



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

(21) PI 0719257-6 A2



\* B R P I 0 7 1 9 2 5 7 A 2 \*

(22) Data de Depósito: 05/12/2007  
(43) Data da Publicação: 29/04/2014  
(RPI 2260)

(51) Int.Cl.:  
G06F 17/28

(54) Título: VERIFICAÇÃO DE ERRO DE  
COMBINAÇÃO DE PALAVRAS COM BASE NA  
INTERNET.

(57) Resumo:

(30) Prioridade Unionista: 05/12/2006 US 11/633.788

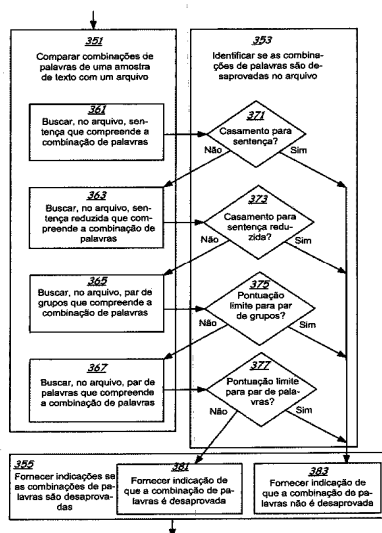
(73) Titular(es): Microsoft Corporation

(72) Inventor(es): HSIAO-WUEN HON, Jianfeng Gao, Ming Zhou,  
William B. Dolan

(74) Procurador(es): Alexandre Ferreira

(86) Pedido Internacional: PCT US2007086560 de  
05/12/2007

(87) Publicação Internacional: WO 2008/070750de  
12/06/2008



## “VERIFICAÇÃO DE ERRO DE COMBINAÇÃO DE PALAVRAS COM BASE NA INTERNET”

### Antecedentes da Invenção

5 Frequentemente, traduções de duas línguas diferentes envolvem dificuldades em casos em que o significado de palavras individuais é adicionalmente restrito pelo contexto do uso. Tipicamente, quando as pessoas aprendem uma segunda língua, que não é sua língua nativa, há muitas escolhas de palavras que parecem ter sido traduzidas corretamente, mas que, realmente, não são usadas em certos agrupamentos ou combinações de palavras no uso típico entre os falantes nativos. Tais combinações de palavras podem estar corretas sob o ponto de vista sintático ou gramatical em um sentido abstrato, mas estão fora do uso normal e, frequentemente, parecerão, intuitivamente, peculiares ou inoportunas aos falantes nativos. Frequentemente, combinações de palavras em uma língua estão de acordo com os padrões estabelecidos que formam itens léxicos exclusivos, de forma que, mesmo grupos de palavras que têm significados similares e são ordenados em um relacionamento gramatical similar, podem não cair em um item léxico no uso nativo da língua, se o item léxico for explicitamente pré-concebido pelos falantes nativos ou não como restringindo a língua dos usos alternativos.

Por exemplo, alguém que aprende inglês como uma segunda língua pode encontrar “forte” (“strong”) e “poderoso” (“powerful”) como duas traduções alternativas para uma única palavra em sua língua nativa, e “produzir” (“to make”) e “fazer” (“to do”) como duas traduções alternativas de uma palavra diferente. Sem muita experiência no uso do inglês, o aprendiz pode escrever “Eu tomei uma xícara de chá poderoso” (“I had a cup of powerful tea”) em vez de “Eu tomei uma xícara de chá forte” (“I had a cup of strong tea”), ou “Eu fiz um plano” (“I did a plan”) em vez de “Eu produzi um plano” (“I made a plan”). Erros de combinação de palavras similares podem ser típicos de um falante nativo de qualquer língua que está aprendendo qualquer outra língua na qual ele não é um falante nativo. Tipicamente, tais combinações de palavras errôneas das palavras são evitadas somente depois de um longo período de experiência contextual e prática interativa com a língua e, tipicamente, não podem ser resolvidas usando um dicionário ou qualquer outro tipo de referência de língua acessível rápida e fácil.

A discussão exposta é fornecida meramente como informação antecedente geral, e não pretende-se que seja usada como um auxílio na determinação do escopo do assunto em questão reivindicado.

### Sumário da Invenção

35 Erros de combinação de palavras podem ser automaticamente verificados usando arquivos locais e com base em rede, tal como a Internet, em uma variedade de modalidades que inclui métodos, sistemas de computação e software que codifica instruções executáveis

por sistemas de computação. Por exemplo, de acordo com um método ilustrativo, uma ou mais combinações de palavras de uma amostra de texto são comparadas com um arquivo, tal como o conteúdo da Internet. As combinações de palavras são identificadas se elas são desaprovadas no arquivo. Indicações são fornecidas por meio de um dispositivo de saída se as combinações de palavras são desaprovadas no arquivo. Então, etapas adicionais podem ser conduzidas, tais como a busca e fornecimento de combinações de palavras potencialmente apropriadas por meio de uma saída do usuário.

Este Sumário é fornecido para introduzir uma seleção de conceitos de uma forma simplificada que é descrita com detalhes a seguir na Descrição Detalhada. Não pretende-se que este Sumário identifique recursos chaves ou recursos essenciais do assunto em questão reivindicado, nem pretende-se que seja usado como um auxílio na determinação do escopo do assunto em questão reivindicado. O assunto em questão reivindicado não é limitado às implementações que resolvem todas e quaisquer desvantagens apontadas nos antecedentes.

#### Descrição Resumida dos Desenhos

A figura 1 representa uma interface de usuário para um sistema de computação que habilita um método de acordo com uma modalidade ilustrativa.

A figura 2 representa um fluxograma de um método de acordo com uma modalidade ilustrativa.

A figura 3 representa um fluxograma de um método de acordo com uma modalidade ilustrativa.

A figura 4 representa uma interface de usuário para um sistema de computação que habilita um método de acordo com uma modalidade ilustrativa.

A figura 5 representa um diagrama de blocos de um ambiente de computação no qual algumas modalidades podem ser praticadas.

A figura 6 representa um diagrama de blocos de um outro ambiente de computação no qual algumas modalidades podem ser praticadas.

#### Descrição Detalhada

A figura 1 representa uma interface de verificação de erro de combinação de palavras 10 para um sistema de computação que habilita um método para verificação de erro de combinação de palavras de acordo com uma modalidade ilustrativa. A interface de verificação de erro de combinação de palavras 10 pode ser ilustrativa das modalidades que incluem sistemas de computação, e instruções executáveis configuradas para ser executadas pelos sistemas de computação, e contextos que habilitam modalidades do método, por exemplo. A seguinte discussão fornece detalhes adicionais das várias modalidades ilustrativas. Embora certos arranjos e rótulos ilustrativos, e amostras do texto que estão sendo comparados por um método implementado por computador sejam incluídos nesta figura e nas figuras subse-

qüentes, pretende-se que eles sejam ilustrativos e indicativos da variedade e do significado mais amplo aqui fornecido na discussão e reivindicações.

A interface de verificação de erro de combinação de palavras 10 é constituída por um painel de exibição gráfica associado com uma aplicação de software, um dos muitos painéis de exibição gráfica que podem ser abertos de uma vez em uma interface gráfica de usuário em um dispositivo de saída, tal como um monitor de computador, na modalidade ilustrativa da figura 1. A interface de verificação de erro de combinação de palavras 10 é rotulada como "Verificação ESL" na barra de título 12, em referência a uma modalidade ilustrativa direcionada à verificação de amostras de texto inseridas pelos aprendizes de inglês como uma segunda língua (ESL). ESL fornece uma aplicação ideal para uma modalidade ilustrativa da verificação de erro de combinação de palavras em virtude de ter sido descoberto que combinações de palavras errôneas constituem uma grande parte dos erros de estudantes intermediários de inglês como uma segunda língua.

A modalidade ilustrativa de uma aplicação de verificação ESL é repetidamente referida na seguinte discussão, embora com o entendimento de que ela é dada a título de exemplo somente, indicativa dos princípios mais amplos que podem ser aplicados a uma extensa faixa de outras modalidades. Por exemplo, a seguir são descritas outras modalidades que podem operar sem a interface de verificação de erro de combinação de palavras 10 mostrada na figura 1, enquanto que ainda outras modalidades podem ser direcionadas a aprendizes de qualquer outra língua como uma segunda língua, e outras modalidades podem ser direcionadas a crianças que estão aprendendo dominar suas próprias línguas nativas, por exemplo.

A interface de verificação de erro de combinação de palavras 10 inclui duas abas no seu topo, que correspondem a dois diferentes modos de tarefas: aba 11, rotulada "Carregar Sentenças ESL (novamente, direcionada a uma modalidade direcionada a ESL em particular), e aba 13, rotulada "Executar Consultas". A figura 1 representa a interface de usuário 10 com a aba "Carregar Sentenças ESL" 11 selecionada, com seus botões, legendas, caixas de texto e outros elementos de interface de usuário (ou "itens de interface gráfica") associados na tela. A figura 4 representa a interface de verificação de erro de combinação de palavras 10 com a aba "Executar Consultas" 13 selecionada e os elementos de interface de usuário associados com aquela tarefa na tela, cuja discussão aparece adicionalmente a seguir.

O painel da interface de usuário sob a aba "Carregar Sentenças ESL" 11 inclui uma variedade de elementos de interface de usuário de entrada e de saída interativos, que, por sua vez, são descritos como segue. No geral, eles são divididos entre uma seção superior com a legenda de título de "Analisar Sentenças", e uma seção inferior com a legenda de título "Analisar Combinações de palavras".

Ao longo do topo da seção de “Analisar Sentenças” aparece uma linha de elementos de interface: um botão “Carregar de Arquivos Gramaticamente Analisados” 15, um botão “Carregar de Arquivos Brutos” 17, um comentário de legenda “Tipo de Combinação de palavras” em frente a uma caixa de combinação de palavras 19 em que atualmente se lê “Combinação de palavras V-N” (discutido a seguir) na representação, e um botão “Extraí Combinações de palavras” 21. Abaixo destes itens de interface gráfica fica uma caixa de texto 23 com barras de rolagem horizontal e vertical 22, 24, e uma barra de título 20 onde se lê “Sentença Original”. Abaixo disto fica uma caixa de texto 25 com a legenda “Sentença Gramaticamente Analisada”. No geral, estes elementos de interface habilitam a funcionalidade de usuário para carregar sentenças de uma amostra de texto e prepará-las para comparação com combinações de palavras no conteúdo disponível em uma rede para verificação de erro.

A caixa de texto 23 inclui inúmeras sentenças de amostra de uma amostra de texto, tal como um documento escrito por um aprendiz ESL. Da forma indicada pelo tamanho comprimido do indicador de rolagem 26 no espaço da barra de rolagem vertical 24 na caixa de texto 23, as sentenças que são atualmente exibidas são uma pequena fração da lista total de sentenças atualmente carregadas na caixa de texto que um usuário pode acessar movendo o indicador de rolagem 26. As sentenças exibidas na caixa de texto 23 representam falhas típicas que um estudante intermediário de inglês como uma segunda língua pode cometer: “Eu fiz um plano” (“I did a plan”), “Eu reconheci esta pessoa há anos” (“I have recognized this person for years”), e “Eu tomei uma xícara de chá poderoso” (“I had a cup of powerful tea”).

Estas sentenças podem ser manualmente carregadas a partir de um documento aberto em uma outra aplicação, ou elas podem ter sido automaticamente selecionadas como tendo erros tipo ESL e importadas de um outro documento ou de múltiplos documentos, depois de um usuário iniciar o processo pela seleção do botão “Carregar de Arquivos gramaticamente Analisados” 15 ou do botão “Carregar de Arquivos Brutos” 17. Em uma outra modalidade, uma aplicação diferente, tal como uma aplicação de processamento de texto, pode ter um gatilho de verificação ESL, que pode ficar ativo como padrão, o que pode criar automaticamente um painel que recomenda o uso da interface de verificação de erro de combinação de palavras 10 quando erros tipo ESL forem detectados no texto que é inserido no programa de processamento de texto.

Como sugerido pelos botões 15 e 17, amostras de texto podem ou não já ser gramaticamente analisadas quando tomadas pela aplicação. Se uma amostra de texto ainda não foi gramaticamente analisada, ela pode ser gramaticamente analisada depois do carregamento e antes de etapas adicionais ser conduzidas, por exemplo. Uma sentença gramaticamente analisada é representada na caixa de texto 25. Em particular, esta é uma versão

gramaticamente analisada da segunda sentença representada na caixa de texto 23, que tem destacamento local ao redor de si para indicar que ela está sendo selecionada por um usuário, tal como sendo clicada com o botão esquerdo de um mouse ou tocada com uma caneta luminosa, por exemplo, ou por algum outro mecanismo de entrada de usuário.

5           Análise gramatical é uma área ativa do desenvolvimento no processamento de língua natural e no entendimento de língua natural. Análise gramatical pode incluir tarefas, tais como rotulação de classe gramatical, agrupamento e classificação semântica, por exemplo. Na modalidade representada na figura 1, a sentença na caixa de texto 25 foi rotulada e agrupada em relação à classe gramatical. Outras modalidades podem incluir outras combina-  
10       ções de palavras de tarefas de análise gramatical, incluindo algumas que são tipicamente direcionadas mais no sentido de algumas línguas não ocidentais, tal como a segmentação de palavras.

Na sentença gramaticamente analisada na caixa de texto 25, cada palavra na sentença original é imediatamente seguida por uma linha inclinada (isto é, um “corte para frente”) que separa a palavra de um rótulo de classe gramatical de duas ou três letras atribuído àquela palavra. Algumas opções ilustrativas para rótulos de classe gramatical são listadas a seguir. O ponto no final da sentença também é acompanhado por sua própria linha inclinada que o separa de um outro ponto, servindo como um rótulo equivalente a um rótulo de classe gramatical para indicar a função da pontuação. As palavras da sentença, em grupos de uma  
15       ou duas (ou mais em outros exemplos), também são circundadas por colchetes para corresponder aos limites do agrupamento, com um rótulo tipo agrupamento de duas letras no início do interior de cada conjunto entre colchetes. Os rótulos de classe gramatical representados nesta sentença incluem “PRP” para um pronome pessoal, “VBP” para um verbo no tempo presente diferente da terceira pessoa do singular; “VBN” para um verbo no particípio passado; “DT” para um determinante; “NN” para um substantivo singular ou substituível comum; “IN” para uma preposição ou junção subordinada; e “NNS” para um substantivo plural comum. Uma lista ilustrativa de rótulos de classe gramatical de acordo com uma modalidade  
20       exemplar é fornecida como segue, enquanto que outros esquemas de rótulos de classe gramatical, com diferentes definições e mais ou menos especificidade, podem ser usados em outras modalidades. Alguns rótulos são ilustrados com exemplos, em inglês, para a presente modalidade exemplar.

Tabela 1: Lista ilustrativa de rótulos de classe gramatical:

1. CC – Conjunção coordenada (por exemplo, “e”, “mas”, “nem”, “ou”, “ainda”, “mais”, “menos”, “vezes”, “por” (como em dividido por))
- 35       2. CD – Número cardinal
3. DT – Determinante (inclui artigos e determinantes indefinidos, por exemplo: “um”, “uma”, “cada um”, “nenhum” (como artigo), “o” e, por exemplo, “um outro”, “qualquer um”,

“algum”, “cada”, “cada um dos”, “nenhum”, “aquele”, “estes”, “este”, “aqueles”, alguns usos de “todos”, “eles”)

4. EX – “Há” existencial
5. FW – Palavra estrangeira
- 5 6. IN – Preposição ou injunção subordinada
7. JJ - Adjetivo
8. JJR – Adjetivo, comparativo
9. JJS – Adjetivo, superlativo
10. LS – Marcador de item de lista
- 10 11. MD – Verbo modal (por exemplo, “pode”, “poderia”, “pode ser”, “é provável”, “deve”, “deverá”, “deveria”, “será”, “seria”)
12. NN – Substantivo, comum, singular ou substituível
13. NNS – Substantivo, comum, plural
14. NNP – Nome próprio, singular
- 15 15. NNPS – Nome próprio, plural
16. PDT – Pré-determinante
17. POS – Terminação possessiva
18. PRP – Pronome, pessoal
19. PRP\$ - Pronome possessivo (por exemplo: do, da, dos)
- 20 20. RB - Advérbio
21. RBR – Advérbio, comparativo
22. RBS – Advérbio, superlativo
23. RP – Partícula
24. SYM – Símbolo
- 25 25. TO – “Para”
26. UH – Exclamação
27. VB – Verbo, forma básica
28. VBD – Verbo, tempo passado
29. VBG – Verbo, gerúndio ou particípio presente
- 30 30. VBN – Verbo, particípio passado
31. VBP – Verbo, tempo passado, diferente da terceira pessoa do singular
32. VBZ – Verbo, tempo presente, terceira pessoa do singular
33. WDT – Determinante Wh
34. WP – Pronome Wh
- 35 35. WP\$ - Pronome Wh possessivo (“cujo”)
36. WRB – Advérbio Wh

A amostra de texto gramaticamente analisada representada pela sentença gramati-

camente analisada na caixa de texto 25 também foi agrupada, nesta representação, em que pares de colchetes circundam cada grupo definido, e um tipo de grupo é indicado com um rótulo no início do conteúdo de cada conjunto entre colchetes. Agrupamento é uma tarefa de análise gramatical relativamente tratável que pode ser realizada com base em informação superficial e local. Agrupamento envolve dividir sentenças em segmentos não sobrepostos de forma que cada grupo contenha uma palavra cabeça principal juntamente com palavras que se associam a ela. Portanto, o agrupamento pode ser concebido como dividindo uma sentença em frases, mas, especificamente, em unidades frasais que podem ser definidas para impedir a sobreposição de frases nesta modalidade ilustrativa. Agrupamento habilita a estrutura sintática de um texto e os relacionamentos das dependências entre as frases a se tornar identificáveis. Por exemplo, uma locução substantiva pode ser o sujeito da locução verbal, e uma segunda locução substantiva pode ser o objeto da locução verbal.

Os grupos ilustrativos da figura 1 fornecem bons exemplos disto. O primeiro grupo é classificado como NP, para uma locução substantiva, e é baseado na palavra cabeça principal isolada, “Eu” (“I”), o sujeito da sentença. O segundo grupo é classificado VP, para uma locução verbal, e inclui “ter” (“have”) e “reconhecido” (“recognized”). Os dois verbos individuais trabalham juntos para formar o conceito integrado de uma forma de particípio passado do verbo “reconhecer” (“to recognize”). O terceiro grupo é um outro NP, ou locução substantiva, que inclui uma palavra cabeça “pessoa” (“person”) juntamente com uma palavra de suporte associada para formar o conceito único de “esta pessoa” (“this person”) como o objeto da sentença. Um quarto grupo classificado PP, para uma locução preposicional, e um quinto grupo classificado NP, para uma outra locução substantiva, cada qual contendo uma única palavra. O ponto, embora classificado como um ponto, é omitido dos grupos.

O sistema de análise gramatical associado com a interface de verificação de erro de combinação de palavras 10, ou usado por ela, pode ser treinado para realizar automaticamente a rotulação de classe gramatical e o agrupamento de forma precisa e confiável, de maneiras tais como aquelas familiares aos versados na técnica do processamento de língua natural.

A seção inferior do painel da interface de usuário sob a aba “Carregar Sentenças ESL” 11, com a legenda de título de “Analisar Combinações de palavras”, inclui um conjunto de caixas de texto 27, 29, 31 e 33, cada qual representada com uma barra de título e pelo menos uma barra de rolagem. Na barra de título da caixa de texto 27 se lê “Combinações de palavras ESL”, e a caixa de texto é mostrada incluindo dois grupos colocados da sentença na caixa de texto 25, uma locução verbal para “reconheceu” (“have recognized”), e uma locução substantiva para “esta pessoa” (“this person”), o objeto da sentença. À medida em que eles aparecem, com rótulos de classe gramatical, e classificações de agrupamento e colchetes, nas frases colocadas se lê “[VP have / VBP recognized / VBN] [NP this / DT per-



son / NN]”. Esta combinação de palavras foi fornecida na caixa de texto 27 em função da caixa de combinação de palavras do tipo de combinação de palavras 19 sendo ajustada em “Combinação de palavras V-N”, uma opção que indica combinações de palavras com uma locução verbal e subsequente locução substantiva localizadas próximas uma da outra. A combinação de palavras correspondente àquele ajuste foi extraída da sentença na caixa de texto 25.

A caixa de texto 29 contém candidatas de combinação de palavras, da forma indicada nesta barra de título. As candidatas de combinação de palavras têm a mesma locução substantiva, “esta pessoa” (“this person”), que a combinação de palavras na caixa de texto 27, mas uma variedade de diferentes verbos no particípio passado precedendo a locução substantiva. Estes verbos candidatos são candidatos para substituir “reconheceu” (“recognized”) como a palavra cabeça de verbo, na forma de particípio passado, e subsequente a “ter” (“have”) em uma locução verbal. Eles incluem “conheceu” (“known”), “admitiu” (“admitted”), “pensou” (“thought”) e assim por diante. Eles também não incluem a combinação de palavras original, como uma comparação da combinação de palavras original com o conteúdo de um arquivo revelado que era escasso ou ausente com conteúdo do arquivo e, portanto, desaprovado e, provavelmente, errôneo.

Portanto, a interface da verificação de erro de combinação de palavras 10 ou uma aplicação associada com ela, ou usada com ela, podem indicar que a combinação de palavras original é desaprovada no conteúdo do arquivo e, portanto, provavelmente errônea. Esta indicação pode ser fornecida para que um usuário veja na caixa de texto 33, por exemplo, como na modalidade ilustrativa da figura 1. A caixa de texto 33 indica o estado da combinação de palavras, como sua barra de título sugere. Ela indica que esta combinação de palavras está desaprovada no conteúdo do arquivo, ou provavelmente errônea, e posta uma linha rabiscada 32 sob a palavra cabeça de verbo na combinação de palavras julgada desaprovada, que é uma candidata à substituição. Um usuário pode selecionar uma das candidatas de combinação de palavras na caixa de texto 29, tal como pelo duplo clique na combinação de palavras de substituição desejada, por exemplo, para substituir a combinação de palavras indicada como desaprovada. Esta indicação também pode ser fornecida, por exemplo, em um programa de processamento de texto separado, programa de navegação em rede ou outra aplicação, tal como pela adição de uma marcação na combinação de palavras imprópria. Uma marcação como esta pode ser destacando ou sublinhando a combinação de palavras desaprovada ou errônea, tais como com uma cor brilhante ou linha rabiscada, para atrair a atenção do usuário.

As candidatas de combinação de palavras na caixa de texto 29 são fornecidas em decorrência da comparação da combinação de palavras na caixa de texto 27 com conteúdo disponível em um grande arquivo. Isto pode incluir um arquivo local e/ou um arquivo com

base em rede que inclui conteúdo hospedado em recursos distribuídos e acessíveis em uma rede, tal como a rede mundial de computadores nesta modalidade ilustrativa. Em outras modalidades, o arquivo pode compreender o conteúdo de outras redes disponíveis, tais como a Internet, uma rede de área ampla, uma rede de área local ou algum outro tipo de rede, por exemplo. Conteúdo derivado de uma busca na Internet usando a Internet como o arquivo, pode ser usado juntamente com uma busca de um arquivo convencional, tal como o arquivo do Wall Street Journal, por exemplo. À ocorrência de uma combinação de palavras no arquivo convencional pode ser dado um peso em particular em relação às ocorrências da combinação de palavras no conteúdo da Internet, já que o arquivo deve ser relativamente garantido de uso consistentemente apropriado, comparado com o conteúdo da Internet, que pode incluir tanto uma maior variedade de fontes quanto uma inclusão de escrita mais casual em relação a um arquivo. Entretanto, o conteúdo da Internet também fornece vantagens substanciais em seu volume absoluto em relação a qualquer arquivo convencional disponível de uso de língua. A combinação de palavras apropriada correspondente a qualquer dada combinação de palavras em uma amostra de texto pode ser improvável de ocorrer em um arquivo convencional. Por outro lado, descobriu-se que, pelo menos em inglês, quase todas as combinações de palavras que podem ocorrer no uso apropriado podem ser encontradas na Internet, ou pelo menos uma combinação de palavras estruturalmente equivalente que pode ser revelada por uma busca flexível de diferentes termos de consulta com base na combinação de palavras, como discutido adicionalmente a seguir. Também descobriu-se que uma combinação de palavras de substituição apropriada pode ser encontrada na Internet para pelo menos a maior parte das combinações de palavras errôneas encontradas na forma de amostras de texto típicas, por exemplo, aprendizes de inglês como uma segunda língua. Uma sentença ou combinação de palavras em particular deixar de aparecer em uma busca na Internet dá uma alta confiança de que a sentença ou combinação de palavras é errônea. Resultados de busca podem ser avaliados para ver se alternativas similares ocorrem com muito mais frequência do que uma combinação de palavras em questão, de forma que uma possível ocorrência de uma combinação de palavras na Internet, apesar de ser errônea, seja revelada como tal.

Na época em que este documento foi escrito, a maior parte do conteúdo de texto disponível na Internet está em inglês. Muitas línguas humanas atuais têm pouco conteúdo disponível na Internet, e mesmo algumas línguas comumente faladas têm representação relativamente modesta entre o conteúdo da Internet. Hindi, por exemplo, é atualmente a terceira língua mais amplamente falada no mundo, mas descobriu-se em pelo menos um estudo ser representada por menos conteúdo na Internet do que islandês, que tem menos de um milésimo de falantes do hindi. A eficácia da verificação de erro de combinação de palavras depende, em parte, do tamanho da amostra do arquivo disponível como uma base de com-

paração para as combinações de palavras em uma amostra de texto a ser verificada. Portanto, podem ser usados métodos especializados para verificação de erro de combinação de palavras com base na Internet em diferentes línguas, que podem não ser necessários para a verificação de uma amostra de texto em inglês.

5 Estes podem incluir, por exemplo, alvejar ou triar particularmente o conteúdo em uma língua alvo que casa com a língua da amostra de texto, ou usar uma rede, arquivo ou biblioteca digitalizada especializados além da Internet ou de outro conteúdo acessado em uma rede para o arquivo usado como uma base para comparação. Estes métodos também podem incluir usar uma utilidade de busca habilitada para indexar e buscar conteúdo não  
10 codificado em Unicode, por exemplo, ou usar esquemas de codificação de caractere especializado ou formatos de transformação para poder ler, indexar e buscar conteúdo em formatos de codificação não padrão. Há indicações de que uma grande quantidade de conteúdo nas línguas que usam sistemas de escrita diferentes do alfabeto romano, mesmo uma grande maioria do conteúdo na Internet para algumas línguas, existe em formatos de codificação  
15 não padrão, de forma que uma busca habilitada a indexar e buscar aqueles formatos de codificação não padrão pode produzir muito mais conteúdo do que em outras circunstâncias. Métodos tais como estes, de acordo com várias modalidades, podem contribuir com a capacidade de um sistema de verificação de combinação de palavras com base na Internet aplicado na faixa mais ampla possível de línguas.

20 Estratégias adicionais para restringir ou influenciar o conteúdo da Internet que está sendo buscado também podem incluir aplicar preferências nas categorias ou classificações de URLs. Alguns URLs podem ser especificamente catalogados para busca preferencial, e/ou para aplicação de maior peso às combinações de palavras encontradas nos seus conteúdos, se seus conteúdos forem conhecidos por incluir grandes bibliotecas ou arquivos de  
25 uso confiável de língua apropriada. Isto também pode incluir influência preferencial de uma busca ou ponderação dos resultados da busca com base nos domínios de nível de topo (TLDs). Por exemplo, ordem de busca adicionada ou ponderação de resultado podem ser aplicadas aos URLs com um TLD de “edu” ou “gov” em oposição a “com” ou “org”. Para buscas envolvidas em verificação de combinação de palavras para línguas diferentes do  
30 inglês, buscas com triagem por conteúdo da Internet na língua alvo também podem incluir influenciar ou restringir uma busca ou o peso aplicado nos seus resultados com base nos TLDs do código de país atribuídos aos países em que a língua alvo é uma língua oficial ou tem uso substancial. Por exemplo, uma aplicação de verificação de combinação de palavras direcionada à verificação de uma amostra de texto em francês pode ser influenciada ou res-  
35 trita a buscas de conteúdo da Internet encontradas em páginas da Internet com um TLD de “fr”, “be”, “ch” ou “ca” (os TLDs do código de país para França, Bélgica, Suíça e Canadá, respectivamente), ou outros TLDs do código de país que pertencem a outras nações ofici-

almente ou substancialmente de língua francesa.

A caixa de texto 31 mostra parte de uma lista maior de todos os verbos disponíveis que são encontrados no conteúdo do arquivo que forma a palavra cabeça de uma locução verbal que precede “esta pessoa” (“this person”) em uma combinação de palavras. Da forma indicada pelo indicador de rolagem na barra de rolagem vertical no lado direito da caixa de texto 31, que é uma fração do tamanho da barra de rolagem, a lista é muito maior do que a fração de verbos disponíveis atualmente visível na caixa de texto 31 na representação da figura 1. A maior parte das combinações de palavras disponíveis listadas na caixa de texto 31 não foi selecionada como candidatas de combinação de palavras, como as listagens na caixa de texto 29, em virtude de elas não pontuarem muito alto em um esquema de pontuação com casamento impreciso, com base em um conjunto de critérios, para a provável propriedade do casamento da combinação de palavras original e sua substituição por uma combinação de palavras que fornece o significado original de uma forma correta.

Métodos ilustrativos de comparação da combinação de palavras na caixa de texto 27, de identificação se ela é desaprovada no conteúdo do arquivo, e de fornecer substituições sugeridas para ela, são descritos com detalhes adicionais em relação às figuras restantes.

A figura 2 representa um fluxograma de um método de verificação de combinação de palavras 50 de acordo com uma modalidade ilustrativa, certos aspectos do qual são análogos aos aspectos da interface de verificação de erro de combinação de palavras 10 representada na figura 1. As etapas centrais do método 50 são representadas pelas linhas cheias no fluxograma da figura 2. Estas incluem a etapa 51 de comparar uma ou mais combinações de palavras de uma amostra de texto com um arquivo, tal como pela submissão de conteúdo que contém a combinação de palavras como o termo de consulta de uma busca na Internet, a etapa 53 de identificar se as combinações de palavras são desaprovadas no arquivo, e a etapa 55 de fornecer indicações se as combinações de palavras são desaprovadas por meio de um dispositivo de saída. Como é representado com uma seta recursiva do fluxograma, as etapas 51 e 53 podem ser realizadas em sucessão repetida para realizar diferentes tipos de comparações das combinações de palavras da amostra de texto com o arquivo e identificar se as combinações de palavras são desaprovadas no arquivo, da forma indicada pelos diferentes tipos de comparações.

As etapas 51, 53 e 55 também podem envolver ou ser precedidas ou seguidas por etapas adicionais, tais como aquelas aqui representadas em linhas tracejadas. Estas incluem, por exemplo, receber uma indicação de combinações de palavras em um texto, ou identificá-las, tal como pela recepção de uma entrada de usuário ou seleção ou outra forma de indicação de uma amostra de texto, como na etapa 61. Elas também podem envolver um recurso padrão, em uma aplicação de processamento de texto ou em alguma outra aplica-

ção de software que envolve entrada humana de texto, na qual a entrada é monitorada e triada em relação a combinações de palavras errôneas ou outras indicações de uso não nativo, e etapas adicionais são disparadas se tais indicações forem encontradas, como na etapa 63. A amostra de texto também pode ser gramaticamente analisada, como na etapa 65 e como ilustrativamente discutido em relação à figura 1, embora outras modalidades do método possam ser aplicadas às amostras de texto que já estão gramaticamente analisadas, por exemplo.

Tipos em particular de combinações de palavras podem ser particularmente alvejados para comparação com o arquivo, como na etapa 67. Esta etapa inclui selecionar combinações de palavras, tais como combinações de palavras verbo – substantivo, combinações de palavras preposição – substantivo, combinações de palavras adjetivo – substantivo ou combinações de palavras verbo – advérbio, para comparação com o arquivo e realização de etapas adicionais de verificação de erro de combinação de palavras. Qualquer uma destas categorias individuais de combinação de palavras pode ser usada exclusivamente em algumas modalidades, embora qualquer um ou todos estes tipos de combinação de palavras, e/ou tipos de combinação de palavras adicionais, possam ser usados em outras modalidades. Descobriu-se que o uso de quatro tipos de combinação de palavras listados na etapa 67 cobre uma parte substancial de todos os erros de combinação de palavras cometidos por usuários típicos de língua não nativa. Selecionar estes tipos de combinação de palavras em particular envolve, primeiro, identificar as classes gramaticais na amostra de texto, tal como pelas etapas de análise gramatical preparatórias, como na etapa 65 ou como discutido em relação à figura 1.

A etapa 53, de identificar se as combinações de palavras são desaprovadas no arquivo, pode ser realizada repetidamente para uma variedade de termos de consulta individuais que são baseados na combinação de palavras em diferentes formatos, e pode tomar a forma de uma determinação final, com base em todos os termos de consulta usados, se combinações de palavras que foram comparadas com o arquivo são ou não desaprovadas no conteúdo daquele arquivo.

Uma modalidade ilustrativa é adicionalmente representada nas etapas análogas 351 e 353 no fluxograma da figura 3. Nesta modalidade, comparar combinações de palavras de uma amostra de texto com um arquivo, como na etapa 51 da figura 2, pode incluir uma estratégia de casamento impreciso que inclui buscar o arquivo por termos de consulta que incluem uma sentença completa, uma sentença reduzida, um par de grupos ou um par de palavras que incluem a combinação de palavras, da forma indicada nas etapas 361, 363, 365 e 367, respectivamente, na figura 3. Uma sentença reduzida é formada pela remoção de auxiliares de uma sentença que inclui a combinação de palavras, em que os auxiliares são palavras selecionadas da sentença que pode ser pré-definida como não exigida para

preservar a combinação de palavras na sentença em uma modalidade ilustrativa. Por exemplo, um gabarito de consulta de sentença reduzida com base na sentença com combinação de palavras errônea “Eu reconheci esta pessoa há anos” (“I have recognized this person for years”) pode ser “reconheci esta pessoa” (“have recognized this person”). Os termos de consulta do par em grupo e do par de palavras podem ser para as palavras em particular ou grupos que contêm a combinação de palavras para ficar imediatamente adjacentes um em relação ao outro, ou em uma faixa pré-selecionada de íntima proximidade na mesma sentença, tal como com não mais do que uma ou duas outras palavras os separando, por exemplo. As buscas de sentença reduzida, as buscas de par de grupo, e as buscas de par de palavra em uma busca de casamento impreciso também podem detectar resultados com uma ordem de palavra diferente, mas equivalente, ou usar espaços reservados para certas palavras de função e outros tipos de palavras, de forma que, por exemplo, “reconhecer esta pessoa” (“recognize this person”) e “reconhecer aquela pessoa” (“recognize that person”) possam ser consideradas como equivalentes para um casamento impreciso.

Uma combinação de palavras pode ser avaliada como desaprovada se termos de consulta que compreendem a combinação de palavras não pontuarem acima de uma pontuação de casamento limite pré-selecionada que indica uma presença significativa de casamentos imprecisos dos termos de consulta no arquivo. Identificar se as combinações de palavras são desaprovadas no arquivo, como na etapa 53 da figura 2, pode incluir realizar tipos diferentes de buscas para os termos de consulta, incluindo buscar um casamento de uma sentença ou de uma sentença reduzida, como nas etapas 371 e 373, respectivamente, ou avaliar uma pontuação limite para casamentos encontrados para pares de grupos ou pares de palavras, como nas etapas 375 e 377, respectivamente, na figura 3. Já que sentenças e sentenças reduzidas são muito mais específicas, um único casamento ou um pequeno número de casamentos com uma daquelas podem ser considerados suficientes para indicar que uma combinação de palavras não é desaprovada ou errônea. Casamentos com pares de grupos ou pares de palavras, desde que estes sejam menos restritos e mais genéricos, podem ter uma maior expectativa de ser encontrados em pelo menos uns poucos resultados de busca, mesmo se errôneos, em função de um resíduo relativamente escasso de conteúdo da Internet escrito por usuários imperitos da língua. Portanto, um limite mais alto pode ser ajustado para casamentos de par de grupos e de par de palavras, suficiente para indicar prevalência relativa do par de grupos ou de palavras indicado, indicativo da combinação de palavras correspondente ao uso normal, em oposição a somente uma quantidade relativamente trivial de representação no arquivo que pode ser esperada para uma combinação de palavras desaprovada, de acordo com esta modalidade.

A pontuação de casamento pode incluir um contador básico que atribui uma certa pontuação cumulativa a cada ocorrência de um casamento de um par de grupos ou de um

par de palavras, por exemplo, e somente indicar que uma combinação de palavras é apropriada em vez de desaprovada se casamentos suficientes forem encontrados para alcançar uma pontuação total mínima. Pontuações para resultados com diferentes consultas também podem ser combinadas, dando mais peso a termos de consulta mais específicos, tal como dando dez vezes mais peso por contagem de casamentos de par de grupo do que por contagem de casamentos de par de palavras, como um exemplo ilustrativo. Sentenças reduzidas ou sentenças completas também podem ser incluídas em uma avaliação de pontuação de casamento em vez de se satisfazerem por conta própria para determinar o estado da combinação de palavras. Por exemplo, a cada casamento encontrado para uma sentença reduzida que contém o par de combinação de palavras da amostra de texto pode ser atribuída uma pontuação que é vinte vezes o peso de um casamento por par de grupos e 200 vezes o peso por casamento de par de palavras, por exemplo. Qualquer que seja o método usado para avaliar casamentos ou pontuações de casamento, o processo pode culminar na etapa 355 de fornecer indicações de que as combinações de palavras são desaprovadas, como na etapa 381, ou de que elas não são desaprovadas, como na etapa 383.

A ponderação dos diferentes gabaritos de consulta e a pontuação de casamento limite usadas para distinguir combinações de palavras de uso comum de combinações de palavras desaprovadas podem ser calibradas ou sintonizadas em maior ou menor sensibilidade, seja por um desenvolvedor do software de verificação de combinação de palavras, seja pelo usuário final como uma opção de usuário. Um usuário pode desejar escalonar o limite mais baixo para ter um processo mais rápido, embora com uma maior possibilidade de combinações de palavras um pouco questionáveis ou desaprovadas em casos limites passando despercebidas. Ou um usuário pode optar por mais minúcia e ajustar o limite mais alto e, possivelmente, gastar mais tempo examinando combinações de palavras identificadas que, contudo, revela-se em uso relativamente comparável a todas as substituições em potencial, mas com uma maior garantia de eliminação de todas as combinações de palavras errôneas da amostra de texto.

Em virtude de a ordem dos termos de consulta na figura 3 ir do mais específico ao mais genérico e menos restrito, é provável que a reunião dos resultados de busca de cada etapa sucessiva dure mais tempo que para a etapa anterior. Embora seja provável que todas as etapas de retornar resultados de busca exijam somente uma fração de um segundo, o desempenho do sistema ainda pode ser tornado mais eficiente e o tempo total pode ser reduzido pela finalização dos processos das etapas 351 e 353 assim que uma indicação suficiente de um casamento for encontrada, sem atravessar as etapas restantes crescentemente demoradas. Se muitos casamentos forem encontrados para uma dada sentença ou sentença reduzida, por exemplo, gastar a fração relativamente maior do que um segundo exigida para realizar buscas de par de grupos ou de par de palavras pode ser evitado. Por-

tanto, buscas para a mesma combinação de palavras usando diferentes termos de busca podem ser perseguidas em sucessão até que um dos termos de consulta forneça resultados de busca que satisfazem um limite pré-selecionado para casar a combinação de palavras, ou até que todos os termos de consulta que compreendem a combinação de palavras sejam  
5 usados sem satisfazer o limite pré-selecionado.

Voltando às etapas adicionais representadas na figura 2, depois que são fornecidas as indicações se certas combinações de palavras são desaprovadas, uma aplicação de verificação de erro de combinação de palavras pode ir adicionalmente para buscar e fornecer palavras de substituição candidatas para verificar as combinações de palavras errôneas, da  
10 forma representada nas etapas 71, 73, 75 e 77, de acordo com uma modalidade ilustrativa. Isto pode incluir compor termos de busca com um caractere curinga que substitui uma ou mais das palavras em uma ou mais das combinações de palavras desaprovadas, como na etapa 71. Um caractere curinga pode servir como um espaço reservado aberto que é inserido em um termo de consulta no lugar de uma palavra suspeita indicada como potencialmente errônea, em que o espaço reservado pode corresponder a qualquer palavra inserida na  
15 posição da palavra suspeita no restante do termo de consulta no arquivo que está sendo buscado. Um caractere curinga pode ser usado em uma busca pela substituição de uma das palavras por um símbolo, tal como um sinal de asterisco, que substitui a palavra errônea suspeita.

Uma referência de combinação de palavras pode ser buscada por termos de consulta que fornecem candidatas de combinação de palavras em potencial, como na etapa 73, tais como aquelas com elementos de caractere curinga. Novamente, a referência de combinação de palavras pode ser a Internet ou um outro arquivo com base em rede ou local, e também pode incluir uma referência de dicionário de combinação de palavras especializado, tal como triando os resultados de busca da Internet para remover combinações de palavras  
25 não incluídas no dicionário de combinação de palavras de palavra, por exemplo.

Então, uma busca da referência de combinação de palavras pode retornar resultados com segmentos de texto que usam qualquer palavra ou frase no lugar do símbolo que representa o espaço reservado do caractere curinga. Por exemplo, novamente em relação à  
30 caixa de texto 29 para candidatas de combinação de palavras na figura 1, as diversas candidatas de combinação de palavras ali listadas podem ter sido todas retornadas de uma busca de caractere curinga com um símbolo de caractere curinga precedente à frase "esta pessoa" ("this person"). Então, o termo de busca pode ser, por exemplo, "\* esta pessoa" ("\* this person"), em que o asterisco "\*" forma o caractere curinga, e os resultados de busca  
35 incluem "conheceu esta pessoa" ("known this person"), "admitiu esta pessoa" ("admitted this person"), "pensou nesta pessoa" ("thought this person"), e assim por diante. A escolha do símbolo usado para indicar uma palavra com caractere curinga pela ferramenta de busca é



arbitrária, e deve, da melhor maneira, ser um e comercial, um ponto de interrogação ou qualquer outra coisa.

Um caractere curinga pode ser usado em cada um de uma série de termos de consulta com base na mesma combinação de palavras, indo, seqüencialmente através de uma série de mais ou menos gabaritos de consulta restritos com base na combinação de palavras, analogamente ao que foi feito originalmente para avaliar a propriedade ou desaprovação de uma combinação de palavras em questão. Por exemplo, para a sentença de combinação de palavras errônea “Eu reconheci esta pessoa há anos” (“I have recongnized this person for years”), uma série de termos de consulta com caractere curinga pode incluir um para a sentença completa, mas com um caractere curinga que substitui o verbo na combinação de palavras, isto é, “Eu \* esta pessoa há anos” (“I have \* this person for years”); um ou mais termos de consulta com base em um formato de sentença reduzida, isto é, “Eu \* esta pessoa” (“I have \* this person”), “\* esta pessoa há anos” (“have \* this person for years”), “Eu \* esta pessoa” (“I have \* this person”); um ou mais termos de consulta com base em um par de grupos que contém a combinação de palavras, isto é, “\* esta pessoa há anos” (“\* this person for years”), “\* esta pessoa” (“\* this person”); e um ou mais para a combinação de palavras como um par de palavras, buscadas como imediatamente adjacente uma em relação à outra ou em uma íntima proximidade, tal como separadas por não mais que uma palavra intermediária, por exemplo, “\* ~pessoa” (“\* ~ person”), em que o til “~” é arbitrariamente selecionado em uma modalidade ilustrativa para ser interpretado pela ferramenta de busca como um espaço reservado de proximidade, que pode representar tanto zero quanto uma palavra, ou possivelmente mais, no seu lugar, separando o caractere curinga e a palavra especificada. Portanto, usar um caractere curinga no termo de consulta fornece um exemplo ilustrativo de uma maneira efetiva de comparar combinações de palavras da próxima amostra com combinações de palavras potencialmente análogas, em uma variedade de níveis, em conteúdo disponível na Internet ou em outro arquivo com base em rede ou local.

Também pode ser usado um caractere curinga especializado que é restrito a um conjunto finito de candidatos preferenciais, ou uma referência de combinação de palavras em relação à qual os resultados são triados pode incluir uma indicação de candidatas de combinação de palavras preferidas. Por exemplo, certas candidatas de combinação de palavras podem ser preferidas com base em um processo de retradução da língua nativa do escritor da amostra de texto. O usuário pode ter uma opção de inserir sua língua nativa, ou o software de verificação de combinação de palavras pode usar uma ferramenta para avaliar pistas na amostra de escrita que podem indicar uma provável língua nativa do escritor, por exemplo, que o software pode solicitar ao usuário para confirmar. Outras pistas podem ser consideradas, tais como opções de língua que estão sendo usadas em outro software instalado no computador, ou indicações do local físico do computador, por exemplo. Por exem-

5 plo, se o software detectar que ele está sendo executado em um computador que também está executando seu sistema operacional em uma versão na língua chinesa, ou se o software detectar que o computador está conectado na Internet por meio de um servidor que uma referência indica que está fisicamente localizado em Pequim, enquanto que a amostra de texto está em inglês, ele pode atribuir tratamento preferencial ou de prioridade às candidatas de combinação de palavras que são típicas de um falante chinês.

10 Se a língua nativa do escritor for conhecida, o software de verificação de combinação de palavras pode considerar se combinações de palavras errôneas são baseadas no mapeamento impróprio da língua nativa do escritor para a língua não nativa da amostra de escrita. Isto pode incluir, por exemplo, erros de mapeamento, tais como aplicação de uma regra gramatical de língua nativa na língua não nativa, uso equivocado de um falso cognato ou de um falso amigável, omitindo uma forma plural quando a língua nativa não distinguir formas plurais, etc.

15 Por exemplo, o escritor da amostra de texto que inclui “Eu reconheci esta pessoa há anos” (“I have recognized this person for years”) pode ser indicado como um falante nativo de chinês. O software de verificação de combinação de palavras pode responder a esta indicação pela tradução desta entrada em chinês e verificando em traduções candidatas de volta em inglês. “Reconhecer” (“to recognize”) pode ser traduzido como a palavra em mandarim chinês “rènshi” (da forma produzida em Hanyu Pinyin), uma palavra que pode ser traduzida para inglês como “reconhecer” (“to recognize”) ou “conhecer” (“to know”). Portanto, uma retradução pode indicar que “conhecer” (“know”) é uma provável candidata de substituição para “reconhecer” (“recognize”), para produzir “Eu conheci esta pessoa há anos” (“I have known this person for years”). Como um outro exemplo, uma amostra de texto em inglês por um escritor indicado como falante nativo de francês pode incluir a sentença “Eu comande

20 comande a fé com cebolas” (“I commanded the faith and onions”) com tanto “comandei” (“commanded”) quanto “fé” (“faith”) rotuladas como combinações de palavras desaprovadas. Isto pode ser traduzido em francês como “J’ai commandé le foi et des oignons”, com “commandé” apropriadamente retraduzido como “pedi” (“ordered”) e “le foi” soletrada novamente como “le foie” para retradução como “o fígado” (“the liver”), de forma que o software priorize

25 ou adicione peso a estas candidatas de combinação de palavras entre todas as outras geradas pelos resultados de busca e possa sugerir candidatas de combinação de palavras de substituição para produzir a sentença “Eu pedi o fígado e cebolas” (“I ordered the liver and onions”).

35 Em ambos os casos, a retradução da língua nativa pode ter sido usada para restringir o busca de substituição candidata no primeiro lugar, ou ela pode ter sido usada como uma referência de combinação de palavras para avaliar ou priorizar entre candidatas de substituição em potencial geradas pela busca.

Depois de buscar uma referência de combinação de palavras para termos de consulta, como na etapa 73, a etapa 75 pode ser realizada para identificar resultados da busca com uma proporção relativamente alta de uma palavra candidata que substitui o caractere curinga, nesta modalidade ilustrativa. Novamente, isto pode usar diferentes ponderações para diferentes tipos de casamento entre resultados de busca que fornecem um casamento com a sentença completa, com a sentença reduzida, com o par de grupos, ou com o par de palavras que contêm a combinação de palavras em uma modalidade ilustrativa. Uma pontuação total pode ser uma soma ponderada dos componentes de pontuação de cada tipo de termo de consulta, em que cada componente de pontuação é um produto do número de resultados de busca encontrados para aquele termo de consulta vezes o peso atribuído àquele termo de consulta, por exemplo. Uma pontuação total relativamente alta usando um método como este pode ser interpretada como uma medida efetiva de uma proporção relativamente alta de uma palavra candidata que substitui o caractere curinga, e que serve como um provável potencial para verificar a combinação de palavras errônea, nesta modalidade ilustrativa.

Um caractere curinga pode ser particularmente direcionado a certas classes gramaticais ou tipos de grupo em um termo de consulta. Por exemplo, em um conjunto de termos de consulta com base em combinações de palavras verbo – substantivo, os verbos em cada uma das combinações de palavras pode ser substituída por um caractere curinga em um conjunto de termos de consulta, e os substantivos em um outro conjunto de termos de consulta. Para uma combinação de palavras preposição – substantivo, as preposições em particular podem ser parte da combinação de palavras selecionada para substituição por um caractere curinga, desde que possa ser mais provável que o substantivo esteja associado com o significado pretendido da sentença e em virtude de a preposição em particular a ser selecionada possa ser mais comumente sujeita ao uso de combinação de palavras errônea pelos aprendizes intermediários da língua. Portanto, o alvejamento da preposição para substituição por um caractere curinga nos termos de consulta com base nas combinações de palavras preposição – substantivo pode suportar uma estratégia para diminuir o tempo exigido para produzir a combinação de palavras de substituição desejada. Similarmente, o caractere curinga pode ser preferivelmente selecionado como o adjetivo em uma combinação de palavras adjetivo – substantivo, e o advérbio em uma combinação de palavras verbo – advérbio, novamente, em virtude de ser mais provável que o substantivo e o verbo possam estar ancorados no significado pretendido, enquanto que adjetivos e advérbios podem ser mais propensos à combinação de palavras errônea do que os substantivos e verbos que eles modificam. Outras classes gramaticais ou tipos de grupo em particular podem ser selecionados para substituição preferencial ou exclusiva pelos caracteres curinga em termos de consulta em outras modalidades.

Então, uma palavra de substituição em potencial que descobriu-se ter uma proporção significativamente maior do que a palavra suspeita entre os resultados de busca pode ser indicada como uma substituição sugerida para a parte aparentemente desaprovada da combinação de palavras. Isto pode ser feito depois de também fazer referência cruzada a uma combinação de palavras com a palavra de substituição em potencial para inclusão em uma referência de combinação de palavras, tais como um dicionário ou combinações de palavras apropriadas ou uma referência cruzada com retraduições sugeridas com base na língua nativa do escritor, como exposto.

Os resultados da busca, com uma ou diversas palavras candidatas de combinação de palavras em potencial, tais como aquelas que descobriu-se ter as pontuações de casamento mais altas, podem ser fornecidos por meio de uma saída de usuário como as combinações de palavras potencialmente apropriadas, como na etapa 77 da figura 2. Isto pode ser análogo às candidatas de combinação de palavras exibidas na caixa de texto 29 da figura 1, por exemplo, de maneira tal que um usuário seja habilitado a selecionar uma das candidatas de combinação de palavras para substituir a combinação de palavras indicada como errônea. Algumas vezes, uma única combinação de palavras terá uma indicação substancialmente melhor da probabilidade da substituição apropriada para a combinação de palavras errônea. Este é o caso na representação da figura 4.

A figura 4 representa uma interface de usuário 410 para um sistema de computação que habilita um método de verificação de combinação de palavras, de acordo com uma modalidade ilustrativa. A interface de usuário 410 inclui as mesmas abas no seu topo que a interface de verificação de erro de combinação de palavras 10 da figura 1. Especificamente, a aba 11 rotulada "Carregar Sentenças ESL" (novamente, direcionada de forma ilustrativa a uma modalidade ESL em particular), e a aba 13, rotulada "Executar Consultas". Embora a figura 1 represente a interface de usuário 10 com a aba "Carregar Sentenças ESL" 11 selecionada, a figura 4 representa a interface de usuário 410 com a aba 13 selecionada, com seus botões, legendas, caixas de texto e outros elementos de interface de usuário associados na tela. Estes são divididos em uma seção superior rotulada "Sentenças ESL Gramaticamente Analisadas", e em uma seção inferior, rotulada "Níveis de Consulta". A interface de usuário 410 é direcionada ao encaixe em tipos de consultas em particular e ao fornecimento de resultados por meio da interface de usuário de buscas realizadas naquelas consultas, em apoio a uma modalidade ilustrativa da verificação de erro de combinação de palavras.

A seção superior "Sentenças ESL Gramaticamente Analisadas" inclui uma caixa de texto 25 análoga à caixa de texto 25 da figura 1 e que, aqui, inclui a mesma sentença gramaticamente analisada. Esta seção também inclui uma linha 41 de botões abaixo da caixa de texto 25 da sentença gramaticamente analisada. Estes incluem botões para uma variedade de diferentes gabaritos de consulta ilustrativos para realizar buscas com base na sen-

tença gramaticamente analisada. Em particular, nesta modalidade ilustrativa, há botões rotulados “Criar Consultas”; “Submeter Consultas S I” e “Submeter Consultas S II”, em relação a diferentes consultas com base em sentença, tais como para sentenças completas ou reduzidas; “Submeter Consultas C I” e “Submeter Consultas C II”, para diferentes tipos de consultas de par de grupos; e “Submeter Consultas W”, para consultas com base no par de palavras individuais da combinação de palavras, de acordo com a discussão exposta.

A seção inferior da interface de usuário 410 inclui uma caixa de texto 43 para consultas do nível de sentença, uma caixa de texto 45 para consultas do nível do primeiro grupo e uma caixa de texto 47 para consultas de nível do segundo grupo, como exemplos ilustrativos de algumas das caixas de texto de resultado de consulta que podem ser chamadas para apresentar palavras candidatas de combinação de palavras de substituição. Cada uma das caixas de texto 43, 45 e 47 também é representada com um contador de acertos para os resultados da busca de consulta para aquela candidata de combinação de palavras, e uma caixa de texto separada à sua direita que pode exibir alguns dos resultados de busca de consulta do arquivo correspondente a cada tipo de consulta. Da forma sugerida pela figura, “conhecer” (“know”) é fortemente indicado como a palavra de substituição apropriada para “reconheci” (“recognized”) em cada um dos níveis de gabarito de consulta na combinação de palavras que está sendo verificada neste exemplo ilustrativo. Isto ilustra que o esquema enfileirado de gabaritos de consulta forneceu uma redundância que era mais do que suficiente, neste caso, para apresentar um caso convincente para um candidato de substituição em particular. Então, um usuário pode considerar o contador de acertos para cada um dos gabaritos de consulta que ele desejar, e selecionar uma das candidatas de combinação de palavras de substituição para substituir a combinação de palavras desaprovada à qual ela corresponde.

Embora as figuras 1 e 4 representem painéis de exibição gráfica associados com uma modalidade de aplicação dedicada, uma outra modalidade ilustrativa pode funcionar de uma maneira menos óbvia em comunicação com uma outra aplicação, tal como um programa de processamento de texto, um navegador da Internet, uma aplicação de correio eletrônico, um programa de apresentação ou outra aplicação que envolve entrada de texto do substancialmente ou uma outra forma de língua natural. Tal outro programa ou aplicação pode cooperar com uma modalidade do programa de verificação de erro de combinação de palavras separado, ou pode incluir um módulo próprio que compreende uma modalidade de verificação de erro de combinação de palavras, por exemplo. Uma ferramenta de verificação de erro de combinação de palavras pode ser acessível sob um menu de ferramentas na outra aplicação, por exemplo, ou pode ser acessível como uma opção em um menu instantâneo quando um usuário clicar com o botão direito do mouse em uma palavra, como um outro exemplo, em diferentes modalidades.

Em um ainda outro modo de operação, o outro programa pode executar um módulo ou programa de verificação de erro de combinação de palavras automaticamente em um ambiente padrão, e pode fornecer automaticamente indicações de combinações de palavras errôneas ou desaprovadas por meio de uma marcação aplicada nas palavras no monitor. A  
 5 marcação pode tomar a forma de uma linha azul dentada sob a combinação de palavras, por exemplo, para indicá-la como errônea ou desaprovada. Em um ainda outro modo de operação, um módulo ou programa separado por executar para detectar se uma entrada de usuário inclui indicações de que um usuário não é um usuário nativo da língua da entrada, e a indicação do uso não nativo pode disparar a ativação de um módulo ou programa de erro  
 10 de combinação de palavras, por exemplo.

A figura 5 ilustra um exemplo de um ambiente de sistema de computação adequado  
 100 no qual várias modalidades podem ser implementadas. Por exemplo, várias modalidades podem ser implementadas como aplicações de software, módulos ou outras formas de instruções que são executáveis pelo ambiente de sistema de computação 100 e que confi-  
 15 guram o ambiente do sistema de computação 100 para realizar várias tarefas ou métodos envolvidos em diferentes modalidades. Uma aplicação de software ou módulo que incorpora uma modalidade da verificação de erro de combinação de palavras pode ser desenvolvida em qualquer um de uma variedade de línguas ou ambientes de programação ou de desen-  
 volvimento de conjunto de instruções. Por exemplo, ele pode ser escrito em C#, F#, C++, C,  
 20 Pascal, Visual Basic, Java, JavaScript, Delphi, Eiffel, Nemerle, Perl, PHP, Python, Ruby, VisualFoxPro, Lua ou qualquer outra língua de programação. Também percebe-se que línguas de programação inéditas e outras formas de criação de instruções executáveis conti-  
 nuarão a ser desenvolvidas, nas quais modalidades adicionais podem ser prontamente de-  
 senvolvidas.

De acordo com uma modalidade ilustrativa, o ambiente do sistema de computação  
 25 100 pode ser configurado para realizar as tarefas de verificação de erro de combinação de palavras em resposta à recepção de uma indicação de uma combinação de palavras em um texto. Então, o ambiente do sistema de computação 100 pode realizar uma busca na Inter-  
 net para cada um de um ou mais gabaritos de consulta associados com a combinação de  
 30 palavras indicada. Vários gabaritos de consulta usados podem incluir uma sentença, uma sentença reduzida, um par de grupos, e/ou um par de palavras individuais, qualquer um dos  
 quais podendo incluir a combinação de palavras de palavra. Então, o ambiente do sistema  
 de computação 100 pode avaliar se os resultados da busca na Internet para cada um dos  
 gabaritos de consulta indicam que a combinação de palavras corresponde ao uso normal,  
 35 ou se ela é desaprovada ou indicativa de provável erro. O uso normal pode ser indicado tan-  
 to por um casamento exato do gabarito de consulta que compreende a sentença quanto por  
 uma pontuação de casamento que é maior do que um limite pré-selecionado. Então, o sis-

tema pode indicar, como parte da saída do ambiente do sistema de computação 100, por meio de um dispositivo de saída perceptível pelo usuário, em decorrência de uma modalidade de um método de verificação de erro de combinação de palavras, se a combinação de palavras corresponde ao uso normal ou se ela é desaprovada e é indicada como uso errôneo.

O ambiente do sistema de computação 100, representado na figura 5, é somente um exemplo de um ambiente de computador para executar e fornecer saída das várias modalidades, e não pretende-se que sugira alguma limitação ao escopo do uso ou da funcionalidade do assunto em questão reivindicado. Nem deve o ambiente de computação 100 ser interpretado como tendo nenhuma dependência ou exigência relacionada a nenhum dos componentes, ou combinação de palavras deles, ilustrados no ambiente operacional exemplar 100.

Modalidades são operacionais com inúmeros outros ambientes ou configurações de sistema de computação de uso geral ou de uso especial. Exemplos de sistemas, ambientes e/ou configurações de computação bem conhecidos que podem ser adequados para uso com várias modalidades incluem, mas sem limitações, computadores pessoais, computadores servidores, dispositivos de mão ou portáteis, sistemas multiprocessadores, sistemas com base em microprocessador, conversores de sinal de frequência, dispositivos eletrônicos programáveis pelo consumidor, PCs em rede, minicomputadores, computadores de grande porte, sistemas de telefonia, ambientes de computação distribuída que incluem qualquer um dos sistemas ou dispositivos expostos, e congêneres.

Modalidades podem ser descritas no contexto geral das instruções executáveis por computador, tais como módulos de programa, que são executadas por um computador. No geral, módulos de programa incluem rotinas, programas, objetos, componentes, estruturas de dados, etc. que realizam tarefas em particular ou implementam tipos de dados abstratos em particular. Algumas modalidades são projetadas para ser praticadas em ambientes de computação distribuída em que tarefas são realizadas por dispositivos de processamento remoto que são ligados por meio de uma rede de comunicações. Em um ambiente de computação distribuída, módulos de programa são localizados em mídia de armazenamento no computador tanto local quanto remota, incluindo dispositivos de armazenamento em memória. Como aqui descrito, tais instruções executáveis podem ser armazenadas em uma mídia de maneira tal que elas sejam capazes de ser lidas e executadas por um ou mais componentes de um sistema de computação, desse modo, configurando o sistema de computação com capacidades inéditas.

Em relação à figura 5, um sistema exemplar para a implementação de algumas modalidades inclui um dispositivo de computação de uso geral na forma de um computador 110. Componentes do computador 110 podem incluir, mas sem limitações, uma unidade de

processamento 120, uma memória do sistema 130 e um barramento do sistema 121 que acopla os vários componentes do sistema, incluindo a memória do sistema, na unidade de processamento 120. O barramento do sistema 121 pode se qualquer um de diversos tipos de estruturas de barramento, incluindo um barramento de memória ou um controlador de memória, um barramento periférico e um barramento local que usam qualquer uma de uma variedade de arquiteturas de barramento. A título de exemplo, e sem limitações, tais arquiteturas incluem barramento Arquitetura Padrão da Indústria (ISA), barramento Arquitetura Micro Canal (MCA, barramento ISA Melhorado (EISA), barramento Associação dos Padrões Eletrônicos de Vídeo (VESA) local, e barramento Interconexão de Componente Periférico (PCI) também conhecido como barramento *Mezzanine*.

Tipicamente, o computador 110 inclui uma variedade de mídias legíveis por computador. Mídia legível por computador pode ser qualquer mídia disponível que pode ser acessada pelo computador 110 e inclui mídia tanto volátil quanto não volátil, mídia tanto removível quanto não removível. A título de exemplo, e sem limitações, mídia legível por computador pode compreender mídia de armazenamento no computador e mídia de comunicação. Mídia de armazenamento no computador inclui mídia tanto volátil quanto não volátil, removível e não removível implementada em qualquer método ou tecnologia para armazenamento de informação, tais como instruções legíveis por computador, estruturas de dados, módulos de programa ou outros dados. Mídia de armazenamento no computador inclui, mas sem limitações, RAM, ROM, EEPROM, memória flash ou outra tecnologia de memória, CD-ROM, discos versáteis digitais (DVD) ou outro armazenamento em disco ótico, cassetes magnéticos, fita magnética, armazenamento em disco magnético ou outros dispositivos de armazenamento magnético, ou qualquer outra mídia que pode ser usada para armazenar a informação desejada e que pode ser acessada pelo computador 110. Tipicamente, mídia de comunicação incorpora instruções legíveis por computador, estruturas de dados, módulos de programa ou outros dados em um sinal de dados modulado, tais como uma onda portadora ou outro mecanismo de transporte, e inclui qualquer mídia de distribuição de informação. O termo "sinal de dados modulado" significa um sinal que tem uma ou mais de suas características ajustadas ou modificadas de uma maneira tal para codificar informação no sinal. A título de exemplo, e sem limitações, mídia de comunicação inclui mídia com fios, tais como uma rede com fios ou conexão direta com fios, e mídia sem fios, tais como acústica, RF, infravermelho e outras mídias sem fios. Combinações de palavras de qualquer um dos expostos também devem ser incluídas no escopo da mídia legível por computador.

A memória do sistema 130 inclui mídia de armazenamento no computador na forma de memória volátil e/ou não volátil, tais como memória exclusiva de leitura (ROM) 131 e memória de acesso aleatório (RAM) 132. Tipicamente, um sistema básico de entrada / saída (BIOS) 133, que contém as rotinas básicas que ajudam a transferir informação entre ele-



mentos no computador 110, tal como durante a inicialização, é armazenado na ROM 131. Tipicamente, a RAM 132 contém dados e/ou módulos de programa que são imediatamente acessíveis pela unidade de processamento 120, e/ou que estão sendo atualmente nela operados. A título de exemplo, e sem limitações, a figura 5 ilustra o sistema operacional 134, programas de aplicação 135, outros módulos de programa 136 e dados de programa 137.

O computador 110 também pode incluir outras mídias de armazenamento no computador removíveis / não removíveis, voláteis / não voláteis. A título de exemplo somente, a figura 5 ilustra uma unidade de disco rígido 141 que lê ou grava em mídia magnética não removível e não volátil, uma unidade de disco magnético 151 que lê ou grava em um disco magnético não removível e não volátil 152, e uma unidade de disco ótico 155 que lê ou grava em um disco ótico removível e não volátil 156, tais como um CD ROM ou outra mídia ótica. Outras mídias de armazenamento no computador removíveis / não removíveis, voláteis / não voláteis que podem ser usadas no ambiente operacional exemplar incluem, mas sem limitações, cassetes de fita magnética, cartões de memória flash, discos versáteis digitais, fita de vídeo digital, RAM em estado sólido, ROM em estado sólido e congêneres. Tipicamente, a unidade de disco rígido 141 é conectada no barramento do sistema 121 por meio de uma interface de memória não removível, tal como a interface 140, e a unidade de disco magnético 151 e a unidade de disco ótico 155 são tipicamente conectadas no barramento do sistema 121 por uma interface de memória removível, tal como a interface 150.

As unidades e suas mídias de armazenamento no computador associadas supradiscutidas e ilustradas na figura 5 fornecem armazenamento de instruções legíveis por computador, estruturas de dados, módulos de programa e outros dados para o computador 110. Na figura 5, por exemplo, a unidade de disco rígido 141 é ilustrada armazenando o sistema operacional 144, os programas de aplicação 145, outros módulos de programa 146 e dados de programa 147. Note que estes componentes podem ser tanto os mesmo quanto diferentes do sistema operacional 134, dos programas de aplicação 135, dos outros módulos de programa 136 e dos dados de programa 137. Aqui, ao sistema operacional 144, aos programas de aplicação 145, aos outros módulos de programa 146 e aos dados de programa 147 são dados diferentes números para ilustrar que, no mínimo, eles são cópias diferentes.

Um usuário pode inserir comandos e informação no computador 110 por meio de dispositivos de entrada, tais como um teclado 162, um microfone 163 e um dispositivo de apontamento 161, tais como um mouse, dispositivo de apontamento com esfera superior, ou plataforma sensível ao toque. Outros dispositivos de entrada (não mostrados) podem incluir uma manete, controlador de jogos, antena parabólica, digitalizador ou congêneres. Estes e outros dispositivos de entrada são freqüentemente conectados na unidade de processamento 120 por meio de uma interface de entrada de usuário 160 que é acoplada no barramento do sistema, mas podem ser conectados por outra interface e estruturas de barramento, tais

como uma porta paralela, uma porta de jogos ou um barramento serial universal (USB). Um monitor 191 ou outro tipo de dispositivo de exibição também é conectado no barramento do sistema 121 por meio de uma interface, tal como uma interface de vídeo 190. Além do monitor, computadores também podem incluir outros dispositivos de saída periféricos, tais como

5 alto-falantes 197 e impressora 196, que podem ser conectados por meio de uma interface periférica de saída 195.

O computador 110 é operado em um ambiente de rede 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 dispositivo portátil, um servidor, um roteador, um PC em rede, um dispositivo par ou outro nó de rede comum e, tipicamente, inclui

10 muitos ou todos os elementos supradescritos em relação ao computador 110. As conexões lógicas representadas na figura 5 incluem uma rede de área local (LAN) 171 e uma rede de área ampla (WAN) 173, mas também podem incluir outras redes. Tais ambientes de rede são corriqueiros em escritórios, redes de computador empresariais, intranets e a Internet.

Quando usado em um ambiente de rede LAN, o computador 110 é conectado na LAN 171 por meio de uma interface ou adaptador de rede 170. Quando usado em um ambiente de rede WAN, o computador 110 inclui tipicamente um modem 172 ou outro dispositivo para estabelecer comunicações na WAN 173, tal como a Internet. O modem 172, que pode ser interno ou externo, pode ser conectado no barramento do sistema 121 por meio da interface de entrada do usuário 160 ou de outro mecanismo apropriado. Em um ambiente de

20 rede, módulos de programa representados em relação ao computador 110, ou a partes deste, podem ser armazenados no dispositivo de armazenamento em memória remoto. A título de exemplo, e sem limitações, a figura 5 ilustra programas de aplicação remotos 185 residentes no computador remoto 180. Percebe-se que as conexões de rede mostradas são exemplares e que outros dispositivos para estabelecer uma ligação de comunicações entre

25 os computadores podem ser usados.

A figura 6 representa um diagrama de blocos de um ambiente de computação móvel geral que compreende um dispositivo de computação móvel e uma mídia legível pelo dispositivo de computação móvel e que compreende instruções executáveis que são executáveis pelo dispositivo de computação móvel, de acordo com uma outra modalidade ilustrativa. A figura 6 representa um diagrama de blocos de um sistema de computação móvel

30 que inclui o dispositivo móvel 201 de acordo com uma modalidade ilustrativa. O dispositivo móvel 201 inclui um microprocessador 202, uma memória 204, componentes de entrada / saída (I/O) 206 e uma interface de comunicação 208 para comunicação com computadores remotos ou outros dispositivos móveis. Em uma modalidade, os componentes supramencionados são acoplados para comunicação uns com os outros em um barramento adequado

35 210.

A memória 204 é implementada como memória eletrônica não volátil, tal como memória de acesso aleatório (RAM) com um módulo de segurança em bateria (não mostrado) de maneira tal que informação armazenada na memória 204 não seja perdida quando a energia geral fornecida ao dispositivo móvel 200 for desligada. Uma parte da memória 204 é  
5 ilustrativamente alocada como memória endereçável para execução do programa, enquanto que uma outra parte da memória 204 é ilustrativamente usada para armazenamento, tal como para simular o armazenamento em uma unidade de disco.

A memória 204 inclui um sistema operacional 212, programas de aplicação 214, bem como um armazenamento de objeto 216. Durante a operação, o sistema operacional  
10 212 é ilustrativamente executado pelo processador 202 a partir da memória 204. O sistema operacional 212, em uma modalidade ilustrativa, é um sistema operacional da marca WINDOWS® CE, centro de distribuição pela Microsoft Corporation. O sistema operacional 212 é ilustrativamente projetado para dispositivos móveis, e implementa recursos de base de dados que podem ser utilizados pelas aplicações 214 por meio de um conjunto de inter-  
15 faces e métodos de programação de aplicação exposto. Os objetos no armazenamento de objetos 216 são mantidos pelas aplicações 214 e sistema operacional 212, pelo menos parcialmente, em resposta às chamadas às interfaces e métodos de programação de aplicação expostos.

A interface de comunicação 208 representa inúmeros dispositivos e tecnologias que  
20 permitem que o dispositivo móvel 200 transmita e receba informação. O dispositivo inclui modems com fios e sem fios, receptores de satélite e sintonizadores de difusão, para citar alguns. O dispositivo móvel 200 também pode ser diretamente conectado em um computador para trocar dados com ele. Em tais casos, a interface de comunicação 208 pode ser um transceptor infravermelho ou uma conexão de comunicação serial ou paralela, todas as  
25 quais podem transmitir informação em fluxo contínuo.

Componentes de entrada / saída 206 incluem uma variedade de dispositivos de entrada, tais como tela sensível ao toque, botões, roletes e um microfone, bem como uma variedade de dispositivos de saída que incluem um gerador de áudio, um dispositivo vibratório e uma tela. Os dispositivos listados anteriormente são dados a título de exemplo e não precisam estar todos presentes no dispositivo móvel 200. Além do mais, outros dispositivos de  
30 entrada / saída podem ser anexados ou encontrados no dispositivo móvel 200.

O sistema de computação móvel 200 também inclui a rede 220. O dispositivo de computação móvel 201 fica ilustrativamente em comunicação sem fios com a rede 220 – que pode ser a Internet, uma rede de área ampla ou uma rede de área local, por exemplo –  
35 pela transmissão e recepção de sinais eletromagnéticos 299 de um protocolo adequado entre a interface de comunicação 208 e a interface sem fios 222. A interface sem fios 222 pode ser um concentrador sem fios ou antena celular, por exemplo, ou qualquer outra interface de

5 sinal. Por sua vez, a interface sem fios 222 fornece acesso por meio da rede 220 a um amplo arranjo de recursos de computação adicionais, ilustrativamente representados pelos recursos de computação 224 e 226. Naturalmente, qualquer número de dispositivos de computação em quaisquer locais pode ficar em conexão comunicativa com a rede 220. O dispositivo de computação 201 é habilitado a fazer uso das instruções executáveis armazenadas na mídia do componente de memória 204, tais como instruções executáveis que habilitam o dispositivo de computação 201 a realizar consultas de verificação em língua não nativa e outras tarefas, em umas poucas modalidades ilustrativas.

10 Embora o assunto em questão tenha sido descrito em língua específica aos recursos estruturais e/ou atos metodológicos, entende-se que o assunto em questão definido nas reivindicações anexas não é necessariamente limitado aos recursos ou atos específicos supradescritos. Em vez disto, os recursos e atos específicos supradescritos são divulgados como formas de exemplo da implementação das reivindicações. Como um exemplo em particular, embora os termos “computador”, “dispositivo de computação” ou “sistema de compu-  
15 tação” possam ser aqui usados algumas vezes somente por conveniência, entende-se que cada um destes pode dizer respeito a qualquer dispositivo de computação, sistema de computação, ambiente de computação, dispositivo móvel ou outro componente ou contexto de processamento de informação, e não é limitado a nenhuma interpretação individual. Como um outro exemplo em particular, embora muitas modalidades sejam apresentadas com elementos ilustrativos que são amplamente familiares no momento do depósito do pedido de  
20 patente, percebe-se que muitas inovações na tecnologia de computação afetarão elementos de diferentes modalidades em aspectos tais como interfaces de usuário, métodos de entrada do usuário, ambientes de computação e métodos de computação, e que elementos definidos pelas reivindicações podem ser incorporados de acordo com estes e outros avanços  
25 inovadores, embora ainda permanecendo consistentes com os elementos definidos pelas reivindicações aqui apresentadas, e abrangidos por elas.

## REIVINDICAÇÕES

1. Método (por exemplo, 50) implementável por um sistema de computação (por exemplo, 100, 110, 180, 200, 224, 226), **CARACTERIZADO** pelo fato de que compreende:

5       comparar uma ou mais combinações de palavras de uma amostra de texto (por exemplo, 23) com um arquivo (por exemplo, 51, 351);

      identificar se as combinações de palavras são desaprovadas no arquivo (por exemplo, 53, 353); e

      fornecer indicações (por exemplo, 32, 33) se as combinações de palavras são desaprovadas por meio de um dispositivo de saída (por exemplo, 33, 32, 43, 45, 47, 55, 355).

10       2. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que o arquivo compreende conteúdo da Internet (por exemplo, 171, 173, 18, 185, 220, 224, 226).

      3. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que uma combinação de palavras é desaprovada tanto se ela não for encontrada no arquivo (por exemplo, 371, 373) quanto se ela não pontuar acima de uma pontuação de casamento limite  
15       pré-selecionada que indica uma presença significativa de casamentos imprecisos da combinação de palavras no arquivo (por exemplo, 375, 377).

      4. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que o arquivo compreende conteúdo acessível em uma rede (por exemplo, 171, 173, 18, 185, 220, 224, 226).

20       5. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que comparar as combinações de palavras com o arquivo compreende realizar uma ou mais buscas na Internet usando um ou mais termos de consulta que compreendem cada uma das uma ou mais combinações de palavras (por exemplo, 361, 363, 365, 367).

      6. Método, de acordo com a reivindicação 5, **CARACTERIZADO** pelo fato de que  
25       para cada uma das uma ou mais combinações de palavras para as quais buscas são realizadas, uma busca é realizada para cada um dos um ou mais termos de consulta que compreendem a combinação de palavras até que cada um dos termos de consulta forneça resultados de busca que satisfazem um limite pré-selecionado para o casamento da combinação de palavras, ou até que todos os termos de consulta que compreendem a combinação de  
30       palavras sejam usados sem satisfazer o limite pré-selecionado (por exemplo, 351, 353).

      7. Método, de acordo com a reivindicação 6, **CARACTERIZADO** pelo fato de que, para cada uma das combinações de palavras para as quais as buscas são realizadas, a combinação de palavras é desaprovada se os termos de consulta que compreendem a combinação de palavras não pontuarem acima de uma pontuação de casamento limite pré-  
35       selecionada que indica uma presença significativa de casamentos imprecisos dos termos de consulta no arquivo (por exemplo, 375, 377).

      8. Método, de acordo com a reivindicação 6, **CARACTERIZADO** pelo fato de que

pelo menos um dos termos de consulta compreende uma sentença que compreende a combinação de palavras (por exemplo, 361).

5 9. Método, de acordo com a reivindicação 6, **CHARACTERIZADO** pelo fato de que pelo menos um dos termos de consulta compreende uma sentença reduzida que compreende a combinação de palavras, em que a sentença reduzida é formada pela remoção de auxiliares de uma sentença que compreende a combinação de palavras (por exemplo, 363).

10 10. Método, de acordo com a reivindicação 6, **CHARACTERIZADO** pelo fato de que pelo menos um dos termos de consulta compreende um par de grupos que compreende a combinação de palavras (por exemplo, 365).

11. Método, de acordo com a reivindicação 6, **CHARACTERIZADO** pelo fato de que pelo menos um dos termos de consulta compreende um par de palavras que compreende a combinação de palavras (por exemplo, 367).

12. Método, de acordo com a reivindicação 1, **CHARACTERIZADO** pelo fato de que compreende adicionalmente rotular a classe gramatical da amostra de texto (por exemplo, 25, 65), e em que adicionalmente as combinações de palavras que são comparadas com o arquivo são selecionadas da amostra de texto para compreender pelo menos uma de uma combinação de palavras verbo – substantivo, de uma combinação de palavras preposição – substantivo, de uma combinação de palavras adjetivo – substantivo, e de uma combinação de palavras verbo – advérbio (por exemplo, 19, 67).

13. Método, de acordo com a reivindicação 1, **CHARACTERIZADO** pelo fato de que a amostra de texto está em uma língua alvo, e o método compreende adicionalmente triar o arquivo para comparar as combinações de palavras somente com conteúdo indicado como estando na língua alvo.

14. Método, de acordo com a reivindicação 1, **CHARACTERIZADO** pelo fato de que compreende adicionalmente triar a amostra de texto para determinar se ela é indicada mostrando uso não nativo (por exemplo, 63), e se a amostra de texto é indicada mostrando uso não nativo, então, iniciar automaticamente a etapa de comparar as combinações de palavras da amostra de texto com o arquivo.

15. Método, de acordo com a reivindicação 1, **CHARACTERIZADO** pelo fato de que compreende adicionalmente:

compor um ou mais termos de consulta com um caractere curinga substituindo uma das palavras em uma das combinações de palavras desaprovadas (por exemplo, 71);

buscar uma referência de combinação de palavras para os termos de consulta (por exemplo, 73);

35 identificar resultados da busca com uma proporção relativamente alta de uma palavra candidata que substitui o caractere curinga (por exemplo, 75); e

fornecer os resultados da busca com a palavra candidata por meio do dispositivo de

saída como as combinações de palavras potencialmente apropriadas (por exemplo, 77).

16. Método, de acordo com a reivindicação 15, **CARACTERIZADO** pelo fato de que os termos de consulta compreendem um ou mais de um gabarito de sentença (por exemplo, 361, 363), um gabarito de grupo (por exemplo, 365), e um gabarito de palavra (por exemplo, 367), e em que os resultados da busca com uma proporção relativamente alta da palavra candidata que substitui o caractere curinga são avaliados pela multiplicação de uma proporção dos resultados de busca que compreendem a palavra candidata por um peso de gabarito de consulta que é o mais alto para o gabarito de sentença, mais baixo para o gabarito de grupo, e o mais baixo para o gabarito de palavra.

17. Método, de acordo com a reivindicação 15, **CARACTERIZADO** pelo fato de que os termos de consulta compreendem uma ou mais de uma combinação de palavras verbo – substantivo, de uma combinação de palavras preposição – substantivo, de uma combinação de palavras adjetivo – substantivo e de uma combinação de palavras verbo – advérbio, e o caractere curinga é selecionado como o verbo em uma combinação de palavras verbo – substantivo, como o substantivo em uma combinação de palavras verbo – substantivo, como a preposição em uma combinação de palavras preposição – substantivo, como o adjetivo em uma combinação de palavras adjetivo – substantivo, ou como o advérbio em uma combinação de palavras verbo – advérbio (por exemplo, 19, 67).

18. Método, de acordo com a reivindicação 15, **CARACTERIZADO** pelo fato de que compreende adicionalmente habilitar um usuário a selecionar uma das combinações de palavras potencialmente apropriadas para substituir a combinação de palavras desaprovada à qual ela corresponde.

19. Mídia (por exemplo, 131, 132, 152, 156, 204), compreendendo instruções executáveis por um sistema de computação (por exemplo, 100, 110, 180, 200, 224, 226), **CARACTERIZADA** pelo fato de que as instruções configuram o sistema de computação para:

receber uma indicação de uma combinação de palavras em um texto (por exemplo, 15, 17, 61, 63);

realizar uma busca na Internet para cada um dos um ou mais gabaritos de consulta associados com a combinação de palavras indicada (por exemplo, 51, 351), em que um dos gabaritos de consulta compreende uma sentença na qual a combinação de palavras foi encontrada (por exemplo, 361, 363), um dos gabaritos de consulta compreende um par de grupos que compreende a combinação de palavras (por exemplo, 365), e um dos gabaritos de consulta compreende um par de palavras individuais que compreende a combinação de palavras (por exemplo, 367);

avaliar se os resultados da busca na Internet para cada um dos um ou mais gabaritos de consulta indicam que a combinação de palavras corresponde ao uso normal, da for-

ma indicada tanto por um casamento exato do gabarito de consulta que compreende a sentença quanto por uma pontuação de casamento que é maior do que um limite pré-selecionado (por exemplo, 53, 353); e

5       indicar, por meio de um dispositivo de saída perceptível pelo usuário (por exemplo, 191, 196, 197, 206) se a combinação de palavras corresponde ao uso normal (por exemplo, 33, 32, 43, 45, 47, 55, 355).

20. Sistema de computação (por exemplo, 100, 110, 180, 200, 224, 226),

**CARACTERIZADO** pelo fato de que é configurado para:

identificar combinações de palavras em um texto (por exemplo, 51, 351);

10       buscar a Internet por um conjunto de gabaritos de consulta com base em cada uma das combinações de palavras (por exemplo, 53, 353); e

indicar, por meio de um dispositivo de saída do usuário, se os resultados da busca indicam que as combinações de palavras são relativamente escassas na Internet (por exemplo, 33, 32, 43, 45, 47, 55, 355).



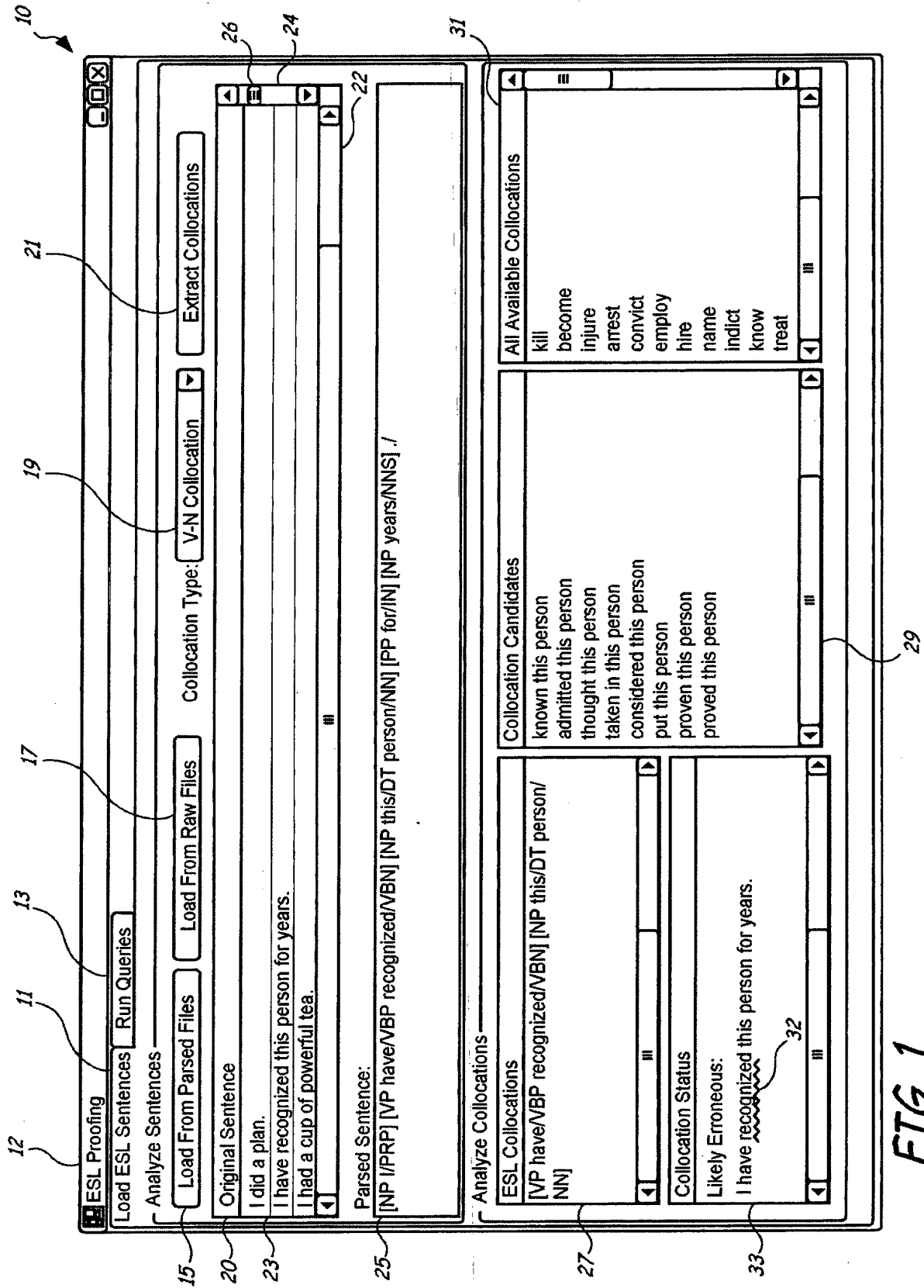


FIG. 1

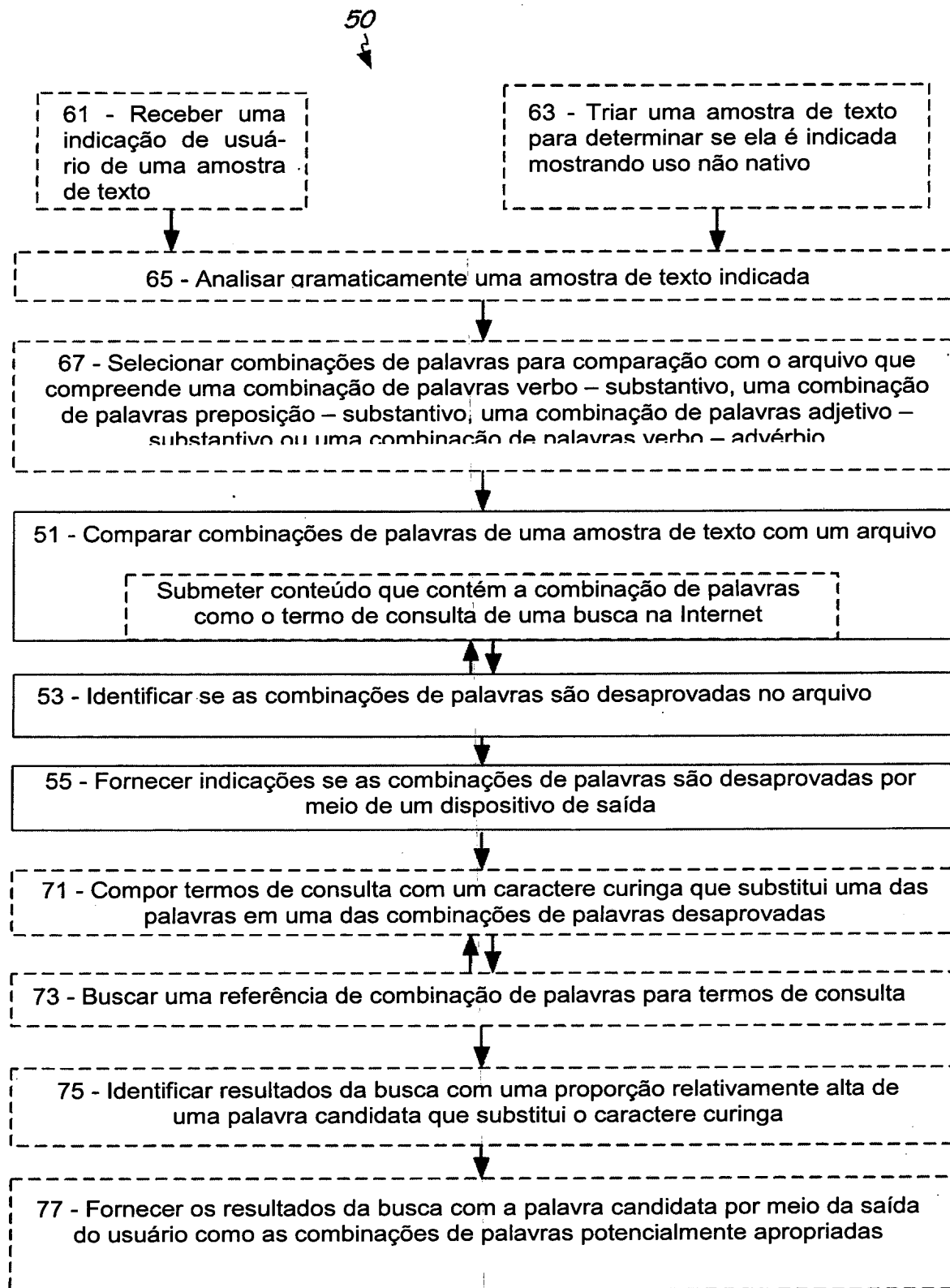


FIG. 2

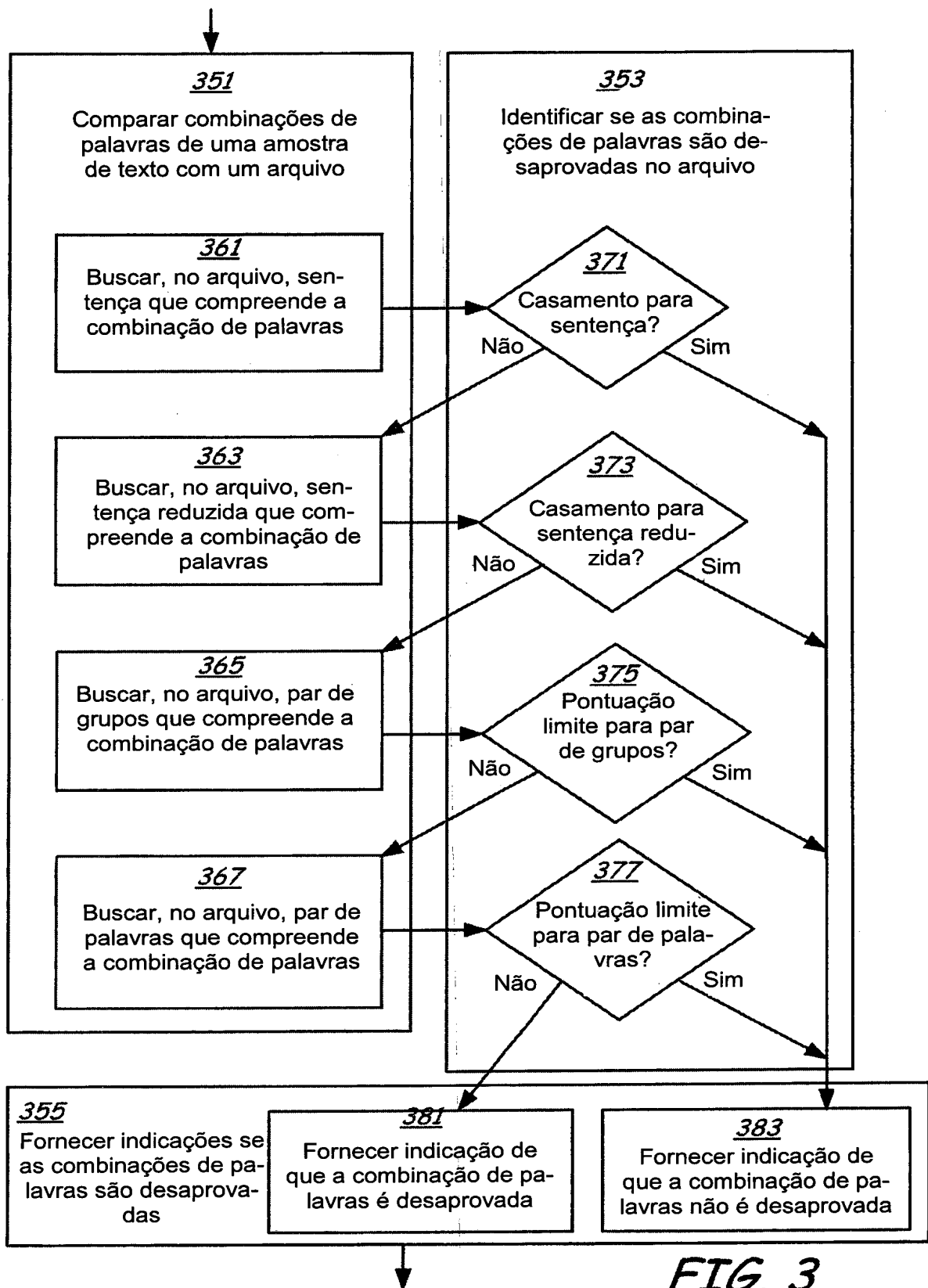


FIG. 3

ESL Proofing

Load ESL Sentences

Run Queries

ESL Parsed Sentences

[NP /PRP] [VP have/VBP recognized/VBN] [NP this/DT person/NN] [PP for/IN] [NP years/NNS] ./.

Create Queries

Submit S-Queries I

Submit S-Queries II

Submit C-Queries I

Submit C-Queries II

Submit W-Queries

Levels of Queries

Sentence Level Queries

	Counts
"known this person for years"	145
"proven this person for years"	0
"put this person for years"	0
"recognized this person for years"	0
"proved this person for years"	0
"thought this person for years"	0
"admitted this person for years"	0
"considered this person for years"	0

Chunk First Level Queries

"this person for years"	
"this person" AND "for years"	
"this person" AND "years"	

Suggested Collocations

"known"	32
---------	----

Query Search Results from Corpus Content

Excellent Patents

Thomas Edison told the patent examiner, "I have known this person for years, and he is also a clever inventor..."  
<http://www.uspto.gov/notareaisearchresult.html>

Albert Einstein

When Einstein applied for a job there, the Swiss Patent Office checked his credentials with Hermann Minkowski, who told them, "I have known this person for years, and he is a competent student..."  
<http://www.uspto.gov/illustrativeexample.html>

Query Search Results from Corpus Content

New CEO

...the chairman has known this person for years, and analysts see the transition as ensuring continuity...  
<http://www.uspto.gov/searchresultexample.html>

Another Search Result Title...

Query Search Results from Corpus Content

Personnel

References: How long have you known this person? Years:  
<http://www.uspto.gov/anotherexample.html>

Lower Frequency Search Result Example

...the Commissioner thought this person had years of relevant experience dealing with patents on new inventions...  
<http://www.uspto.gov/illustrativeexample.html>

Document Title

...search result from document body...  
[http://www.uspto.gov/document\\_url.html](http://www.uspto.gov/document_url.html)

Chunk Second Level Queries

	Counts
"known this person" AND "years"	4,631
"put this person" AND "years"	2,583
"thought this person" AND "years"	1,888
"considered this person" AND "years"	372
"recognized this person" AND "years"	185
"proved this person" AND "years"	17
"proven this person" AND "years"	14
"admitted this person" AND "years"	12
"taken in this person" AND "years"	2

FIG. 4

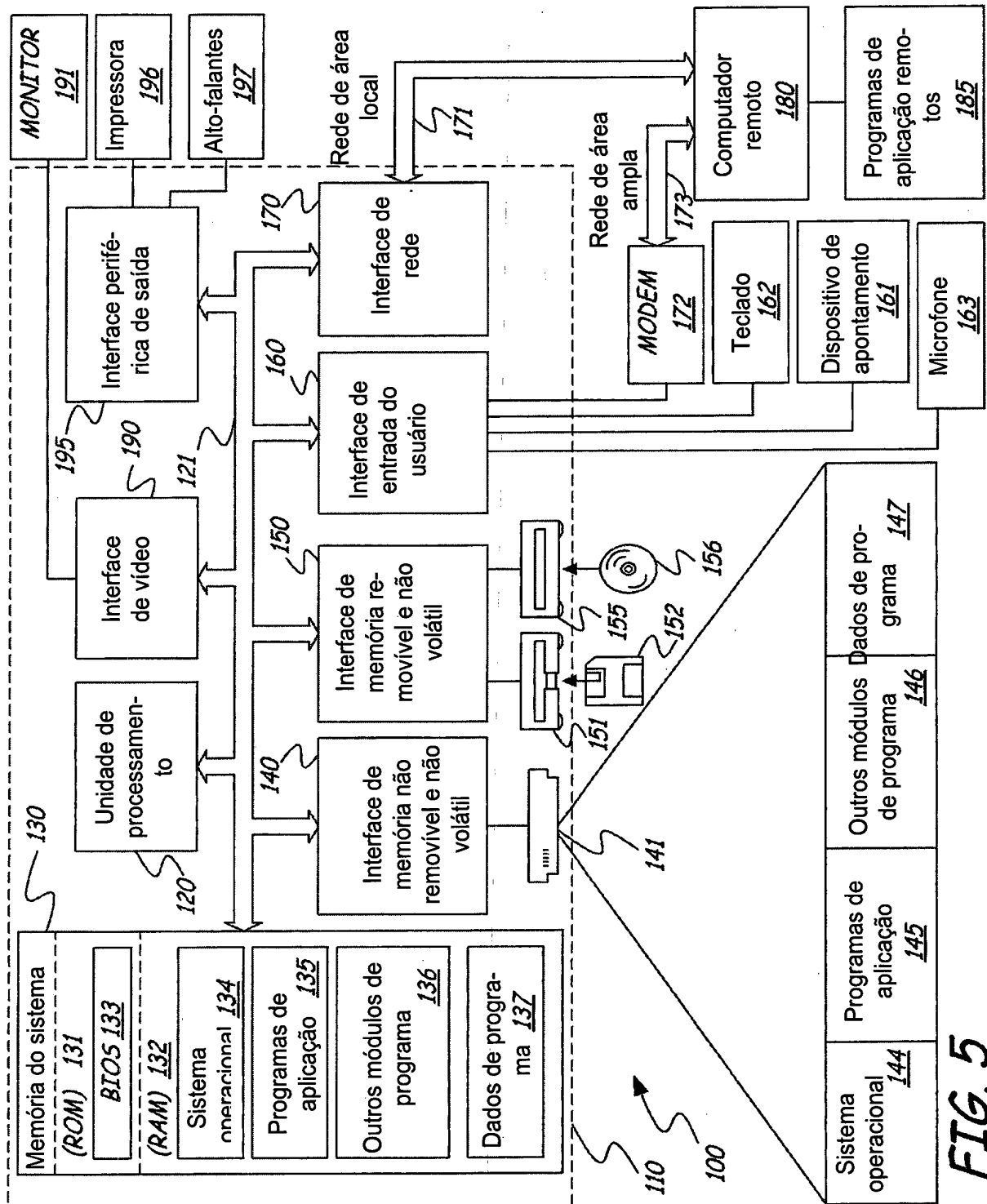
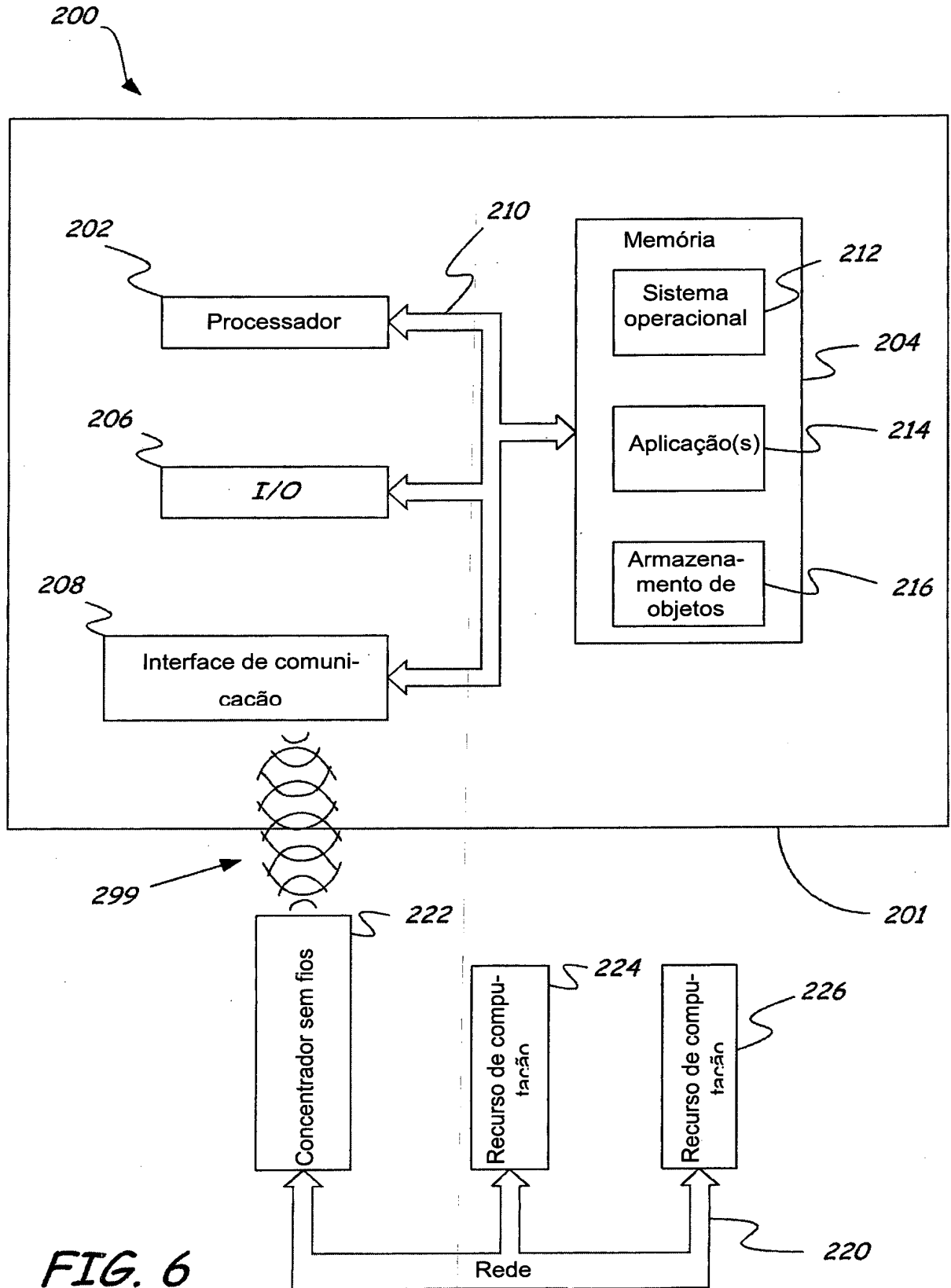


FIG. 5



## RESUMO

### "VERIFICAÇÃO DE ERRO DE COMBINAÇÃO DE PALAVRAS COM BASE NA INTERNET"

Erros de combinação de palavras podem ser automaticamente verificados usando arquivos locais e com base em rede, incluindo a Internet. Por exemplo, de acordo com um método ilustrativo, uma ou mais combinações de palavras de uma amostra de texto são comparadas com um arquivo, tal como o conteúdo da Internet. As combinações de palavras são identificadas se elas são desaprovadas no arquivo. São fornecidas indicações por meio de um dispositivo de saída se as combinações de palavras são desaprovadas no corpo. Então, etapas adicionais podem ser conduzidas, tais como busca e fornecimento de combinações de palavras potencialmente apropriadas por meio de uma saída de usuário.