（630、いいえ）、フェッチサーバは、埋め込みアイテムテーブル内のエントリが古いか否かを判断してもよい（645）。エントリが古いか否かは、アイテムの変更率、埋め込みアイテムのタイプ（例えば、スクリプト、スタイルシート、画像など）、ブラウザレンダリングエンジンがレンダリングしているウェブページの重要度、などのいくつかの要因に依存してもよい。いくつかの実施形態において、埋め込みアイテムテーブルは、例えばブラックリストに載っている埋め込みアイテムのURLのリダイレクトURLのためのように、エントリが古くならないことを示すフィールドあるいは値を有してもよい。エントリが古くない場合（645、いいえ）、フェッチサーバは、書き換えられたURLについての埋め込みアイテムテーブルのコンテンツを、要求しているバッチレンダリングエンジン（650）に返してもよく、この埋め込みアイテムについてのプロセス600は終了する。いくつかの実施形態において、コンテンツを返すことは、フェッチサーバが埋め込みアイテムテーブル内のエントリの位置を応答として提供することと、バッチレンダリングプロセスが位置を使用してコンテンツにアクセスすることと、を含んでもよい。

【0057】

埋め込みアイテムテーブル内のエントリが古い場合（645、はい）または書き換えられたURLが埋め込みアイテムデータストア内に無い場合（625、いいえ）、フェッチサーバは、例えば図1のウェブ巡回エンジン130のようなウェブ巡回からURLのフェッチを要求してもよい（635）。フェッチサーバがクロールコンテンツを受信するとき、通信あるいはさらなる処理をせずに、受信したコンテンツを埋め込みアイテムデータストアにエントリとして格納してもよい（640）。いくつかの実施形態において、フェッチサーバは、埋め込みアイテムの以前のクロールのコンテンツおよびクロール時間に上書きせずに、埋め込みアイテムのコンテンツおよびクロール時間を保存することができる。いくつかの実施形態において、フェッチサーバは、埋め込みアイテムテーブルに1つのエントリをキープしてもよく、および埋め込みアイテムの以前のクロールを維持しなくてもよい。いずれにせよ、埋め込みアイテムテーブル内に一旦保存されると、コンテンツはキャッシュされ、および古くなるまで再度フェッチされる必要はない。フェッチサーバは、フェッチされたコンテンツを要求しているバッチレンダリングエンジンに返してもよく（650）、プロセス600は終了する。

【0058】

図7は、一実施形態に従った、フェッチサーバがモック画像をバッチレンダリングエンジンに提供するプロセス例700を図示する流れ図である。プロセス700は、図1のシステム100のようなシステムによって実行されてもよい。システムは、複数のバッチレンダリングエンジンからの、ウェブページに埋め込まれた画像のためのフェッチ要求に回答するために、プロセス700を使用してもよい。いくつかの実施形態において、プロセス700は、フェッチサーバによって実行されてもよく、およびバッチレンダリングエン

ジンの１つからの要求に応答して開始されてもよい。いくつかの実施形態において、フェッチサーバは、他の埋め込みアイテム（例えば、図６のプロセス６００）とは無関係にプロセス７００を実行してもよい。別の実施形態において、フェッチサーバは、プロセス７００の要素を他の埋め込みアイテムを含むプロセス、例えば図６のプロセス６００に組み込んでよい。

【００５９】

プロセス７００は、要求された画像がサイズテーブル（７０５）にエントリを有するか否かをフェッチサーバが判断することから開始してもよい。画像サイズテーブルは、図１の画像サイズテーブル１５６であってもよい。画像サイズテーブルは、画像のためのサイズを含み、これはＵＲＬのような画像のための識別子によって格納される。画像がサイズテーブルにない場合（７０５、いいえ）、あるいは画像がサイズテーブルにあるが（７０５、はい）古い（７１０、はい）場合、フェッチサーバは、例えば図１のウェブ巡回エンジン１３０のようなウェブ巡回エンジンを介して、画像のフェッチ（７１５）をスケジュールしてもよい。いくつかの実施形態において、フェッチサーバは、エントリが古いかなを判断するためにサイズテーブルの情報を使用してもよい。いくつかの実施形態において、フェッチサーバは、サイズが古いかなを判断するために、図６のステップ６４５に関して上述したように、別個の埋め込みアイテムテーブルからの情報を使用してもよい。したがって、いくつかの実施形態において、フェッチサーバは図６のステップ６４５と併せて、あるいはその一部としてステップ７１０を実行してもよい。画像のためのコンテンツが受信されたとき、フェッチサーバは、サイズテーブルに画像のためのエントリを追加してもよく、エントリはフェッチされた画像のサイズを含む（７２０）。いくつかの実施形態において、フェッチサーバは、図６のステップ６４０の一部として上述されたように、フェッチされたコンテンツを埋め込みアイテムテーブルに格納してもよい。

【００６０】

画像がサイズテーブルにあり（７０５、はい）、かつ古くない（７１０、いいえ）場合、あるいは画像がフェッチされて格納された（７２０）後、システムはサイズテーブル（７２５）からサイズを使用してモック画像を生成してもよい。モック画像は、要求された画像と同じサイズであるが空のコンテンツを指定する画像ファイルフォーマットデータを有してもよい。システムは、モック画像（７３０）を要求バッチレンダリングエンジンに返し、プロセス７００は終了する。

【００６１】

いくつかの実施形態において、プロセス７００のステップのいくつかは、任意であるかまたは他の処理の一部として実行されてもよいことが理解される。例えば、画像のサイズが古いかなを判断するステップは、図６のステップ６４５の一部として実行されてもよく、埋め込みアイテムテーブルの情報に基づいてもよい。さらに、ステップ７１５は、図６のステップ６３５の一部として、あるいはそれと併せて実行されてもよい。言い換えると、フェッチサーバは、画像のためのコンテンツをフェッチする、キャッシュされフェッチされたコンテンツが古いかなを判断する、などのようなプロセス６００の態様とプロセス７００の態様を組み合わせてもよい。もちろん、フェッチサーバは、プロセス６００と完全に独立したプロセス７００を実行してもよい。したがって、実装はプロセス７００の変形を含んでもよい。

【００６２】

図６は、汎用コンピュータデバイス８００の例を示しており、それは、図１のシステム１００、及び／又はクライアント１７０として動作し得、また、本明細書に記載される技法を用いて使用され得る。コンピューティングデバイス８００は、コンピューティングデバイスの種々の形態例、例えば、ラップトップ、デスクトップ、ワークステーション、パーソナルデジタルアシスタント、携帯電話、スマートフォン、タブレット、サーバ、ならびにウェアラブルデバイスを含む他のコンピューティングデバイスなどを表わすことが意図される。本明細書に示される構成要素、それらのつながりや関係、およびそれらの機能は、単なる例であることを意味するものであり、この文書において記載される及び／又は

10

20

30

40

50

請求される発明の実施形態を限定することを意味するものではない。

【0063】

コンピューティングデバイス800は、インターフェース808経由で接続される、例えばシリコンベースのハードウェアプロセッサ等のプロセッサ802、メモリ804、ストレージデバイス806、および拡張ポート810を含む。いくつかの実施形態において、コンピューティングデバイス800は、インターフェース808経由で接続される、いくつかある構成要素の中でも、送受信機846、通信インターフェース844、およびGPS（全地球測位システム）レシーバモジュール848を含み得る。デバイス800は、通信インターフェース844を通して無線で通信し得、それは、必要に応じて、デジタル信号処理回路を含み得る。構成要素802、804、806、808、810、840、844、846、および848のそれぞれは、必要に応じて、一般的なマザーボード上にまたは他の様式で搭載され得る。

10

【0064】

プロセッサ802は、メモリ804内またはストレージデバイス806上に記憶された命令を含む、コンピューティングデバイス800内の実行のための命令を処理することができ、外部入出力デバイス、例えばディスプレイ816などの上にGUIのためのグラフィカル情報を表示する。ディスプレイ816は、モニタまたはフラットタッチスクリーンディスプレイであり得る。いくつかの実施形態において、複数のプロセッサ及び/又は複数のバスが、必要に応じて、複数のメモリや複数の種類のメモリと共に、使用され得る。また、複数のコンピューティングデバイス800は、（例えば、サーババンク、ブレードサーバのグループ、またはマルチプロセッサシステムとして）必要な動作の一部分を提供する各デバイスと、接続され得る。

20

【0065】

メモリ804は、コンピューティングデバイス800内に情報を記憶する。一実施形態において、メモリ804は、揮発性メモリのユニットまたは複数ユニットである。別の実施形態において、メモリ804は、不揮発性メモリのユニットまたは複数ユニットである。メモリ804はまた、コンピュータで読み取り可能な媒体の別の形態、例えば磁気もしくは光学ディスクなどであり得る。いくつかの実施形態において、メモリ804は、拡張インターフェースを通して提供される拡張メモリを含み得る。

30

【0066】

ストレージデバイス806は、コンピューティングデバイス800のためのマストレージを提供することができる。一実施形態において、ストレージデバイス806は、コンピュータで読み取り可能な媒体、例えば、フロッピーディスクデバイス、ハードディスクデバイス、光学ディスクデバイス、またはテープデバイス、フラッシュメモリもしくは他の類似の固体メモリデバイス、あるいはストレージエリアネットワークもしくは他の構成におけるデバイスを含む、デバイスのアレイなどであり得るか、それらを含み得る。コンピュータプログラム製品は、そのようなコンピュータで読み取り可能な媒体に明白に具体化され得る。コンピュータプログラム製品はまた、実行されるときに、例えば上記したものなどの、1つ以上の方法を行う命令を含み得る。コンピュータまたはマシンで読み取り可能な媒体は、ストレージデバイス、例えばメモリ804、ストレージデバイス806、またはプロセッサ802上のメモリなどである。

40

【0067】

インターフェース808は、コンピューティングデバイス800のための帯域消費型の動作を扱う高速コントローラ、またはより低い帯域消費型の動作を扱う低速コントローラ、あるいはそのようなコントローラの組み合わせであり得る。外部インターフェース840は、他のデバイスとのデバイス800の近接領域通信を可能にするように、提供され得る。いくつかの実施形態において、コントローラ808は、ストレージデバイス806および拡張ポート814に結合され得る。種々の通信ポート（例えば、USB、ブルートゥース（登録商標）、イーサネット（登録商標）、無線イーサネット（登録商標））を含み得る拡張ポートは、例えば、ネットワークアダプタを通して、1つ以上の入出力デバイス

50

、例えば、キーボード、ポインティングデバイス、スキャナ、または例えばスイッチもしくはルータなどのネットワーキングデバイスに結合され得る。

【0068】

コンピューティングデバイス800は、図面に示されるように、多くの異なる形態で実装され得る。例えば、それは、標準サーバ830として、またはそのようなサーバのグループにおいて複数回、実装され得る。それはまた、ラックサーバシステムの一部として実装され得る。加えて、それは、パーソナルコンピュータ、例えばラップトップコンピュータ832、デスクトップコンピュータ834、あるいはスマートフォン836などで実装され得る。システム全体は、互いに通信する複数のコンピューティングデバイス800で構成され得る。他の構成が可能である。

10

【0069】

図9は、汎用コンピュータデバイス900の例を示しており、それは、図1のシステム100であり得るとともに本明細書に記載される技法を用いて使用され得る。コンピューティングデバイス900は、大規模データ処理デバイス、例えばサーバ、ブレードサーバ、データセンター、メインフレーム、および他の大規模コンピューティングデバイスなどの種々の形態例を表わすことが意図される。コンピューティングデバイス900は、場合によっては1つ以上の通信ネットワークによって相互接続されるネットワーク接続ストレージノードを含む、複数プロセッサを有する分散型システムであり得る。本明細書に示された構成要素、それらのつながりや関係、およびそれらの機能は、単なる例示であることを意味するものであり、この文書において記載されるか請求される発明の実施形態を限定

20

【0070】

分散型コンピューティングシステム900は、任意の数のコンピューティングデバイス980を含み得る。コンピューティングデバイス980は、ローカルもしくはワイドエリアネットワーク、専用光学リンク、モデム、ブリッジ、ルータ、スイッチ、有線もしくは無線ネットワーク等の上で通信する、サーバまたはラックサーバ、メインフレーム等を含み得る。

【0071】

いくつかの実施形態において、各コンピューティングデバイスは、複数のラックを含み得る。例えば、コンピューティングデバイス980aは、複数のラック958a~958nを含む。各ラックは、1以上のプロセッサ、例えばプロセッサ952a~952nおよび962a~962nなどを含み得る。プロセッサは、データプロセッサ、ネットワーク接続ストレージデバイス、および他のコンピュータ制御デバイスを含み得る。いくつかの実施形態において、1つのプロセッサは、マスタープロセッサとして動作し得、スケジューリングおよびデータ分散タスクを管理し得る。プロセッサは、1以上のラックスイッチ958を通して相互接続され得、1以上のラックが、スイッチ978を通して接続され得る。スイッチ978は、複数の接続されたコンピューティングデバイス900間の通信に対処し得る。

30

【0072】

各ラックは、メモリ、例えばメモリ954やメモリ964など、およびストレージ、例えば956および966などを含み得る。ストレージ956および966は、マストレージを提供し得、揮発性もしくは不揮発性ストレージ、例えばネットワーク接続ディスク、フロッピーディスク、ハードディスク、光学ディスク、テープ、フラッシュメモリもしくは他の類似の固体メモリデバイス、あるいはストレージエリアネットワークまたは他の構成におけるデバイスを含む、デバイスのアレイなどを含み得る。ストレージ956または966は、複数のプロセッサ、複数のラック、または複数のコンピューティングデバイス間で共有され得、1つ以上のプロセッサによって実行可能な命令を記憶するコンピュータで読み取り可能な媒体を含み得る。メモリ954および964は、例えば、揮発性メモリのユニットまたは複数ユニット、不揮発性メモリのユニットもしくは複数ユニット、及び/又はコンピュータで読み取り可能な媒体の他の形態、例えば、磁気もしくは光学ディ

40

50

スク、フラッシュメモリ、キャッシュ、ランダムアクセスメモリ（ＲＡＭ）、読み取り専用メモリ（ＲＯＭ）、およびそれらの組み合わせなどを含み得る。メモリ、例えばメモリ 954などはまた、プロセッサ952a～952n間で共有され得る。索引などのデータ構造は、例えば、ストレージ956やメモリ954にわたって、記憶され得る。コンピュータデバイス700は、図示されない他の構成要素、例えば、コントローラ、バス、入出力デバイス、通信モジュール等を含み得る。

【0073】

システム100などのシステム全体は、互いと通信する複数のコンピュータデバイス900で構成され得る。例えば、デバイス980aは、デバイス980b、980c、および980dと通信し得、これらは、集合的にシステム100として知られ得る。別の例として、図1のシステム100は、1以上のコンピュータデバイス900を含み得る。コンピュータデバイスのうちのいくつかは、互いに地理的に近く位置し得、その他は、地理的に離れて位置し得る。コンピュータデバイス900のレイアウトは、単なる例であり、システムは、他のレイアウトまたは構成を取り得る。

10

【0074】

さまざまな実施形態は、ストレージシステム、少なくとも1つの入力デバイス、および少なくとも1つの出力デバイスからデータや命令を受信するように、並びにそれらにデータや命令を送信するように、結合される、特殊目的もしくは一般の目的であり得る、回路基板に形成された少なくとも1つのプログラム可能なプロセッサを含むプログラム可能なシステム上で実行可能または解釈可能である1つ以上のコンピュータプログラムにおける実施形態を含むことができる。

20

【0075】

これらのコンピュータプログラム（プログラム、ソフトウェア、ソフトウェアアプリケーション、またはコードとしても知られる）は、プログラム可能なプロセッサに対する機械語命令を含み、高レベル手続言語及び／又はオブジェクト指向プログラミング言語で、ならびに／あるいはアセンブリ／機械言語で実装することができる。本明細書で使用される場合、「機械可読媒体」「コンピュータ可読媒体」という用語は、機械語命令及び／又はデータをプログラム可能なプロセッサに提供するために使用される任意の非一時的コンピュータプログラム製品、装置、及び／又はデバイス（例えば、磁気ディスク、光ディスク、メモリ（読み取り専用メモリを含む）、プログラマブル論理デバイス（PLD））を指す。

30

【0076】

本明細書で述べたシステムおよび技術は、バックエンドコンポーネント（例えば、データサーバとして）を含むコンピューティングシステム、またはミドルウェアコンポーネント（例えば、アプリケーションサーバ）を含むコンピューティングシステム、またはフロントエンドコンポーネント（例えば、ユーザが本明細書で述べたシステムおよび技術の実装と相互作用することができるグラフィカルユーザインタフェースまたはウェブブラウザを備えたクライアントコンピュータ）を含むコンピューティングシステム、またはこのようなバックエンド、ミドルウェア、またはフロントエンドコンポーネントの任意の組合せで、実装されることができる。前記システムのコンポーネントは、デジタルデータ通信（例えば通信ネットワーク）の任意の形式または媒体で接続されることができる。通信ネットワークの例としては、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”）、およびインターネットを含む。

40

【0077】

コンピューティングシステムは、クライアントおよびサーバを含むことができる。クライアントおよびサーバは、通常、相互に離れており、および通常は、通信ネットワークを通じて相互作用する。クライアントとサーバの関係は、それぞれのコンピュータ上で動作し、互いにクライアント・サーバ関係を有するコンピュータプログラムの効力によって生じる。

【0078】

50

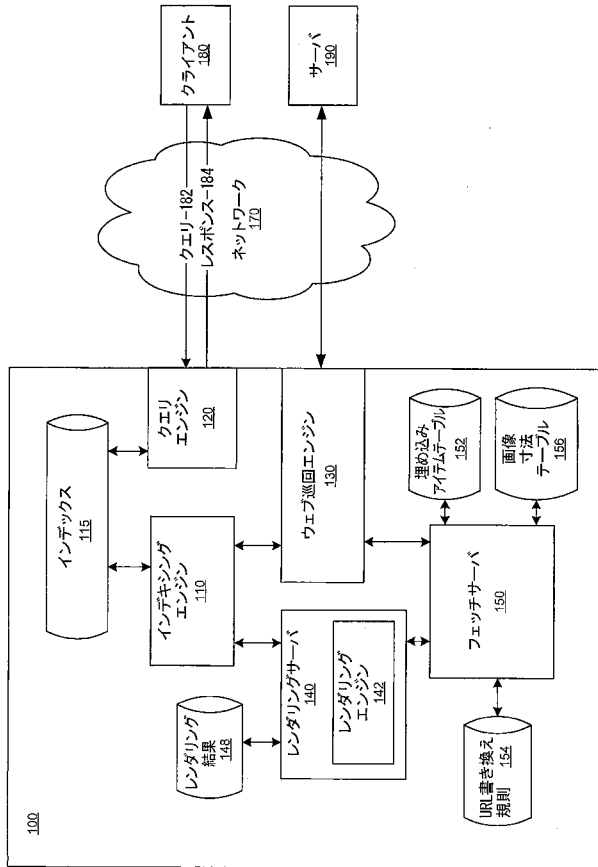
多くの実施形態が説明されてきた。それにもかかわらず、本発明の精神および範囲から逸脱することなく、さまざまな変更がなされ得る。さらに、図に示された論理フローは、望ましい結果を達成するために、図示された特定の順序または連続する順序である必要はない。さらに、他のステップを設けても、または記載されたフローから複数のステップを除いてもよく、また、記載されたシステムに他の要素を追加してもよく、またはそこから除去してもよい。従って、他の実施形態は、添付の特許請求の範囲に含まれる。

【符号の説明】

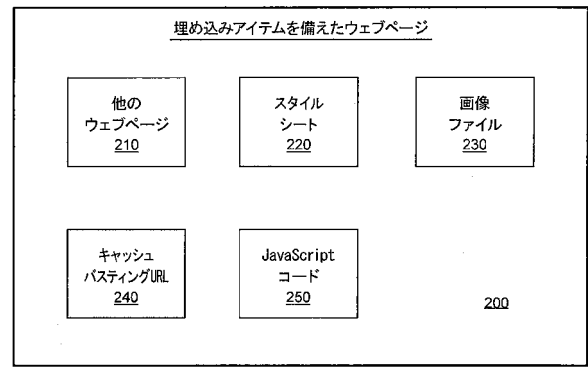
【 0 0 7 9 】

1 1 0	インデキシングエンジン	
1 1 5	インデックス	10
1 2 0	クエリエンジン	
1 3 0	ウェブ巡回エンジン	
1 4 0	レンダリングサーバ	
1 4 2	レンダリングエンジン	
1 4 8	レンダリング結果	
1 5 0	フェッチサーバ	
1 5 2	埋め込みアイテムテーブル	
1 5 4	URL書き換え規則	
1 5 6	画像寸法テーブル	
1 7 0	ネットワーク	20
1 8 0	クライアント	
1 8 2	クエリ	
1 8 4	レスポンス	
1 9 0	サーバ	
2 2 0	スタイルシート	
2 3 0	画像ファイル	
2 4 0	キャッシュバスティングURL	
3 0 5	ページタスクリスト	
3 1 0	バーチャル時計	
3 1 5	レンダリング結果	30
3 1 6	画像	
3 1 7	DOMツリー	
3 1 8	レイアウト	
3 1 9	埋め込みリソース	
3 2 0	エラー	

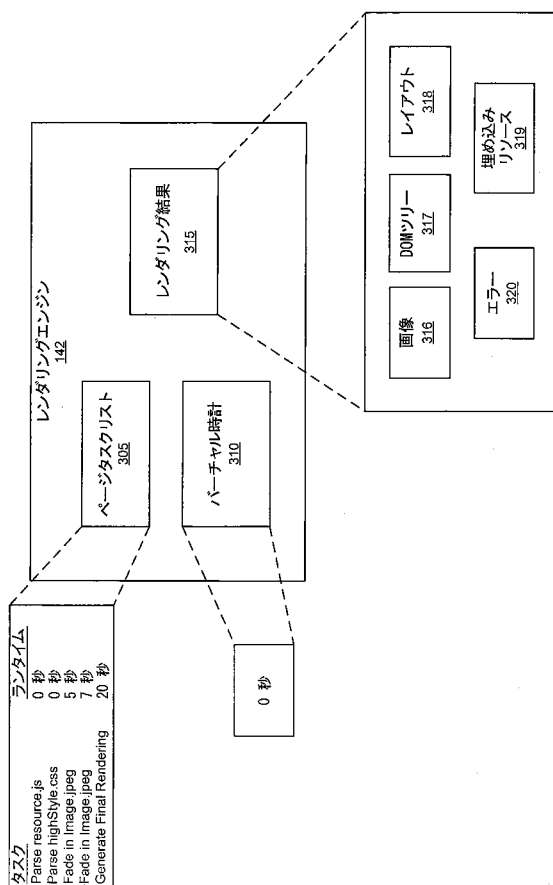
【図 1】



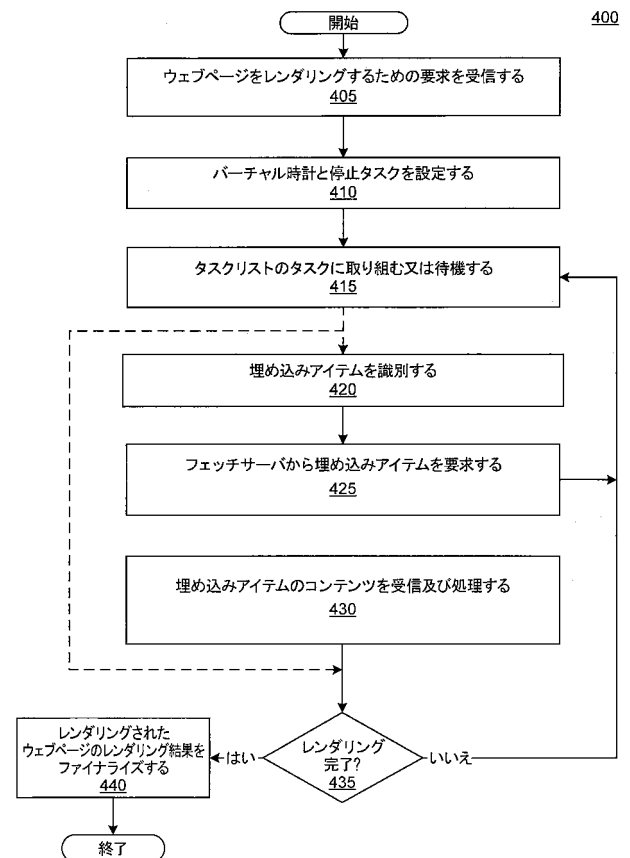
【図 2】



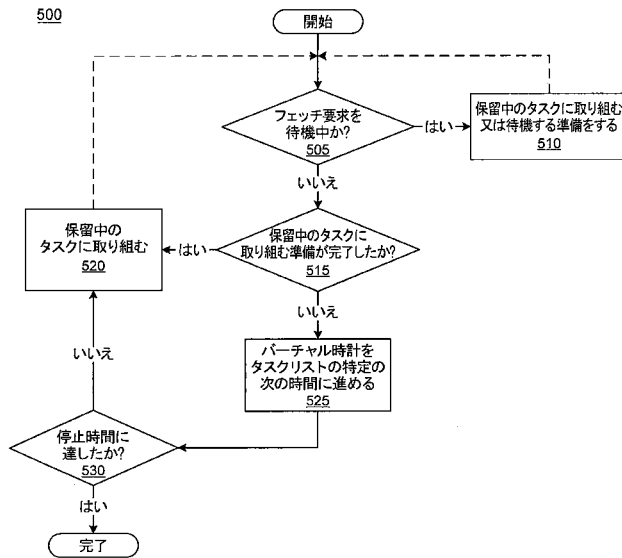
【図 3】



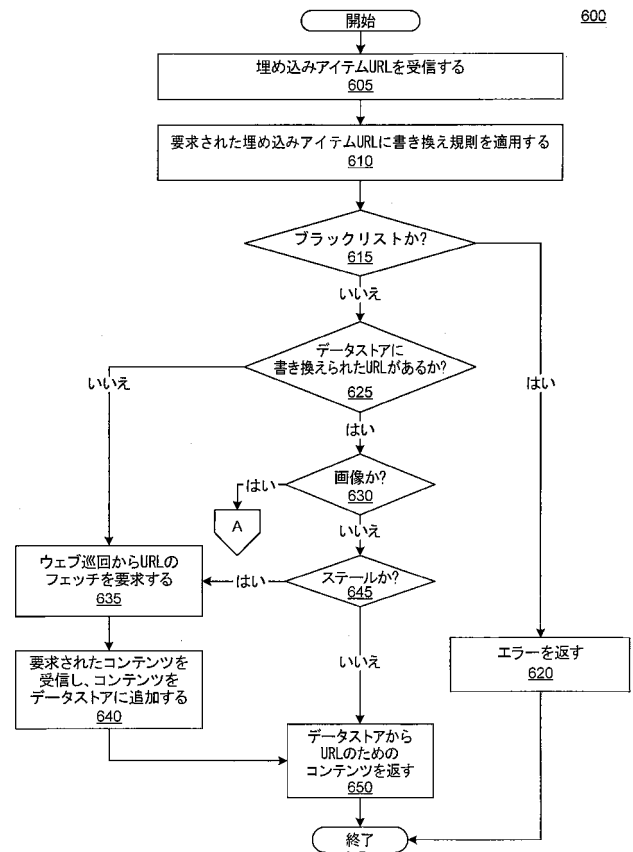
【図 4】



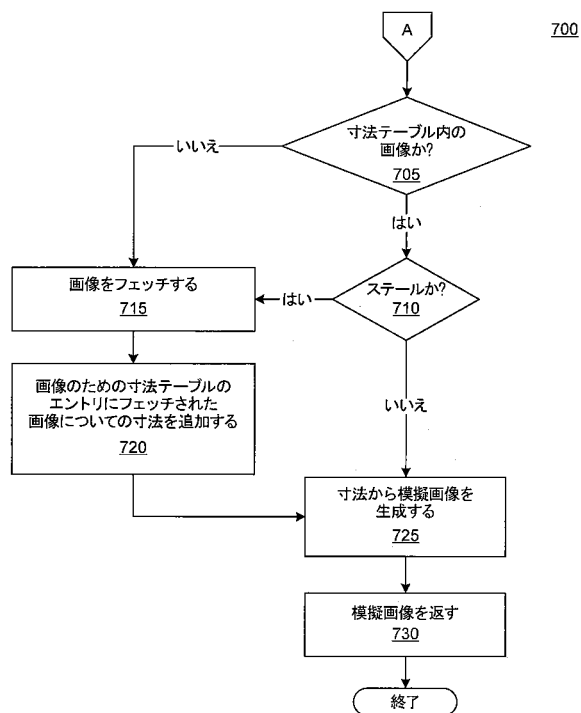
【図 5】



【図 6】



【図 7】



【図 8】

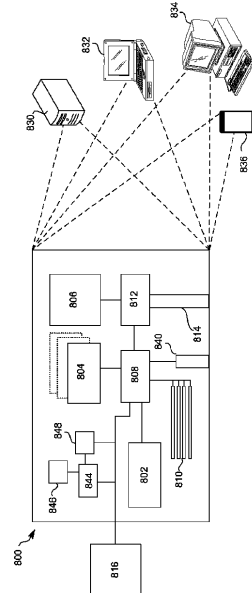


FIG. 8

【図 9】

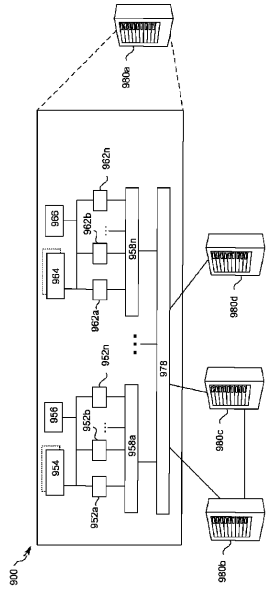


FIG. 9

【 国際調査報告 】

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2014/080832

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, WPI, EPDOC: item, embed+, rewrite, rule, store, duplicat+, result, fetch

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 101201827 A (ALIBABA CO.) 18 June 2008 (2008-06-18) description, page 2, line 2 - page 5, line 8, and figures 1-2	1-4, 15-19
A	CN 101075236 A (TENGXUN SCIENCE & TECHNOLOGY SHENZHEN CO., LTD.) 21 November 2007 (2007-11-21) the whole document	1-19
A	CN 101587495 A (WU, DIZHOU) 25 November 2009 (2009-11-25) the whole document	1-19
A	CN 101763357 A (BEIJING SOGOU TECHNOLOGY DEVELOPMENT CO., LTD.) 30 June 2010 (2010-06-30) the whole document	1-19
A	US 2006106844 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 18 May 2006 (2006-05-18) the whole document	1-19
A	EP 2728481 A1 (RIGHTWARE OY) 07 May 2014 (2014-05-07) the whole document	1-19

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

- “A” document defining the general state of the art which is not considered to be of particular relevance
- “E” earlier application or patent but published on or after the international filing date
- “L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- “O” document referring to an oral disclosure, use, exhibition or other means
- “P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

12 March 2015

Date of mailing of the international search report

25 March 2015

Name and mailing address of the ISA/CN

STATE INTELLECTUAL PROPERTY OFFICE OF THE
P.R.CHINA(ISA/CN)
6,Xitucheng Rd., Jimen Bridge, Haidian District, Beijing
100088, China

Authorized officer

ZHANG, Qianzhen

Facsimile No. (86-10)62019451

Telephone No. (86-10)82245953

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2014/080832

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	101201827	A	18 June 2008	EP	2116939	A1	11 November 2009
				JP	5230642	B2	10 July 2013
				JP	2010512589	A	22 April 2010
				JP	5536846	B2	02 July 2014
				HK	1117612	A1	19 July 2013
				EP	2116939	A4	25 April 2012
				CN	101201827	B	20 February 2013
				WO	2008071097	A1	19 June 2008
				US	2010088612	A1	08 April 2010
				JP	2013016209	A	24 January 2013
CN	101075236	A	21 November 2007	Non			c
CN	101587495	A	25 November 2009	Non			e
CN	101763357	A	30 June 2010	Non			e
US	2006106844	A1	18 May 2006	CN	1776678	A	24 May 2006
				CN	100430939	C	05 November 2008
				US	8812551	B2	19 August 2014
				TW	200636514	A	16 October 2006
EP	2728481	A1	07 May 2014	CN	103810097	A	21 May 2014
				US	2014129697	A1	08 May 2014

フロントページの続き

(81)指定国 AP(BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, RU, TJ, TM), EP(AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US

(72)発明者 ハオ・ファン

中華人民共和国・100084・ベイジン・ハイディエン・ディストリクト・ジョングアンクン・イースト・ロード・ナンバー・1・ツインファ・サイエンス・パーク・ビルディング・6

(72)発明者 エリック・アルジャン・ヘンドリクス

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 ファイ・シュ

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 クリスティアン・タブス

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 ルペシュ・カプア

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

Fターム(参考) 5B084 AA01 AA26 AB04 AB06 AB16 AB30 BB12 CB12 CB24 DB09

DB12 DC02 DC17