

República Federativa do Brasil
Ministério do Desenvolvimento, Indústria
e do Comércio Exterior
Instituto Nacional da Propriedade Industrial.

(21) PI0616233-9 A2



* B R P I 0 6 1 6 2 3 3 A 2 *

(22) Data de Depósito: 20/09/2006
(43) Data da Publicação: 14/06/2011
(RPI 2110)

(51) Int.CI.:
G06F 17/30 2006.01

(54) Título: FUNÇÕES DE CLASSIFICAÇÃO USANDO ESTATÍSTICAS DE USO DE DOCUMENTO

(30) Prioridade Unionista: 21/09/2005 US 11/231,955

(73) Titular(es): MICROSOFT CORPORATION

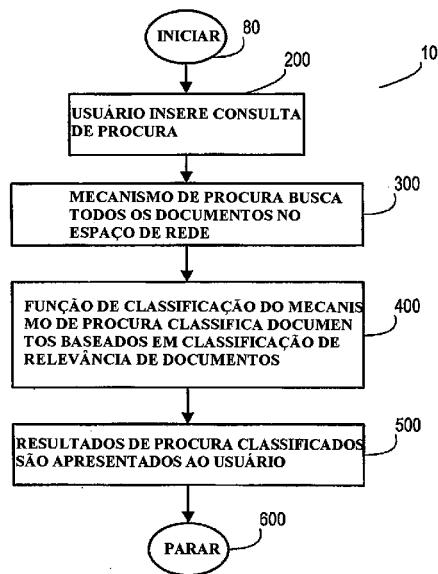
(72) Inventor(es): HUGO ZARAGOZA, ANDREW DEBRUYNE, DMITRIY MEYERZON, KYLE PELTONEN

(74) Procurador(es): NELLIE ANNE DANIEL SHORES

(86) Pedido Internacional: PCT US2006037206 de 20/09/2006

(87) Publicação Internacional: WO 2007/035919de 29/03/2007

(57) Resumo: FUNÇÕES DE CLASSIFICAÇÃO USANDO ESTATÍSTICAS DE USO DE DOCUMENTO. São discutidos aqui métodos que fornecem uma classificação de relevância de documento a um documento em uma rede. O meio legível por computador que tem armazenado neste instruções executáveis por computador para executar um método para fornecer uma classificação de relevância de documento a um documento em uma rede também é discutido. Ademais, são discutidos sistemas de computação que contêm no mínimo um módulo de aplicação, onde no mínimo um módulo aplicativo compreende o código de aplicação para executar métodos de fornecimento de uma classificação de relevância de documento a um documento em uma rede.



"FUNÇÕES DE CLASSIFICAÇÃO USANDO ESTATÍSTICAS DE
USO DE DOCUMENTO"

Fundamentos da Invenção

As funções de classificação que classificam documentos de acordo com sua relevância em uma dada consulta de procura são conhecidas. Esforços continuam na técnica para desenvolver funções de classificação que fornecem melhores resultados de procura para uma dada consulta de procura comparada aos resultados de procura gerados por mecanismos de procura que usam funções de classificação conhecidas.

Sumário da Invenção

São descritas aqui, entre outras coisas, várias tecnologias que determinam uma classificação de relevância de documento para um dado documento em uma rede. A classificação de relevância de documento é gerada via uma função de classificação que compreende um ou mais componentes independentes de consultas, onde no mínimo um componente independente de consulta inclui um parâmetro de uso que leva em conta dados de uso de documentos reais mantidos e armazenados em um servidor da Rede para um ou mais documentos na rede. As funções de classificação podem ser usadas por um mecanismo de procura para classificar múltiplos documentos em ordem (normalmente, na ordem decrescente) baseados nas classificações de relevância de documento dos múltiplos documentos.

Este sumário é fornecido para geralmente introduzir ao leitor um ou mais conceitos selecionados descritos abaixo mais detalhadamente na seção "Descrição Detalhada" de

uma forma simplificada. Este sumário não pretende identificar as características chave e/ou solicitadas do assunto reivindicado.

Breve Descrição dos Desenhos

5 A FIG. 1 representa um fluxograma lógico exemplificado que mostra etapas exemplificadas em um método de produção de resultados de procura classificados em resposta a uma consulta de procura inserida por um usuário;

10 A FIG. 2 representa um diagrama de bloco de alguns dos componentes primários de um ambiente operacional exemplificado para a implementação dos métodos e processos descritos aqui;

15 A FIG. 3 representa um fluxograma lógico exemplificado que mostra etapas exemplificadas do método de determinação de classificação de relevância de documento para documentos em uma rede; e

20 A FIG. 4 representa um fluxograma lógico exemplificado que mostra etapas exemplificadas em um método de classificação de resultados de procura gerados por uma função de classificação que contém um parâmetro de uso de documento.

Descrição Detalhada da Invenção

Para promover um entendimento dos princípios dos métodos e processos descritos aqui, as descrições de modalidades específicas seguem e linguagem específica é usada para descrever as modalidades específicas. Todavia entende-se que nenhuma limitação do escopo dos métodos e processos descritos é pretendida pelo uso de linguagem específica. Altera-

ções, modificações adicionais e tais aplicações adicionais dos princípios dos métodos descritos e processos discutidos são observadas como normalmente ocorreria a um versado na técnica ao qual os métodos e processos descritos pertencem.

5 Métodos de determinar uma classificação de relevância de documento para documentos em uma rede são descritos. Cada classificação de relevância de documento é calculada usando uma função de classificação que desejavelmente contém um ou mais componentes independentes de consulta (por 10 exemplo, um componente de função que não depende de uma dada consulta de procura ou termo de consulta de procura), ou uma combinação desses. As classificações de relevância de documento determinadas pela função de classificação podem ser usadas para classificar documentos dentro de um espaço de 15 rede (por exemplo, um espaço de intranet corporativa) de acordo com cada classificação de relevância de documento. Um processo de procura exemplificado no qual os métodos descritos podem ser usados é mostrado como o processo exemplificado 10 na FIG. 1.

20 A FIG. 1 descreve o processo de procura exemplificado 10, que inicia com a etapa de processo 80, onde um usuário insere uma consulta de procura. A partir da etapa 80, o processo de procura exemplificado 10 prossegue para a etapa 200, onde um mecanismo de procura busca todos os documentos 25 dentro de um espaço de rede por um ou mais termos da consulta de procura. A partir da etapa 200, o processo de procura exemplificado 10 prossegue para a etapa 300, onde uma função de classificação do mecanismo de procura classifica os docu-

mentos dentro do espaço de rede baseado na classificação de relevância de cada documento, a classificação de relevância de documento baseia-se em um ou mais componentes independentes de consulta, um ou mais componentes dependentes de consulta, ou a combinação desses. A partir da etapa 300, o processo de procura exemplificado 10 prossegue para a etapa 400, onde os resultados de procura classificados são apresentados ao usuário, tipicamente na ordem decrescente, identificando os documentos dentro do espaço de rede que são 10 mais relevantes à consulta de procura.

Como discutido mais detalhadamente abaixo, em alguns métodos exemplificados de determinar uma classificação de relevância de documento, no mínimo um componente independente de consulta de uma função de classificação usada para 15 determinar uma classificação de relevância de documento leva em conta "dados de uso de documento" ou "estatísticas de uso de documento" relacionados ao uso atual de um ou mais documentos dentro de um espaço de rede por um ou mais usuários. Os dados e/ou estatísticas de uso de documento são gerados e 20 armazenados por código de aplicação em um servidor da Rede, que é separado de um dado mecanismo de procura. Por exemplo, os dados de uso de documento podem ser mantidos por um sítio da Rede tal que, cada vez que um usuário solicita uma URL, o servidor atualiza um contador de uso. O contador de uso pode 25 manter os dados relacionados a documentos obtidos para um dado intervalo de tempo, tal como última semana, o último mês, o último ano, ou o tempo de vida de um dado documento ou conjunto de documentos. O código de aplicação pode ser

usado para obter os dados de uso do sítio da Rede via (i) uma interface de programação de aplicação especial (API), (ii) uma solicitação de serviço da Rede, ou (iii) solicitando uma pagina da Rede de administração que retorna dados de uso para cada URL no sítio da Rede.

Os sítios da Rede específicos podem ser usados para gerar e manter os dados de uso dentro de um espaço de rede, bem como armazenar os dados de uso em um sistema de armazenamento remoto ou local. Os sítios da Rede adequados para gerar, manter e armazenar os dados de uso de documentos dentro de um espaço de rede incluem, mas não são limitados a, sítios de Serviços WINDOWS® SHAREPOINT®.

Os métodos descritos de determinar uma classificação de relevância de documento podem adicionalmente utilizar uma função de classificação que compreende um ou mais componentes adicionais independentes de consulta. Os componentes adicionais adequados independentes de consulta incluem, mas não são limitados a, um componente independente de consulta que leva em consideração uma distância de clique para cada documento dentro de um espaço de rede como descrito no Pedido de Patente Norte-Americana Número Serial 10/995.983 intitulado "SISTEMA E MÉTODO PARA CLASSIFICAR RESULTADOS DE PROCURA USANDO DISTÂNCIA DE CLIQUE" depositado em 30 de Agosto de 2004, um componente independente de consulta que leva em conta uma distância de clique induzida para cada documento dentro de um espaço de rede como descrito no Pedido de Patente Norte-Americana Número Serial 11/206.286 intitulado "FUNÇÕES DE CLASSIFICAÇÃO QUE USAM UMA DISTÂNCIA DE

CLIQUE INDUZIDA DE UM DOCUMENTO EM UMA REDE" depositado em 15 de Agosto de 2005, e um componente independente de consulta que leva em conta a URL para cada documento dentro de um espaço de rede como descrito no Pedido de Patente Norte-Americana Número Serial 10/955.983 intitulado "SISTEMA E MÉTODO PARA CLASSIFICAR RESULTADOS DE PROCURA USANDO DISTÂNCIA DE CLIQUE" depositado em 30 de Agosto de 2004. O assunto de cada um dos Pedidos de Patente Norte-Americana acima mencionados, que são atribuídos ao cessionário do presente pedido de patente, é aqui incorporado como referência em sua integridade.

Ainda em uma modalidade exemplificada, os métodos descritos de determinar uma classificação de relevância de documento utilizam uma função de classificação que compreende no mínimo um componente independente de consulta, que inclui ambos o parâmetro de uso acima descrito e um ou mais componentes adicionais independentes de consulta descritos acima.

A classificação de relevância de documento pode ser usada para classificar documentos em um espaço de rede. Por exemplo, um método de classificar documentos em uma rede pode compreender as etapas de determinar uma classificação de relevância de documento para cada documento na rede usando o método descrito acima; e classificar os documentos em uma ordem desejada (tipicamente, na ordem decrescente) baseada nas classificações de relevância de documento de cada documento.

A classificação de relevância de documento pode ser também usada para classificar os resultados de procura de uma consulta de procura. Por exemplo, um método de classificar resultados de procura de uma consulta de procura pode compreender as etapas de determinar uma classificação de relevância de documento para cada documento nos resultados de procura de uma consulta de procura usando o método descrito acima, e classificar os documentos em uma ordem desejada (tipicamente, na ordem decrescente) baseada nas classificações de relevância de documento de cada documento.

Os programas de aplicação que usam os métodos descritos aqui podem ser carregados e executados em uma variedade de sistemas de computador que compreendem uma variedade de componentes de hardware. Um sistema de computador exemplificado e um ambiente operacional exemplificado para praticar os métodos descritos aqui são descritos abaixo.

Ambiente Operacional Exemplificado

A FIG. 2 ilustra um exemplo de um ambiente de sistema de computação adequado 100 no qual os métodos descritos aqui podem ser implementados. O ambiente de sistema de computação 100 é somente um exemplo de um ambiente de computação adequado e não pretende sugerir qualquer limitação quanto ao escopo do uso ou funcionalidade dos métodos descritos aqui. O ambiente de computação 100 nem deveria ser interpretado como tendo qualquer dependência ou exigência relacionada a qualquer componente ou combinação de componentes ilustrados no ambiente operacional exemplificado 100.

Os métodos descritos aqui são operacionais com numerosos outros ambientes ou configurações de sistema de computação de propósito geral ou de propósito especial. Exemplos de sistemas de computação, ambientes, e/ou configurações bem conhecidos que podem ser adequados ao uso com os métodos descritos aqui incluem, porém não são limitados a, computadores pessoais, computadores servidores, computadores portáteis ou dispositivos laptops, sistemas multiprocessadores, sistemas baseados em microprocessador, aparelho de conexão à internet via TV, eletrônicos programáveis por consumidor, PCs de rede, minicomputadores, computadores de grande porte, ambientes de computação distribuídos que incluem qualquer um dos sistemas ou dispositivos acima, e seus similares.

Os métodos e processos descritos aqui podem ser descritos no contexto geral de instruções executáveis por computador, tal como módulos de programa, que são executados por um computador. Geralmente, os módulos de programa incluem rotinas, programas, objetos, componentes, estruturas de dados, etc. que executam tarefas particulares ou implementam tipo de dados abstratos particulares. Os métodos e processos descritos aqui também podem ser praticados em ambientes de computação distribuídos onde tarefas são executadas por dispositivos de processamento remoto que são ligados através de uma rede de comunicações. Em um ambiente de computação distribuído, os módulos de programa podem ser localizados em ambos meios de armazenamento por computador local e remoto incluindo dispositivos de armazenamento em memória.

Com relação à FIG. 2, um sistema exemplificado para implementar os métodos e processos descritos aqui inclui um dispositivo de computação de propósito geral na forma de um computador 110. Os componentes do computador 110 podem 5 incluir, porém não estão limitados a, uma unidade de processamento 120, uma memória de sistema 130, e um barramento de sistema 121 que acopla vários componentes de sistema, incluindo, porém não limitados a, memória de sistema 130, à unidade de processamento 120. O barramento de sistema 121 pode 10 ser qualquer um dos vários tipos de estruturas de barramentos incluindo um barramento de memória ou controlador de memória, um barramento periférico, e um barramento local usando qualquer uma de uma variedade de arquiteturas de barramento. A título de exemplo, e não limitação, tais arquiteturas 15 incluem barramento de Arquitetura Padrão de Indústria (ISA), barramento de Arquitetura de Micro Canal (MCA), barramento ISA Aperfeiçoado (EISA), barramento local de Associação de Padrões Eletrônicos de Vídeo (VESA), e barramento de Componentes Periféricos Interconectados (PCI), também conhecido como barramento Mezzanine.

O computador 110 tipicamente inclui uma variedade de meios legíveis por computador. Esses podem ser quaisquer meios disponíveis que podem ser acessados pelo computador 110 e incluem ambos meios voláteis e não voláteis, removíveis 25 e não removíveis. A título de exemplo, e não limitação, os meios legíveis por computador podem compreender meios de armazenamento por computador e meios de comunicação. Os meios de armazenamento por computador incluem ambos meios volá-

teis e não voláteis, removíveis e não removíveis implementados em qualquer método ou tecnologia para armazenamento de informação tal como instruções legíveis por computador, estruturas de dados, módulos de programa ou outros dados. Os 5 meios de armazenamento por computador incluem, mas não estão limitados a, RAM, ROM, EEPROM, memória rápida ou outra tecnologia de memória, CD-ROM, discos versáteis digitais (DVD) ou outro armazenamento em disco óptico, cassetes magnéticos, fita magnética, armazenamento em disco magnético ou outros 10 dispositivos de armazenamento magnéticos, ou qualquer outro meio que pode ser usado para armazenar a informação desejada e que pode ser acessado pelo computador 110. Os meios de comunicação tipicamente incorporam instruções legíveis por computador, estruturas de dados, módulos de programa ou outros 15 dados em um sinal modulado de dados tal como uma onda portadora ou outro mecanismo de transporte, e incluem quaisquer meios de entrega de informação. O termo "sinal modulado de dados" significa um sinal que tem uma ou mais de suas características ajustadas ou alteradas de tal maneira a codificar 20 informação no sinal. A título de exemplo, e não limitação, os meios de comunicação incluem meios por fios tais como uma rede por fios ou conexão direta por fios, e meios sem fio tais como acústicos, RF, infravermelhos e outros meios sem fio. Combinações de qualquer um dos acima deveriam 25 ser também incluídas no escopo dos meios legíveis por computador como usado aqui.

A memória de sistema 130 inclui meios de armazenamento por computador na forma de memória volátil e/ou não

volátil, tal como a memória somente de leitura (ROM) 131 e a memória de acesso aleatório (RAM) 132. Um sistema de entrada/saída básico (BIOS) 133, contendo as rotinas básicas que ajudam a transferir informação entre elementos no computador 5 110, tal como durante a inicialização, é tipicamente armazenado na ROM 131. A RAM 132 tipicamente contém dados e/ou módulos de programa que estão imediatamente acessíveis e/ou estão presentemente sendo operados pela unidade de processamento 120. A título de exemplo, e não limitação, a FIG. 2 10 ilustra o sistema operacional 134, programas de aplicação 135, outros módulos de programa 136, e dados de programa 137.

O computador 110 pode também incluir outros meios de armazenamento por computador removíveis/não removíveis, 15 voláteis/não voláteis. A título de exemplo somente, a FIG. 2 ilustra uma unidade de disco rígido 141 que lê a partir de e escreve em meios magnéticos não removíveis e não voláteis, uma unidade de disco magnético 151 que lê a partir de e escreve em discos magnéticos removíveis e não voláteis 152, e 20 uma unidade de disco óptico 155 que lê a partir de e escreve em um disco óptico removível e não volátil 156, tal como um CD-ROM ou outro meio óptico. Outros meios de armazenamento por computador removíveis/não removíveis, voláteis/não voláteis que podem ser usados no ambiente operacional exemplificado 25 incluem, mas não estão limitados a, fitas cassetes magnéticas, cartões de memória rápida, discos versáteis digitais, fita de vídeo digital, RAM de estado sólido, ROM de estado sólido, e seus similares. A unidade de disco rígido

141 é tipicamente conectada ao barramento de sistema 121 a-
través de uma interface de memória não removível tal como a
interface 140, e a unidade de disco magnético 151 e a unida-
de de disco óptico 155 são tipicamente conectadas ao barra-
5 mento de sistema 121 por uma interface de memória removível,
tal como a interface 150.

As unidades e seus meios de armazenamento por com-
putador associados discutidos acima e ilustrados na FIG. 2
fornecem armazenamento de instruções legíveis por computa-
10 dor, estruturas de dados, módulos de programa e outros dados
para o computador 110. Na FIG. 2, por exemplo, a unidade de
disco rígido 141 é ilustrada como armazenando o sistema ope-
racional 144, programas de aplicação 145, outros módulos de
programa 146, e dados de programa 147. Nota-se que esses
15 componentes podem ou ser os mesmos ou diferentes do sistema
operacional 134, programas de aplicação 135, outros módulos
de programa 136, e dados de programa 137. Ao sistema opera-
cional 144, aos programas de aplicação 145, a outros progra-
mas 146 e aos dados de programa 147 são atribuídos números
20 diferentes aqui para ilustrar que, no mínimo, eles são có-
pias diferentes.

Um usuário pode inserir comandos e informação no
computador 110 através de dispositivos de entrada, tal como
um teclado 162 e um dispositivo de apontamento 161, comumen-
25 te referido como um mouse, mouse estacionário ("trackball")
ou mesa sensível ao toque. Outros dispositivos de entrada
(não mostrados) podem incluir um microfone, um comando de
jogos, mesa de jogos, antena de satélite, digitalizador, ou

seus similares. Esses e outros dispositivos de entrada são freqüentemente conectados à unidade de processamento 120 através de uma interface de entrada de usuário 160 que é acoplada ao barramento de sistema 121, mas pode ser conectada 5 por outra interface e estruturas de barramento, tal como uma porta paralela, porta de jogos ou um barramento serial universal (USB). Um monitor 191 ou outro tipo de dispositivo de exibição está também conectado ao barramento de sistema 121 via uma interface, tal como uma interface de vídeo 190. Em 10 adição ao monitor 191, o computador 110 pode também incluir outros dispositivos periféricos de saída tal como alto-falantes 197 e impressora 196, que podem ser conectados através de uma interface periférica de saída 195.

O computador 110 pode operar em um ambiente de rede 15 usando conexões lógicas a um ou mais computadores remotos, tal como um computador remoto 180. O computador remoto 180 pode ser um computador pessoal, um servidor, um roteador, um PC de rede, um dispositivo não hierárquico ou outro nó de rede comum, e tipicamente inclui muitos ou todos os 20 elementos descritos acima em relação ao computador 110, embora somente um dispositivo de armazenamento em memória 181 foi ilustrado na FIG. 2. As conexões lógicas representadas na FIG. 2 incluem uma rede de área local (LAN) 171 e uma rede de área ampla (WAN) 173, mas podem também incluir outras 25 redes. Tais ambientes de rede são comuns em escritórios, redes de computador de grandes empresas, intranets e a Internet.

Quando usado em um ambiente de rede LAN, o computador 110 é conectado à LAN 171 através de uma interface de rede ou adaptador 170. Quando usado em um ambiente de rede WAN, o computador 110 tipicamente inclui um modem 172 ou outro dispositivo para estabelecer comunicações pela WAN 173, tal como a Internet. O modem 172, que pode ser interno ou externo, pode estar conectado ao barramento de sistema 121 via a interface de entrada de usuário 160, ou outro mecanismo apropriado. Em um ambiente de rede, os módulos de programa representados em relação ao computador 110, ou partes desses, podem ser armazenados no dispositivo de armazenamento em memória remota. A título de exemplo, e não limitação, a FIG. 2 ilustra programas de aplicação remota 185 como residindo no dispositivo de memória 181. Será apreciado que as conexões de rede mostradas são exemplificadas e outros meios de estabelecer uma ligação de comunicação entre os computadores podem ser usados.

Métodos e processos discutidos aqui podem ser implementados usando um ou mais programas de aplicação que incluem, mas não estão limitados a, um aplicativo de software de sistema servidor (por exemplo, aplicativo de software WINDOWS SERVER SYSTEMTM), um aplicativo de classificação de procura, e um aplicativo para gerar, manter e armazenar dados de uso de documentos em um espaço de rede (por exemplo, aplicativo de serviços WINDOWS[®] SHAREPOINT[®]), qualquer um dos quais poderia ser um dos vários programas de aplicação designados como programas de aplicação 135, programas de a-

plicação 145 e programas de aplicação remota 185 no sistema exemplificado 100.

Como mencionado acima, aqueles versados na técnica apreciarão que os métodos descritos de gerar uma classificação de relevância de documento, para um dado documento, podem ser implementados em outras configurações do sistema de computador, que inclui dispositivos portáteis, sistemas multiprocessadores, eletrônicos programáveis de consumidor ou baseados em microprocessador, computadores pessoais em rede, minicomputadores, computadores de grande porte, e seus similares. Os métodos descritos de gerar uma classificação de relevância de documento para um dado documento podem também ser praticados em ambientes de computação distribuídos, onde tarefas são executadas por dispositivos de processamento remoto que são ligados através de uma rede de comunicações. Em um ambiente de computação distribuído, os módulos de programa podem estar localizados em ambos dispositivos de armazenamento em memória local e remoto.

Implementação de Modalidades Exemplificadas

Como discutido acima, métodos de determinar uma classificação de relevância de documento para um documento em uma rede são fornecidos. Os métodos descritos podem classificar um documento em uma rede utilizando uma função de classificação que leva em conta um valor de uso de documento de cada documento na rede.

Os métodos descritos de determinar uma classificação de relevância de documento para um documento em uma rede podem compreender um número de etapas. Em uma modalidade e-

xemplificada, o método de determinar uma classificação de relevância de documento para um documento em uma rede compreende as etapas de atribuir um valor de uso atual (U_A) a um ou mais documentos em uma rede compreendendo N documentos, onde o valor de uso atual (U_A) é baseado em dados de uso atual mantidos e armazenados em um servidor; e a menos do que N documentos são atribuídos um valor de uso atual (U_A), atribuir um valor de uso padrão (U_D) aos documentos que não têm dados de uso atual associados a eles; e usar o valor de uso (isto é, U_A ou U_D) para cada documento para determinar a classificação de relevância de documento de um dado documento na rede.

Como usado aqui, o termo "dados de uso atual" representa um ou mais tipos de dados associados com o "uso" do documento por um ou mais usuários. Tipos de dados de uso atual para um dado documento ou conjunto de documentos podem incluir, mas não estão limitados, o número de visualizações de documentos por todos os usuários em um dado período de tempo, o número médio de visualizações de documento por usuário em um dado período de tempo, o tempo total gasto em um documento particular em um dado período de tempo, o tempo médio gasto em um documento particular em um dado período de tempo, etc. O dado período de tempo pode ser, por exemplo, a última semana, o último mês, o último ano, ou o tempo de vida do documento, ou qualquer outro período desejado de tempo.

As etapas de gerar, manter e armazenar dados ou estatísticas de uso de documento para documentos em um espa-

ção de rede podem ser executadas por código de aplicação comumente encontrado em sistemas de computação. Os dados de uso de documento são gerados, mantidos e armazenados independentemente de uma dada consulta de procura ou mecanismo de procura, e são tipicamente gerados, mantidos e armazenados por código de aplicação no servidor que mantém o documento (ou página) e torna o documento (ou página) disponível para um usuário. Os programas de aplicação adequados para gerar, manter e armazenar dados ou estatísticas de uso de documento incluem, mas não estão limitados, Serviços WINDOWS® SHAREPOINT® e outros programas de aplicação similares.

Os dados de uso de documento armazenados e mantidos nesses sítios de serviço, bem como outros sítios da Rede executando uma função similar, podem ser acessados usando código de aplicação como discutido acima. Por exemplo, os dados de uso de documento podem ser acessados a partir de um dado sítio da Rede (por exemplo, um sítio de Serviços WINDOWS® SHAREPOINT®) via (i) uma interface de programação de aplicação especial (API), (ii) uma solicitação de serviço da Rede, ou (iii) solicitando uma pagina da Rede de administração que retorna dados de uso para cada URL no sítio da Rede.

Os métodos descritos de determinar uma classificação de relevância de documento para um documento em uma rede podem compreender um número de etapas adicionais incluindo, mas não limitadas, monitorar um ou mais documentos em um espaço de rede para uso atual de documento; armazenar dados de

uso atual de documento para um ou mais documento em um arquivo de armazenamento de dados local ou remoto; calcular um valor de uso atual (U_A) para um documento baseado nos dados de uso atual para o documento ou uma pasta contendo o documento; 5 armazenar valores de uso atual (U_A) para um ou mais documentos em um arquivo de armazenamento de dados local ou remoto; solicitar dados de uso de documento armazenado ou valores de uso atual (U_A) a partir de um arquivo de armazenamento de dados local ou remoto (por exemplo, uma solicitação por tais dados a partir de um mecanismo de procura depois de uma consulta de procura específica por um usuário); restaurar os dados de uso de documento atual ou um valor de uso atual (U_A) para um ou mais documentos de um arquivo de armazenamento de dados local ou remoto; e opcionalmente, 10 combinar um valor de uso de documento (isto é, atual ou padrão) com uma ou mais propriedades de documento adicionais para determinar uma classificação de relevância de documento 15 para um documento.

A FIG. 3 representa um fluxograma lógico que mostra etapas exemplificadas em um método exemplificado de fornecer valores de uso atual ou padrão para documentos em uma rede seguidos por um procedimento opcional de rebaixamento/atualização por um administrador do sistema. Como mostrado na FIG. 3, o método exemplificado 401 inicia no bloco 402 e prossegue para a etapa 403. Na etapa 403, um primeiro documento na rede é rastreado por dados de uso atual.

A etapa de rastrear um primeiro documento por dados de uso atual (etapa 403) pode ser executada usando uma

aplicação rastreadora capaz de determinar se o primeiro documento tem quaisquer dados de uso atual associados a ele, e se o primeiro documento tem dados de uso atual associados a ele, restaurando os dados de uso atual. As aplicações rastreadoras adequadas para uso nos métodos descritos de fornecer valores de uso atual ou padrão para documentos em uma rede incluem, mas não estão limitadas, aplicações rastreadoras descritas nas Patentes Norte-Americanas Nos. 6.463.455 e 6.631.369, o assunto de ambas é aqui incorporado em sua integridade como referência.

Como discutido acima, os dados de uso atual podem ser obtidos a partir de um ou mais arquivos que armazenam dados de uso atual para um ou mais documentos em uma rede. Os dados de uso atual podem ser armazenados, junto com o documento, como um componente do documento, ou podem ser armazenados em um arquivo de armazenamento de dados separado do documento atual. Os sistemas de armazenamento remoto adequados incluem, mas não estão limitados a, produtos de Serviços de WINDOWS® SHAREPOINT® (WSS) comercialmente disponíveis a partir da Microsoft Corporation (Redmond, WA), bem como qualquer outro sistema de armazenamento remoto. Por exemplo, o sistema de armazenamento remoto WSS grava dados de uso atual incluindo, por exemplo, o número de solicitações a cada documento em uma dada rede através de todos os usuários, e produz estatísticas de números de cliques por documentos durante a última semana, o último mês, o último ano, ou o tempo de vida total do documento, ou qualquer outro período de tempo. Ademais, como notado acima, deveria ser entendido que

os métodos descritos aqui não estão limitados a um sistema de armazenamento remoto WSS, mas podem utilizar um sistema de armazenamento remoto WSS ou qualquer outro sistema de dados de documento similar nos métodos descritos.

5 Uma vez que o documento é rastreado, o método exemplificado 401 prossegue para o bloco de decisão 404. No bloco de decisão 404, uma determinação é feita pelo código de aplicação como se o documento tem dados de uso atual associados a ele. Se uma decisão é feita de que o documento
10 tem dados de uso atual associados a ele, o método exemplificado 401 prossegue para a etapa 405, onde um valor de uso baseado no uso atual (U_A) é atribuído ao documento. O valor de uso atual (U_A) pode ser determinado usando um ou mais componentes dos dados de uso atual associados com o documento.
15 Por exemplo, em algumas modalidades, o valor de uso atual (U_A) atribuído ao documento pode ser relacionado somente ao número de usuários visualizando o documento. Em outras modalidades, o valor de uso atual (U_A) atribuído ao documento pode estar relacionado ao número de visualizações do documento por todos os usuários em um dado período de tempo, o número médio de visualizações do documento por usuário em um dado período de tempo, o tempo total gasto em um documento particular em um dado período de tempo, o tempo médio gasto em um documento particular em um dado período de tempo, ou
20 uma combinação de qualquer um dos critérios acima, onde o dado período de tempo compreende a última semana, o último mês, o último ano, ou o tempo de vida do documento, ou qualquer outro período desejado de tempo.
25

Em alguns casos, os dados de uso atual associados com um dado documento sugerem que o documento não foi usado ou visualizado durante um dado período de tempo. Em tal caso, ao documento poderia ser atribuído um valor de uso (U_A) igual a zero para indicar nenhum uso durante o período de tempo; entretanto, tipicamente, aos valores de uso (U_A) baseados no uso atual ou nenhum uso atual são atribuídos um número além de zero.

Ademais, em alguns casos, os dados de uso atual podem ser associados a um conjunto de documentos como oposto a documentos individuais. Por exemplo, uma pasta pode conter um conjunto de documentos, e um servidor associado pode somente rastrear dados de uso relacionados ao acesso à pasta (isto é, uso da pasta), e não aos documentos individuais dentro da pasta. Nessa modalidade, se houver dados de uso atual associados com uma pasta, um valor de uso (U_A) pode ser fornecido para cada documento dentro da pasta baseado nos dados de uso atual da pasta. Tipicamente, cada valor de uso (U_A) será o mesmo para cada documento dentro da pasta; entretanto, diferentes valores de uso (U_A) podem ser atribuídos a diferentes documentos em uma pasta se assim desejado.

A partir da etapa 405, o método exemplificado 401 prossegue para o bloco de decisão 406 descrito abaixo.

Retornando ao bloco de decisão 404, se uma decisão é feita de que o documento não tem dados de uso atual associados a ele, o método exemplificado 401 prossegue para a etapa 407, onde um valor de uso padrão (U_D) é atribuído ao documento. Por exemplo, um valor de uso padrão (U_D) pode ser

atribuído a um documento que é parte de um sítio da Rede que não mantém dados de uso de documento. O valor de uso padrão (U_D) atribuído ao documento pode ser usado para fornecer uma importância inicial ao documento em relação aos documentos que têm dados de uso atual. Por exemplo, se um valor de uso mais alto para um dado documento indica importância relativa do documento na rede, atribuir um valor de uso padrão mais baixo (U_D) ao documento rebaixa a importância do documento em relação a outros documentos na rede.

Em uma modalidade exemplificada onde um valor de uso mais alto para um dado documento indica importância relativa do documento na rede, o valor de uso padrão (U_D) atribuído ao documento pode ser relativo aos valores de uso atual (U_A) atribuídos a outros documentos na rede. Por exemplo, de modo a abaixar a importância relativa do documento, um valor de uso padrão (U_D) pode ser atribuído ao documento, onde este é menor do que qualquer valor de uso atual (U_A) atribuído a outros documentos na rede como descrito acima. Se for desejado aumentar a importância relativa do documento, um valor de uso padrão (U_D) pode ser atribuído ao documento, onde este é maior do que qualquer valor de uso atual (U_A) atribuído a outros documentos na rede ou maior do que alguns dos valores de uso atual (U_A) atribuídos a alguns dos outros documentos na rede.

Em outras modalidades, um valor de uso padrão (U_D) pode ser atribuído a um documento sem dados de uso atual tal que ao documento é dada uma importância relativa média comparada aos documentos que têm um valor de uso atual atribuí-

do (U_A). Por exemplo, nessa modalidade, os valores de uso padrão (U_D) para documentos sem dados de uso atual podem estar na faixa de um valor de uso atual atribuído mínimo (U_{A-min}) a um valor de uso atual atribuído máximo (U_{A-max}), ou podem estar em uma faixa específica entre o valor de uso atual atribuído mínimo (U_{A-min}) e o valor de uso atual atribuído máximo (U_{A-max}). Nessa modalidade, os documentos sem dados de uso atual são fornecidos com uma importância relativa média, sugerindo uso médio, comparado a documentos que têm dados de uso atual associados a eles.

A partir da etapa 407, o método exemplificado 401 prossegue para o bloco de decisão 406. Neste, uma determinação é feita pelo código de aplicação como se todos os documentos em uma rede têm um valor de uso atual (U_A) ou padrão (U_D), o método exemplificado prossegue para a etapa 408, onde o próximo documento é rastreado por dados de uso atual. A partir da etapa 408, o método exemplificado 401 retorna para o bloco de decisão 404 e prossegue como discutido acima.

Retornando ao bloco de decisão 406, se uma determinação é feita pelo código de aplicação de que todos os documentos na rede têm um valor de uso atual (U_A) ou padrão (U_D), o método exemplificado 401 prossegue para o bloco de decisão 409. Neste, uma determinação é feita por um administrador do sistema se rebaixar quaisquer valores de uso atual (U_A) ou padrão (U_D) de modo a representar de forma mais próxima a importância de um dado documento em um espaço de rede. Se uma decisão é feita para rebaixar um ou mais valores de uso atual (U_A) ou padrão (U_D) de modo a representar de

forma mais próxima a importância de um dado documento em um espaço de rede, o método exemplificado 401 prossegue para a etapa 410, onde os valores de uso atual (U_A) ou padrão (U_D) de um ou mais documentos (ou URLs) são ajustados ou negativamente ou positivamente. A partir da etapa 410, o método exemplificado 401 prossegue para a etapa 411 descrita abaixo.

Retornando ao bloco de decisão 409, se uma decisão é feita para não rebaixar (ou atualizar) um ou mais valores de uso atual (U_A) ou padrão (U_D), o método exemplificado 401 prossegue diretamente para a etapa 411. Na etapa 411, os valores de uso atual (U_A) ou padrão (U_D) são utilizados em uma função de classificação para determinar uma classificação de relevância de documento total para cada documento em um espaço de rede. A partir da etapa 411, o método exemplificado 401 prossegue para o bloco final 412.

Uma vez que todos os valores de uso atual (U_A) ou padrão (U_D) tenham sido determinado e opcionalmente rebaixados (ou atualizados), se assim desejado, o valor de uso atual (U_A) ou padrão (U_D) para cada documento pode ser usado como um parâmetro em uma função de classificação para fornecer classificação de relevância para cada documento. Tal classificação de relevância de documento pode ser usada para classificar resultados de procura de uma consulta de procura. Um método exemplificado de classificar resultados de procura gerados usando uma função de classificação contendo um parâmetro de valor de uso de documento é mostrado na FIG.

A FIG. 4 fornece um fluxograma lógico que mostra etapas exemplificadas no método exemplificado 20 que compreende um método de classificar resultados de procura gerados usando uma função de classificação contendo um parâmetro de 5 valor de uso. Como mostrado na FIG. 4, o método exemplificado 20 inicia no bloco 201 e prossegue para a etapa 202. Na etapa 202, um usuário solicita uma procura inserindo uma consulta de procura. Antes da etapa 202, valores de uso atual ou padrão para cada um dos documentos na rede foram pre-10 viamente calculados. A partir da etapa 202, o método exemplificado 20 prossegue para a etapa 203.

Na etapa 203, o valor de uso atual ou padrão para cada documento em uma rede é combinado com qualquer outras estatística de documento (por exemplo, outra estatística in-15 dependente de consulta) para cada documento armazenado no índice. Combinando os valores de uso atual ou padrão com outras estatísticas de documento permite um tempo de resposta à consulta mais rápido já que toda a informação relacionada à classificação é agrupada junta. Conseqüentemente, cada do-20 cumento listado no índice tem um valor de uso atual ou pa- drão associado depois da combinação. Uma vez que a combina- ção está completa, o método exemplificado 20 prossegue para a etapa 204.

Na etapa 204, as estatísticas de documento inde-25 pendente de consulta para um dado documento, incluindo um parâmetro de uso, são fornecidas como um componente de uma função de classificação. Dados dependentes de consulta são também fornecidos para o dado documento, tipicamente como um

componente separado da função de classificação. Os dados dependentes de consulta ou partes relacionada a conteúdo da função de classificação dependem dos termos de procura atual e do conteúdo do dado documento.

5 Em uma modalidade, a função de classificação compreende pelo menos um componente independente de consulta (QID) que compreende um parâmetro de uso. Em uma modalidade, o componente independente de consulta (QID) pode ser representado pela seguinte equação:

$$10 \quad QID(doc) = w_u \frac{k_u U}{k_u + U}$$

(1)

onde:

U representa um valor de uso atual ou um valor de uso padrão; e

15 w_u e k_u representam parâmetros de ajuste para o valor de uso. Em uma modalidade adicional, o componente independente de consulta (QID) pode ser representado pela seguinte equação:

$$QID(doc) = w_u U + k_u$$

20 (2)

onde:

U representa um valor de uso atual ou um valor de uso padrão; e

25 w_u e k_u representam parâmetros de ajuste para o valor de uso. Em ainda uma modalidade adicional, o componente independente de consulta (QID) pode ser representado pela seguinte equação:

$$QID(doc) = w_u [1 + \exp(-k_u U - B)] + C$$

(3)

onde:

U representa um valor de uso atual ou um valor de uso padrão; e

w_u , k_u , B e C representam parâmetros de ajuste (isto é, constantes escalares) para o valor de uso.

Em uma modalidade adicional, a função de classificação compreende a soma do componente independente de consulta (QID) acima descrito e pelo menos um componente dependente de consulta (QD) tal como

$$\text{Classificação} = QD(doc, consulta) + QID(doc).$$

O componente QD pode ser qualquer função de classificação de documento. Em uma modalidade, o componente QD corresponde a uma função de classificação ponderada por campo descrita no Pedido de Patente Norte-Americana No. Serial 10/804.326 intitulada "PONDERAÇÃO POR CAMPO EM PROCURA DE DOCUMENTO DE TEXTO", depositado em 18 de Março de 2004, o assunto da qual é aqui incorporado como referência em sua integridade. Como fornecido no Pedido de Patente Norte-Americana No. Serial 10/804.326, uma equação que pode ser usada como uma representação da função de classificação ponderada por campo é como segue:

$$QD(doc, consulta) = \sum \frac{wtf'(k_i + 1)}{k_i + wtf'} \times \log\left(\frac{N}{n}\right)$$

25

onde:

wtf' representa uma freqüência de termo ponderada ou soma de freqüências de termo de dados termos na consulta de procura multiplicada pelos pesos através de todos os cam-

pos (por exemplo, o título, o corpo, etc. do documento) e normalizados de acordo com o comprimento de cada campo e o comprimento médio correspondente,

N representa um número de documento na rede,

5 n representa um número de documentos contendo um termo de consulta, e

k₁ é uma constante ajustável.

Os termos e equação acima são adicionalmente des- critos em detalhes no Pedido de Patente Norte-Americana No.

10 Serial 10/804.326, o assunto da qual é aqui incorporado como referência em sua integridade.

Em algumas modalidades, a função de classificação pode adicionalmente compreender um componente QID que leva em conta (i) um valor de distância de clique determinado pe- 15 los métodos descritos no Pedido de Patente Norte-Americana No. Serial 10/955.983, intitulada "SISTEMA E MÉTODO PARA CLASSIFICAR RESULTADOS DE PROCURA USANDO DISTÂNCIA DE CLIQUE", depositado em 30 de Agosto de 2004, (ii) um valor de distância de clique induzida determinado pelos métodos 20 descritos no Pedido de Patente Norte-Americana No. Serial 11/206.286 intitulada "FUNÇÕES DE CLASSIFICAÇÃO USANDO UMA DISTÂNCIA DE CLIQUE INDUZIDA DE UM DOCUMENTO EM UMA REDE" depositado em 15 de Agosto de 2005, o assunto de ambos das quais é incorporado aqui como referência em sua integridade, 25 (iii) uma profundidade URL de um documento, ou (iv) uma combinação de (i) ou (ii) e (iii). Por exemplo, esse componente QID adicional pode compreender uma função como segue:

$$QID(doc) = w_{cd} \frac{k_{cd}}{b_{cd} \frac{CD}{k_{ew}} + b_{ud} UD} \\ k_{cd} + \frac{k_{ew}}{b_{cd} + b_{ud}}$$

onde:

w_{cd} representa um peso de um componente independente de consulta tal como um componente contendo um parâmetro de distância de clique ou distância de clique induzida,

b_{cd} representa um peso de uma distância de clique ou distância de clique induzida em relação à profundidade URL,

b_{ud} representa um peso de uma profundidade URL,

CD representa uma distância de clique computada ou atribuída ou distância de clique induzida para um documento,

k_{ew} representa uma constante de ajuste que é determinada pela otimização da precisão da função de classificação, similar a outros parâmetros de ajuste (isto é, k_{ew} pode representar o valor de peso de margem quando todas as margens têm o mesmo valor de peso de margem, ou k_{ew} pode representar o valor de margem médio quando os valores de peso de margem diferem uns dos outros),

UD representa uma profundidade de URL, e

k_{cd} é a constante de saturação de distância de clique.

Os termos ponderados (w_{cd} , b_{cd} e b_{ud}) auxiliam na definição da importância de cada um dos seus termos relacionados (isto é, o componente contendo um parâmetro de distância de clique ou distância de clique induzida, a distância de clique ou distância de clique induzida para um dado docu-

mento, e a profundidade URL do dado documento, respectivamente) e por fim o resultado da função de classificação.

A profundidade de URL (UD) é uma adição opcional ao componente independente de consulta acima referido para suavizar o efeito que o valor de distância de clique ou de distância de clique induzida pode ter na função de classificação. Por exemplo, em alguns casos, um documento que não é muito importante (isto é, tem uma grande profundidade de URL) pode ter uma curta distância de clique ou valor de distância de clique induzida. A profundidade de URL é representada pelo número de barras diagonais em uma URL do documento. Por exemplo, www.example.com\d1\d2\d3\d4.htm inclui quatro barras diagonais e, portanto, teria uma profundidade de URL de 4. Esse documento, entretanto, pode ter uma ligação diretamente a partir da página principal www.example.com dando a ela um valor de distância de clique ou de distância de clique induzida relativamente baixo. Incluindo o termo de profundidade de URL na função acima referida e ponderando o termo de profundidade de URL contra o valor de distância de clique ou de distância de clique induzida compensa um valor de distância de clique ou de distância de clique induzida relativamente alto para refletir mais precisamente a importância do documento na rede. Dependendo da rede, uma profundidade de URL de 3 ou mais pode ser considerada uma ligação profunda.

Em uma modalidade, a função de classificação usada para determinar uma classificação de relevância de documento para um dado documento compreende uma função como segue:

$$\text{Classificação} = \sum \frac{wtf'(k_1+1)}{k_1 + wtf'} \times \log\left(\frac{N}{n}\right) + w_{cd} \frac{\frac{k_{cd}}{b_{cd} \frac{CD}{k_{ew}} + b_{ud} UD}}{k_{cd} + \frac{k_{ew}}{b_{cd} + b_{ud}}} + w_u \frac{k_u U}{k_u + U}$$

onde os termos são como descrito acima.

Em outras modalidades, a profundidade de URL pode ser removida da função de classificação ou outros componentes podem ser adicionados à função de classificação para aperfeiçoar a precisão do componente dependente de consulta, do componente independente de consulta, ou de ambos. Além disso, o componente independente de consulta acima descrito contendo um parâmetro de uso pode ser incorporado em outras funções de classificação (não mostradas) para aperfeiçoar a classificação de resultados de procura.

Uma vez que as estatísticas de documento para um dado documento são fornecidas a uma função de classificação na etapa 204, o método exemplificado 20 prossegue para a etapa 205. Na etapa 205, uma classificação de relevância de documento é determinada para um dado documento, armazenada na memória, e associada com o dado documento. A partir da etapa 205, o método exemplificado 20 prossegue para o bloco de decisão 206.

No bloco de decisão 206, uma determinação é feita pelo código de aplicação se uma classificação de relevância de documento foi calculada para cada documento em uma rede. Se uma determinação é feita de que uma classificação de relevância de documento não foi calculada para cada documento em uma rede, o método exemplificado 20 retorna para a etapa 204 e continua como descrito acima. Se uma determinação é

feita de que uma classificação de relevância de documento foi calculada para cada documento em uma rede, o método exemplificado 20 prossegue para a etapa 207.

Na etapa 207, os resultados de procura da consulta 5 compreendendo numerosos documentos são classificados de acordo com suas classificações de relevância de documento associadas. As classificações de relevância de documento resultantes levam em conta o valor de uso atual ou padrão de cada um dos documentos na rede. Uma vez que os resultados de 10 procura são classificados, o método exemplificado 20 prossegue para a etapa 208 onde resultados classificados são exibidos a um usuário. A partir da etapa 208, o método exemplificado 20 prossegue para a etapa 209 onde os resultados classificados mais altos são selecionados e visualizados pelo 15 usuário. A partir da etapa 209, o método exemplificado 20 prossegue para a etapa 210 onde o método exemplificado 20 termina.

Em adição aos métodos acima descritos de gerar uma classificação de relevância de documento para documentos em 20 uma rede e de usar classificações de relevância de documento para classificar resultados de procura de uma consulta de procura, o meio legível por computador tendo armazenado neste instruções executáveis por computador para executar os métodos acima descritos é descrito também aqui.

25 Sistemas de computação são também descritos aqui. Um sistema de computação exemplificado contém pelo menos um módulo de aplicação utilizado no sistema de computação, onde pelo menos um módulo de aplicação compreende código de apli-

cação carregado neste, onde o código de aplicação executa um método de gerar uma classificação de relevância de documento para documentos em uma rede. O código de aplicação pode ser carregado no sistema de computação usando qualquer um dos 5 meios legíveis por computador acima descritos tendo nestes instruções executáveis por computador para gerar uma classificação de relevância de documento para documentos em uma rede e para usar classificações de relevância de documento para classificar resultados de procura de uma consulta de 10 procura como descrito acima.

Enquanto a especificação foi descrita em detalhes com relação às modalidades específicas dessa, aqueles versados na técnica apreciarão, mediante um entendimento do anterior, que podem prontamente conceber as alterações, variações, e equivalentes a essas modalidades. Conseqüentemente, 15 o escopo dos métodos descritos, o meio legível por computador, e os sistemas de computação deveriam ser avaliados como aqueles das reivindicações em anexo e quaisquer equivalentes a esses.

REIVINDICAÇÕES

1. Meio legível por computador, **CARACTERIZADO** pelo fato de que compreende instruções executáveis por computador armazenadas nele para classificar documentos em uma rede, as 5 ditas instruções executáveis por computador utilizam uma função de classificação que compreende um ou mais componentes independentes de consulta, onde pelo menos um componente independente de consulta inclui um parâmetro de uso que leva em conta dados de uso gerados por servidor, armazenados em 10 servidor para um ou mais documentos na rede.

2. Meio legível por computador, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que o valor de uso compreende (i) um valor de uso atual baseado nos dados de uso atual mantidos por um servidor ou (ii) um valor de 15 uso padrão que não é baseado em dados de uso atual.

3. Meio legível por computador, de acordo com a reivindicação 2, **CARACTERIZADO** pelo fato de que o valor de uso atual é dependente de uma ou mais propriedades relacionadas ao uso de um documento ou de uma pasta contendo um 20 conjunto de documentos, as ditas uma ou mais propriedades relacionadas ao uso compreendem um número total de visualizações de documento ou de pasta por usuários em um dado período de tempo, um número médio de visualizações de documento ou de pasta por usuário em um dado período de tempo, um 25 tempo total gasto em um documento ou pasta particular em um dado período de tempo, um tempo médio gasto em um documento ou pasta particular em um dado período de tempo, onde o dado período de tempo compreende a última semana, o último mês, o

último ano, ou o tempo de vida total do documento ou pasta, ou qualquer outro período de tempo.

4. Meio legível por computador, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que pelo menos 5 um componente independente de consulta é representado por uma fórmula:

$$QID(doc) = w_u \frac{k_u U}{k_u + U}$$

onde:

U representa um valor de uso atual ou um valor de 10 uso padrão; e

w_u e k_u representam parâmetros de ajuste para o valor de uso.

5. Meio legível por computador, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que pelo menos 15 um componente independente de consulta inclui ambos (i) o parâmetro de uso e (ii) um parâmetro de distância de clique ou de distância de clique induzida.

6. Meio legível por computador, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que pelo menos 20 um componente independente de consulta inclui ambos o parâmetro de uso e um parâmetro de profundidade de URL.

7. Meio legível por computador, de acordo com a reivindicação 1, **CARACTERIZADO** adicionalmente pelo fato de que compreende instruções executáveis por computador para 25 atribuir uma classificação gerada pela função de classificação a cada documento na rede, a dita classificação é usada para classificar documentos em ordem.

8. Meio legível por computador, de acordo com a reivindicação 7, **CARACTERIZADO** pelo fato de que a classificação para cada documento é gerada usando uma fórmula:

$$\text{Classificação} = \sum \frac{\text{wtf}'(k_1+1)}{k_1 + \text{wtf}'} \times \log\left(\frac{N}{n}\right) + w_{cd} \frac{k_{cd}}{b_{cd} \frac{CD}{k_{ew}} + b_{ud} UD} + w_u \frac{k_u U}{k_u + U}$$

$$k_{cd} + \frac{k_{ew}}{b_{cd} + b_{ud}}$$

5 onde:

wtf' representa uma freqüência de termo ponderada,

N representa um número de documentos na rede,

n representa um número de documentos contendo um termo de consulta,

10 w_{cd} representa um peso de um componente independente de consulta,

b_{cd} representa um peso de uma distância de clique,

b_{ud} representa um peso de uma profundidade de URL,

CD representa uma distância de clique computada ou
15 distância de clique induzida atribuída a um documento,

k_{ew} representa uma constante de ajuste relacionada a pesos de margem,

UD representa uma profundidade de URL,

U representa um valor de uso atual ou um valor de
20 uso padrão,

w_u e k_u representam parâmetros de ajuste para o valor de uso, e

k_{cd} e k_1 são constantes.

9. Meio legível por computador, de acordo com a

25 reivindicação 1, **CARACTERIZADO** adicionalmente pelo fato de que compreende instruções executáveis por computador para

aceitar um pedido de consulta inserido por um usuário, conduzir uma procura dos documentos na rede para gerar resultados de procura compreendendo múltiplos documentos, classificar os múltiplos documentos dos resultados de procura usando 5 a função de classificação para gerar resultados de procura classificados, e exibir os resultados de procura classificados ao usuário.

10. Meio legível por computador, de acordo com a reivindicação 1, **CARACTERIZADO** adicionalmente pelo fato de 10 que comprehende instruções executáveis por computador para habilitar um administrador a manualmente ajustar os resultados de classificação gerados pela função de classificação.

11. Sistema de computação, **CARACTERIZADO** pelo fato de que contém pelo menos um módulo de aplicação utilizado no 15 sistema de computação, onde pelo menos um módulo de aplicação comprehende código de aplicação carregado neste a partir do meio legível por computador de acordo com a reivindicação 1.

12. Método de determinar uma classificação de relevância de documento para um documento em uma rede, 20 **CARACTERIZADO** pelo fato de que comprehende as etapas de:

atribuir um valor de uso atual (U_A) a um ou mais documentos em uma rede compreendendo N documentos, onde o valor de uso atual (U_A) é baseado nos dados de uso real mantidos e armazenados em um servidor;

se a menos do que N documentos são atribuídos um valor de uso atual (U_A), atribuir um valor de uso padrão

(U_D) aos documentos que não têm dados de uso atual associados a eles; e

usar o valor de uso para cada documento para determinar a classificação de relevância de documento de um dado documento na rede.

13. Método, de acordo com a reivindicação 12, **CARACTERIZADO** adicionalmente pelo fato de que compreende a etapa de:

restaurar os dados de uso atual ou um valor de uso atual (U_A) para um documento a partir de um arquivo de armazenamento de dados no servidor.

14. Método, de acordo com a reivindicação 12, **CARACTERIZADO** adicionalmente pelo fato de que compreende a etapa de:

15. armazenar os dados de uso atual ou um valor de uso atual (U_A) para um documento em um arquivo de armazenamento de dados.

15. Método, de acordo com a reivindicação 12, **CARACTERIZADO** pelo fato de que a classificação de relevância de documento para cada documento na rede é gerada usando uma fórmula:

$$\text{Classificação} = \sum \frac{\text{wtf}'(k_l+1)}{k_l + \text{wtf}'} \times \log\left(\frac{N}{n}\right) + w_{cd} \frac{k_{cd}}{b_{cd} \frac{CD}{k_{ew}} + b_{ud} UD} + w_u \frac{k_u U}{k_u + U}$$

$$k_{cd} + \frac{k_{ew}}{b_{cd} + b_{ud}}$$

onde:

wtf' representa uma freqüência de termo ponderada,

25 N representa um número de documentos na rede,

n representa um número de documentos contendo um termo de consulta,

w_{cd} representa um peso de um componente independente de consulta,

5 b_{cd} representa um peso de uma distância de clique,

b_{ud} representa um peso de uma profundidade de URL,

CD representa uma distância de clique computada ou distância de clique induzida atribuída a um documento,

k_{ew} representa uma constante de ajuste relacionada

10 a pesos de margem,

UD representa uma profundidade de URL,

U representa um valor de uso atual ou um valor de uso padrão,

w_u e k_u representam parâmetros de ajuste para o

15 valor de uso, e

k_{cd} e k_l são constantes.

16. Método de classificar documentos em uma rede,

CARACTERIZADO pelo fato de que compreende as etapas de:

determinar uma classificação de relevância de documento para cada documento na rede usando o método da reivindicação 12; e

classificar os documentos em ordem decrescente baseados nas classificações de relevância de documento para cada documento.

25 17. Método de classificar resultados de procura de uma consulta de procura, **CARACTERIZADO** pelo fato de que compreende as etapas de:

determinar uma classificação de relevância de documento para cada documento nos resultados de procura de uma consulta de procura usando o método da reivindicação 12; e

5 classificar os documentos em ordem decrescente baseados nas classificações de relevância de documento de cada documento.

18. Meio legível por computador, **CARACTERIZADO** pelo fato de que compreende instruções executáveis por computador armazenadas neste para executar o método da reivindicação 12.

19. Sistema de computação que contém pelo menos um módulo de aplicação utilizado no sistema de computação, onde pelo menos um módulo de aplicação compreende código de aplicação para executar um método de determinar uma classificação de relevância de documento para um documento em uma rede, o dito método, **CARACTERIZADO** pelo fato de que compreende as etapas de:

attribution a um valor de uso atual (U_A) a um ou mais documentos em uma rede compreendendo N documentos, onde o 20 valor de uso atual (U_A) é baseado nos dados de uso real mantidos e armazenados em um servidor;

se a menos do que N documentos são atribuídos um valor de uso atual (U_A), atribuir um valor de uso padrão (U_D) aos documentos que não têm dados de uso atual associados a eles; e

usar o valor de uso para cada documento para determinar a classificação de relevância de documento de um dado documento na rede.

20. Sistema de computação, de acordo com a reivindicação 19, **CARACTERIZADO** pelo fato de que o valor de uso atual é dependente de uma ou mais propriedades relacionadas ao uso de um documento ou de uma pasta contendo um conjunto 5 de documentos, as ditas uma ou mais propriedades relacionadas ao uso compreendem um número total de visualizações de documento ou de pasta por usuários em um dado período de tempo, um número médio de visualizações de documento ou de pasta por usuário em um dado período de tempo, um tempo total 10 gasto em um documento ou pasta particular em um dado período de tempo, um tempo médio gasto em um documento ou pasta particular em um dado período de tempo, onde o dado período de tempo compreende a última semana, o último mês, o último ano, ou o tempo de vida total do documento ou pasta, ou 15 qualquer outro período de tempo.

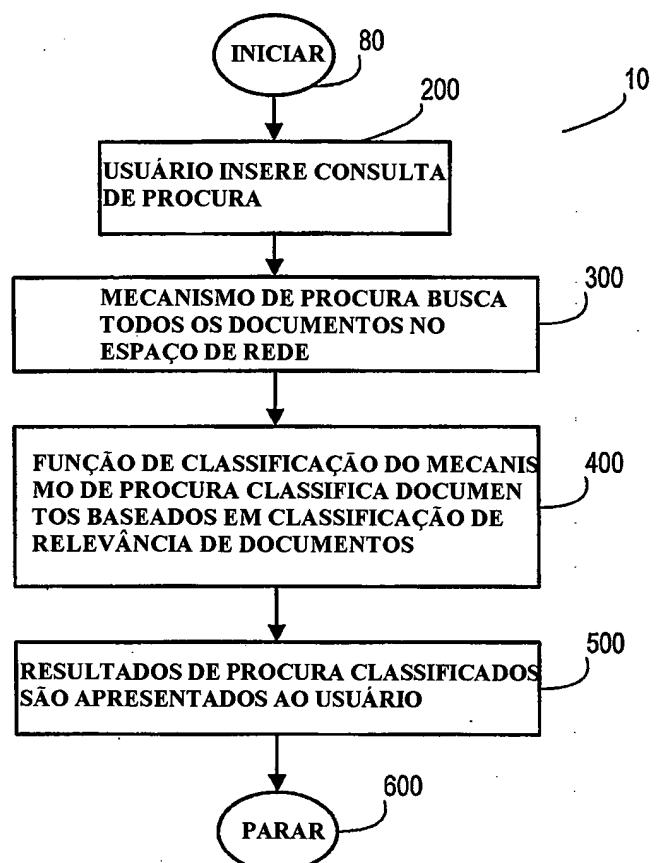


FIG. 1

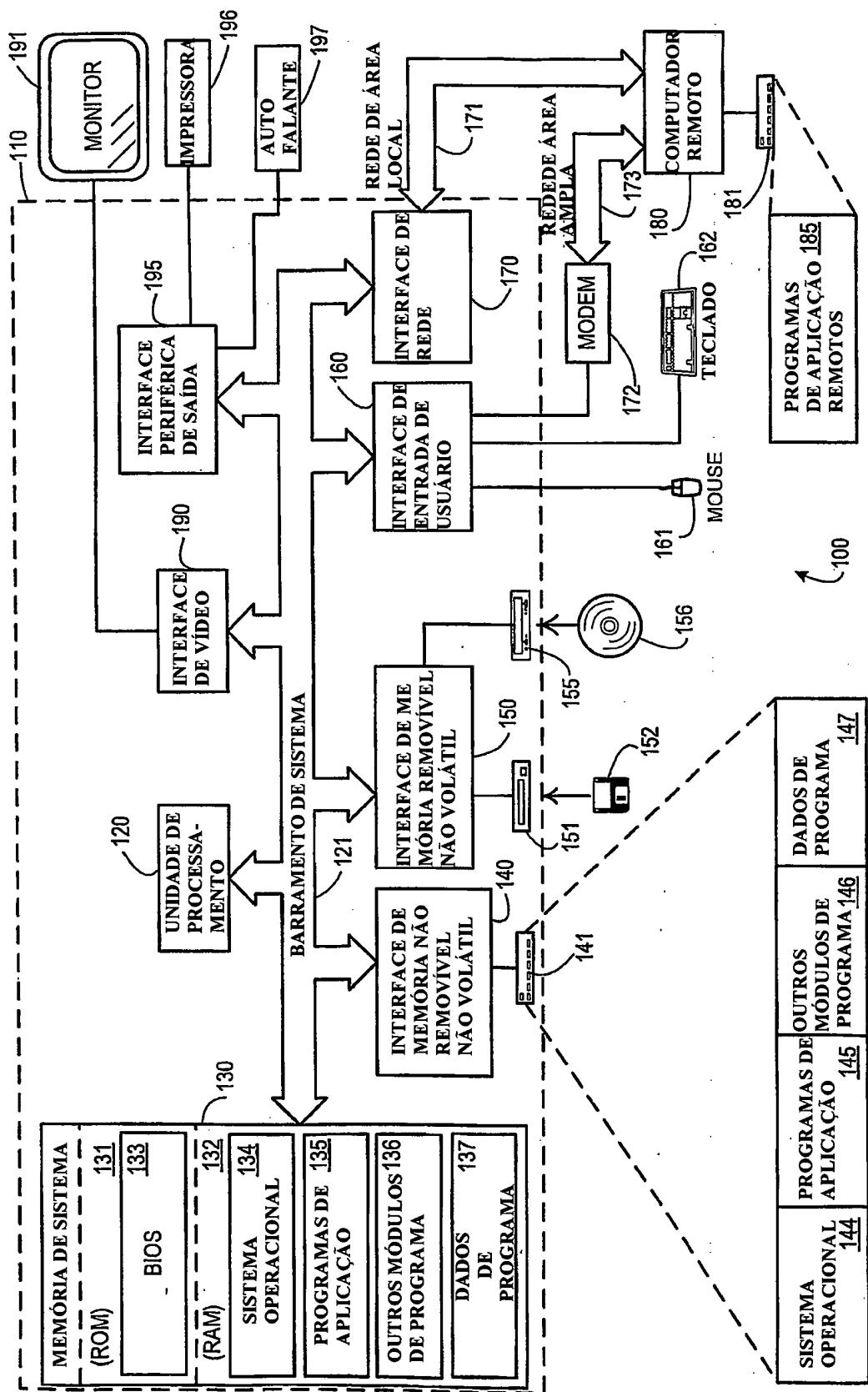


FIG. 2

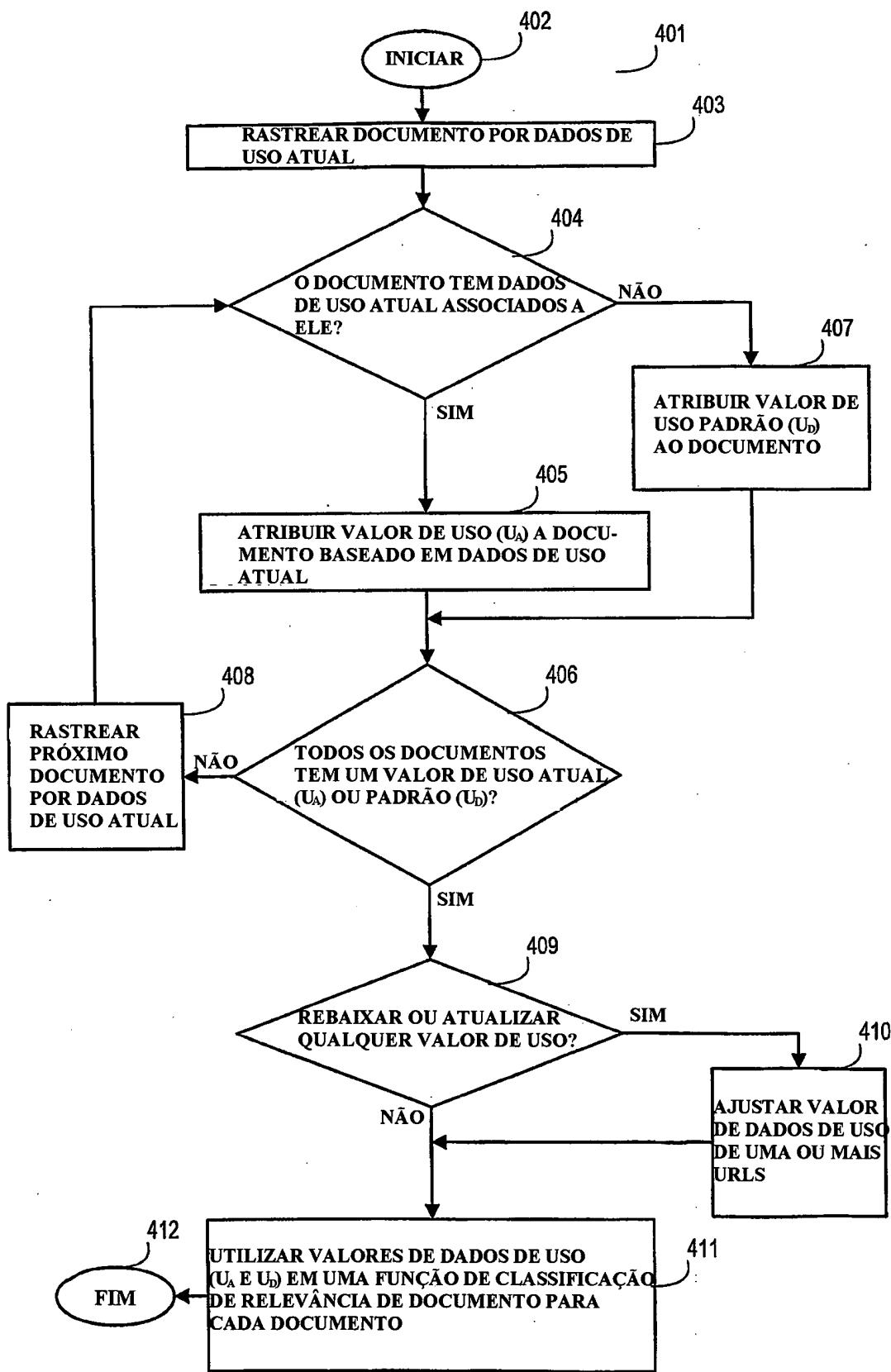


FIG. 3

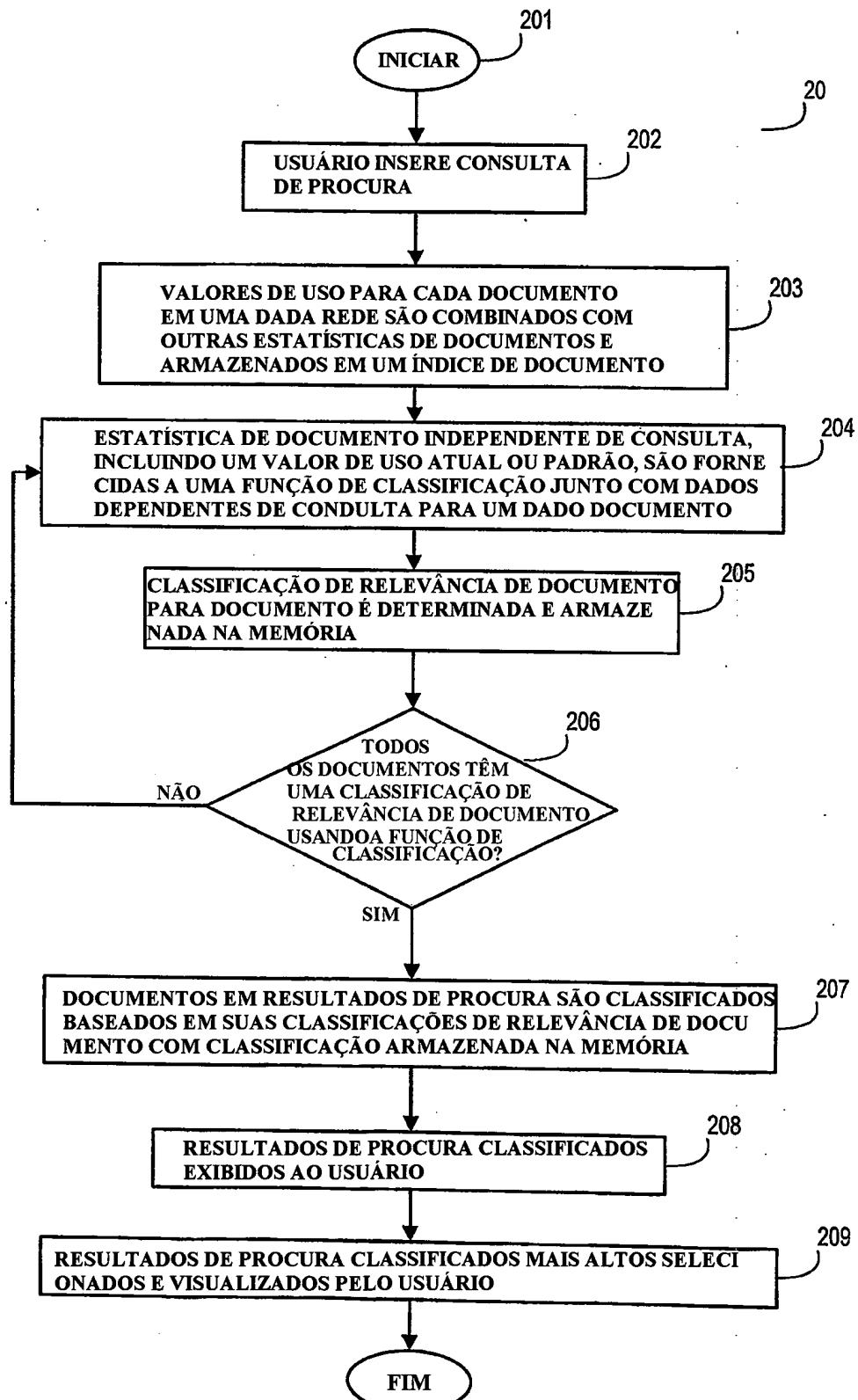


FIG. 4

RESUMO

"FUNÇÕES DE CLASSIFICAÇÃO USANDO ESTATÍSTICAS DE
USO DE DOCUMENTO"

São discutidos aqui métodos que fornecem uma classificação de relevância de documento a um documento em uma rede. O meio legível por computador que tem armazenado neste 5 instruções executáveis por computador para executar um método para fornecer uma classificação de relevância de documento a um documento em uma rede também é discutido. Ademais, 10 são discutidos sistemas de computação que contêm no mínimo um módulo de aplicação, onde no mínimo um módulo aplicativo compreende o código de aplicação para executar métodos de fornecimento de uma classificação de relevância de documento a um documento em uma rede.