

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5439028号
(P5439028)

(45) 発行日 平成26年3月12日(2014.3.12)

(24) 登録日 平成25年12月20日(2013.12.20)

(51) Int.Cl.

F I

G O 6 F 17/30 (2006.01)

G O 6 F 17/30 3 3 O C

G O 6 F 3/0482 (2013.01)

G O 6 F 17/30 3 5 O C

G O 6 F 3/048 6 5 4 B

請求項の数 8 (全 40 頁)

(21) 出願番号 特願2009-116025 (P2009-116025)
 (22) 出願日 平成21年5月12日(2009.5.12)
 (65) 公開番号 特開2010-266970 (P2010-266970A)
 (43) 公開日 平成22年11月25日(2010.11.25)
 審査請求日 平成24年4月25日(2012.4.25)

(73) 特許権者 000102728
 株式会社エヌ・ティ・ティ・データ
 東京都江東区豊洲三丁目3番3号
 (73) 特許権者 509132738
 合同会社シンタックス
 東京都足立区新田三丁目36番2-506号
 (73) 特許権者 513253917
 株式会社NTTデータ・スマートソーシ
 グ
 東京都新宿区西新宿六丁目5番1号
 (74) 代理人 110001634
 特許業務法人 志賀国際特許事務所

最終頁に続く

(54) 【発明の名称】 情報検索装置、情報検索方法、およびプログラム

(57) 【特許請求の範囲】

【請求項1】

複数の単語から構成される検索キー文が入力される入力部と、

前記検索キー文を解析して、前記検索キー文を構成する前記単語に関する解析結果を得る解析部と、

少なくとも1つの前記単語によって構成される文節を部分木ノードとして木構造に構成した辞書であって、少なくとも1つの前記文節によって構成される文に関するマッチ辞書情報として、当該文に含まれる前記文節に関する情報を表すルール情報を記憶するマッチ辞書記憶部と、

前記マッチ辞書記憶部に記憶されている前記マッチ辞書情報と前記検索キー文との関係性を照合するためのマッチング条件が関連付けられており、当該マッチング条件を満たす単語に対して前記検索キー文との照合の度合いを評価するための評価基準を有するマッチプロファイル情報を記憶するマッチプロファイル記憶部と、

前記マッチプロファイル情報に基づき、関連付けられている前記マッチング条件に従って、前記検索キー文と前記マッチ辞書情報との照合を行い、照合の結果、前記マッチング条件を満たす前記文に対して、当該マッチプロファイル情報に関連付けられている前記評価基準に従って、前記検索キー文と前記マッチ辞書情報との照合の度合いを表すスコアを算出する検索処理部と、

を備えることを特徴とする情報検索装置。

【請求項2】

10

20

前記評価基準は、

前記マッチング条件を満たす単語に対して前記照合の度合いに応じた点数を与えるか否かを表し、

前記検索処理部は、

前記評価基準に従って、前記マッチング条件を満たす単語に対して与えられる前記点数を、前記マッチング条件を満たす文毎に算出して、前記スコアを得ることを特徴とする請求項 1 に記載の情報検索装置。

【請求項 3】

前記マッチプロファイル記憶部は、

それぞれ異なる特性を有する複数の前記マッチング条件のうち、少なくとも 1 つのマッチング条件と関連付けられている前記マッチプロファイル情報を、検索の目的に応じて複数備えることを特徴とする請求項 1 あるいは 2 に記載の情報検索装置。

10

【請求項 4】

前記マッチプロファイル記憶部は、

前記マッチング条件として、単語要素マッチング、属性マッチング、あるいは係り受けマッチングのうち、少なくとも 1 つが関連づけられていることを特徴とする請求項 1 から 3 のうちいずれか一項に記載の情報検索装置。

【請求項 5】

前記入力部は、複数の単語から構成される検索対象文を入力し、

前記解析部は、前記検索対象文を解析して、前記検索対象文を構成する前記単語に関する解析結果を得て、

20

前記解析結果に基づき、前記単語の文字列に関する文字情報、および前記単語の属性を表す属性情報を含む前記ルール情報を、少なくとも 1 つの前記単語によって構成される文節と対応付けて、部分木ノードとして木構造に構成した辞書情報であって、少なくとも 1 つの前記文節によって構成される文に関する前記マッチ辞書情報を作成して前記マッチ辞書記憶部に記憶させる辞書作成部を、

さらに備えることを特徴とする請求項 1 から 4 のうちいずれか一項に記載の情報検索装置。

【請求項 6】

入力部が、

30

複数の単語から構成される検索キー文の入力を受け付け、

解析部が、

前記検索キー文を解析して、前記検索キー文を構成する前記単語に関する解析結果を得て、

検索処理部が、

前記マッチ辞書情報と前記検索キー文との関係性を照合するためのマッチング条件が関連付けられており、当該マッチング条件を満たす単語に対して前記検索キー文との照合の度合いを評価するための評価基準を有するマッチプロファイル情報を記憶するマッチプロファイル記憶部から前記マッチプロファイル情報を読み出し、

少なくとも 1 つの前記単語によって構成される文節を部分木ノードとして木構造に構成した辞書であって、少なくとも 1 つの前記文節によって構成される文に関するマッチ辞書情報として、当該文に含まれる前記文節に関する情報を表すルール情報を記憶するマッチ辞書記憶部の前記マッチ辞書情報を利用して、前記マッチプロファイル情報に基づき、関連付けられている前記マッチング条件に応じた前記検索キー文と前記マッチ辞書情報との照合を行い、

40

照合の結果、前記マッチング条件を満たす前記文に対して、当該マッチプロファイル情報に関連付けられている前記評価基準に従って、前記検索キー文と前記マッチ辞書情報との照合の度合いを表すスコアを算出する

ことを特徴とする情報検索方法。

【請求項 7】

50

コンピュータを、

複数の単語から構成される検索キー文が入力される入力手段、

前記検索キー文を解析して、前記検索キー文を構成する前記単語に関する解析結果を得る解析手段、

前記マッチ辞書情報と前記検索キー文との関係性を照合するためのマッチング条件が関連付けられており、当該マッチング条件を満たす単語に対して前記検索キー文との照合の度合いを評価するための評価基準を有するマッチプロファイル情報を記憶するマッチプロファイル記憶部から前記マッチプロファイル情報を読み出し、少なくとも1つの前記単語によって構成される文節を部分木ノードとして木構造に構成した辞書であって、少なくとも1つの前記文節によって構成される文に関するマッチ辞書情報として、当該文に含まれる前記文節に関する情報を表すルール情報を記憶するマッチ辞書記憶部の前記マッチ辞書情報を利用して、前記マッチプロファイル情報に基づき、関連付けられている前記マッチング条件に応じた前記検索キー文と前記マッチ辞書情報との照合を行い、照合の結果、前記マッチング条件を満たす前記文に対して、当該マッチプロファイル情報に関連付けられている前記評価基準に従って、前記検索キー文と前記マッチ辞書情報との照合の度合いを表すスコアを算出する検索処理手段として機能させるためのプログラム。

10

【請求項8】

前記入力手段は、複数の単語から構成される検索対象文を入力し、

前記解析手段は、前記検索対象文を解析して、前記検索対象文を構成する前記単語に関する解析結果を得て、

20

前記コンピュータを、さらに

前記解析結果に基づき、前記単語の文字列に関する文字情報、および前記単語の属性を表す属性情報を含む前記ルール情報を、少なくとも1つの前記単語によって構成される文節と対応付けて、部分木ノードとして木構造に構成した辞書情報であって、少なくとも1つの前記文節によって構成される文に関する前記マッチ辞書情報を作成して前記マッチ辞書記憶部に記憶させる辞書作成手段

として機能させるための請求項7に記載のプログラム。

【発明の詳細な説明】

【技術分野】

30

【0001】

本発明は、入力されたテキストデータを解析した結果に応じた検索を行う情報検索装置、情報検索方法、およびプログラムに関する。

【背景技術】

【0002】

例えば、電子メールによるユーザの意見や企業内の電子文書を収集し、テキストデータに変換してデータベース等に蓄積しておく。そして、オペレーターによるユーザからの問い合わせ対応や苦情対応の際に、オペレーターが入力した文章を検索キーとして、該当する対応方法をデータベースから検索して、オペレーターが使用する端末に送信する検索方法がある。

40

このようなオペレーターによる対応シーンでは、お客様を待たせないためにも、検索キーに適した検索結果をより絞り込んで検索しなければならない。しかし、このような対応シーンでの検索では、ユーザからの問い合わせや苦情等の、文の意味やユーザの意図を含む文章が検索キーとして利用されている。このため、単語やキーワード等を検索キーとして用いる一般的な検索ではなく、文章の意味やユーザの意図をより重視した検索を行う必要がある。

例えば、検索装置によって検索される辞書データを作成する際に、検索対象となる文章の構文構造や意味を解析し、解析した文章の語の品詞や係りうけ関係を抽出し、抽出された情報をマッチングの条件とする木構造の辞書データを作成しておく。そして、検索キーとして入力された検索文章を解析し、マッチングの条件と合致する解析結果を辞書データ

50

から検索して、得られた検索結果としてユーザの端末に表示させるものがある（例えば、特許文献1参照）。

【先行技術文献】

【特許文献】

【0003】

【特許文献1】特開2003-58537号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかしながら、特許文献1のような辞書データを利用した検索においては、マッチングの目的や条件に応じて、辞書データをそれぞれ作り込む必要があるという問題がある。

例えば、「携帯電話がつながりにくい」「携帯電話がつながらない」という二つの文章を利用して、“使用方法に関する問い合わせ”に適したマッチングの条件に応じた辞書データを作成する場合は、以下のような辞書データを作り込む。例えば、上記二つの文章は、携帯電話の使用法に関する問い合わせの意味合いを持っている。このため、上記二つの文章は、ここでのマッチングの条件として関連付けられる文章である。よって、“使用方法に関する問い合わせ”のための辞書データを作成する場合、上記二つの文章は、共に検索される対象としてそれぞれが関連付けられるマッチング条件の下、辞書データが作成される必要がある。

【0005】

一方、“企業に対する意見に関する対応”に適したマッチングの条件に基づく辞書データを作成する場合、「携帯電話がつながりにくい」という文章は、ユーザによる改善の要望という意味合いを持ち、「携帯電話がつながらない」という文章は、ユーザからの苦情という意味合いを持っている。このため、ユーザからの改善の要望を検索する場合、後者ではなく前者のみが検索結果として得られることが好ましい。従って、このような場合、“企業に対する意見に関する対応”に適したマッチングの条件に基づく辞書データを作成する必要がある。

【0006】

つまり、上述のように、検索の目的が異なる場合、その目的に応じたマッチングの条件に基づく辞書データをそれぞれ作成する必要があったため、辞書データを作るための労力が増大するという問題があった。

また、膨大なデータ量の辞書データを、マッチングの条件に応じてそれぞれ記憶しておかなければならないため、効率的な記憶領域の活用が図られないという問題があった。

【0007】

本発明は、このような事情を考慮してなされたものであり、その目的は、一つの辞書データを利用して異なるマッチング条件に基づく検索ができる情報検索装置、情報検索方法、およびプログラムを提供することにある。

【課題を解決するための手段】

【0008】

上述した課題を解決するために、本発明の情報検索装置は、複数の単語から構成される検索キー文が入力される入力部と、前記検索キー文を解析して、前記検索キー文を構成する前記単語に関する解析結果を得る解析部と、少なくとも1つの前記単語によって構成される文節を部分木ノードとして木構造に構成した辞書であって、少なくとも1つの前記文節によって構成される文に関するマッチ辞書情報として、当該文に含まれる前記文節に関する情報を表すルール情報を記憶するマッチ辞書記憶部と、前記マッチ辞書記憶部に記憶されている前記マッチ辞書情報と前記検索キー文との関係性を照合するためのマッチング条件が関連付けられており、当該マッチング条件を満たす単語に対して前記検索キー文との照合の度合いを評価するための評価基準を有するマッチプロファイル情報を記憶するマッチプロファイル記憶部と、前記マッチプロファイル情報に基づき、関連付けられている前記マッチング条件に従って、前記検索キー文と前記マッチ辞書情報との照合を行い、照

合の結果、前記マッチング条件を満たす前記文に対して、当該マッチプロファイル情報に関連付けられている前記評価基準に従って、前記検索キー文と前記マッチ辞書情報との照合の度合いを表すスコアを算出する検索処理部と、を備えることを特徴とする。

【0009】

また、この情報検索装置は、前記評価基準が、前記マッチング条件を満たす単語に対して前記照合の度合いに応じた点数を与えるか否かを表し、前記検索処理部が、前記評価基準に従って、前記マッチング条件を満たす単語に対して与えられる前記点数を、前記マッチング条件を満たす文毎に算出して、前記スコアを得ることを特徴とする。

【0010】

また、この情報検索装置は、前記マッチプロファイル記憶部が、それぞれ異なる特性を有する複数の前記マッチング条件のうち、少なくとも1つのマッチング条件と関連付けられている前記マッチプロファイル情報を、検索の目的に応じて複数備えることを特徴とする。

10

【0011】

また、この情報検索装置は、前記マッチプロファイル記憶部は、前記マッチング条件として、単語要素マッチング、属性マッチング、あるいは係り受けマッチングのうち、少なくとも1つが関連づけられていることを特徴とする。

【0012】

また、この情報検索装置は、前記入力部が、複数の単語から構成される検索対象文を入力し、前記解析部が、前記検索対象文を解析して、前記検索対象文を構成する前記単語に関する解析結果を得て、前記解析結果に基づき、前記単語の文字列に関する文字情報、および前記単語の属性を表す属性情報を含む前記ルール情報を、少なくとも1つの前記単語によって構成される文節と対応付けて、部分木ノードとして木構造に構成した辞書情報であって、少なくとも1つの前記文節によって構成される文に関する前記マッチ辞書情報を作成して前記マッチ辞書記憶部に記憶させる辞書作成部を、さらに備えることを特徴とする。

20

【0013】

上述した課題を解決するために、本発明の情報検索方法は、入力部が、複数の単語から構成される検索キー文の入力を受け付け、解析部が、前記検索キー文を解析して、前記検索キー文を構成する前記単語に関する解析結果を得て、検索処理部が、前記マッチ辞書情報と前記検索キー文との関係性を照合するためのマッチング条件が関連付けられており、当該マッチング条件を満たす単語に対して前記検索キー文との照合の度合いを評価するための評価基準を有するマッチプロファイル情報を記憶するマッチプロファイル記憶部から前記マッチプロファイル情報を読み出し、少なくとも1つの前記単語によって構成される文節を部分木ノードとして木構造に構成した辞書であって、少なくとも1つの前記文節によって構成される文に関するマッチ辞書情報として、当該文に含まれる前記文節に関する情報を表すルール情報を記憶するマッチ辞書記憶部の前記マッチ辞書情報を利用して、前記マッチプロファイル情報に基づき、関連付けられている前記マッチング条件に応じた前記検索キー文と前記マッチ辞書情報との照合を行い、照合の結果、前記マッチング条件を満たす前記文に対して、当該マッチプロファイル情報に関連付けられている前記評価基準に従って、前記検索キー文と前記マッチ辞書情報との照合の度合いを表すスコアを算出することを特徴とする。

30

40

【0014】

また、本発明は、コンピュータを、複数の単語から構成される検索キー文が入力される入力手段、前記検索キー文を解析して、前記検索キー文を構成する前記単語に関する解析結果を得る解析手段、前記マッチ辞書情報と前記検索キー文との関係性を照合するためのマッチング条件が関連付けられており、当該マッチング条件を満たす単語に対して前記検索キー文との照合の度合いを評価するための評価基準を有するマッチプロファイル情報を記憶するマッチプロファイル記憶部から前記マッチプロファイル情報を読み出し、少なくとも1つの前記単語によって構成される文節を部分木ノードとして木構造に構成した辞書

50

であって、少なくとも1つの前記文節によって構成される文に関するマッチ辞書情報として、当該文に含まれる前記文節に関する情報を表すルール情報を記憶するマッチ辞書記憶部の前記マッチ辞書情報を利用して、前記マッチプロファイル情報に基づき、関連付けられている前記マッチング条件に応じた前記検索キー文と前記マッチ辞書情報との照合を行い、照合の結果、前記マッチング条件を満たす前記文に対して、当該マッチプロファイル情報に関連付けられている前記評価基準に従って、前記検索キー文と前記マッチ辞書情報との照合の度合いを表すスコアを算出する検索処理手段として機能させるためのプログラムであることを特徴とする。

【0015】

また、前記入力手段は、複数の単語から構成される検索対象文を入力し、前記解析手段は、前記検索対象文を解析して、前記検索対象文を構成する前記単語に関する解析結果を得て、前記コンピュータを、さらに前記解析結果に基づき、前記単語の文字列に関する文字情報、および前記単語の属性を表す属性情報を含む前記ルール情報を、少なくとも1つの前記単語によって構成される文節と対応付けて、部分木ノードとして木構造に構成した辞書情報であって、少なくとも1つの前記文節によって構成される文に関する前記マッチ辞書情報を作成して前記マッチ辞書記憶部に記憶させる辞書作成手段として機能させるためのプログラムである。

【発明の効果】

【0016】

この発明によれば、一つの辞書データを利用して異なるマッチング条件に基づく検索を実現することができる。

【図面の簡単な説明】

【0017】

【図1】本実施の形態に係る情報検索システムの一例を示すブロック図である。

【図2】本実施の形態に係るクライアント端末装置の一例を示すブロック図である。

【図3】本実施の形態に係るWEBサーバの一例を示すブロック図である。

【図4】本実施の形態に係る日本語解析サーバの一例を示すブロック図である。

【図5】本実施の形態に係るマッチプロファイル記憶部に記憶されているマッチプロファイルの一例を示す概略図である。

【図6】本実施の形態に係るマッチ辞書記憶部に記憶されているマッチ辞書データの一例を示す概略図である。

【図7】本実施の形態に係る日本語解析サーバの一例を示すブロック図である。

【図8】本実施の形態に係る構文解析部によって作成される構文木の一例を示す概略図である。

【図9】単語要素マッチングについて説明するための概略図である。

【図10】係り受けマッチングについて説明するための概略図である。

【図11】属性マッチングについて説明するための概略図である。

【図12】本実施の形態に係る情報検索システムにおけるマッチ辞書データの作成方法の一例を示すフローチャートである。

【図13】本実施の形態に係る情報検索システムにおける検索方法の一例を示すフローチャートである。

【図14】本実施の形態に係る情報検索システムにおけるマッチング処理とスコアリング処理の一例について詳細に説明するフローチャートである。

【図15】本実施の形態に係る検索結果データの一例について説明する概略図である。

【図16】本実施の形態に係る検索結果について説明するための参考図である。

【図17】本実施の形態に係る情報検索システムにおける検索開始処理の一例を示すフローチャートである。

【図18】本実施の形態に係る情報検索システムにおける検索結果の表示方法の一例を示すフローチャートである。

【図19】本実施の形態に係るクライアント端末装置の表示部に表示される検索結果を表

10

20

30

40

50

す画像の一例を示す概略図である。

【図 20】図 19 に示す検索結果から、絞り込み検索を行った後に表示される画面の一例を示す概略図である。

【図 21】検索キー文の一例を示す参考図である。

【図 22】マッチした単文の一例を示す参考図である。

【図 23】マッチプロファイルの設定の一例を説明するための参考図である。

【図 24】マッチプロファイル A に基づき得られた検索結果を表示されている画面の一例を示す概略図である。

【図 25】マッチプロファイル B に基づき得られた検索結果を表示されている画面の一例を示す概略図である。

10

【図 26】マッチプロファイル C に基づき得られた検索結果を表示されている画面の一例を示す概略図である。

【図 27】マッチプロファイルの設定の他の例を説明するための参考図である。

【図 28】特定のマッチモードによって得られた検索結果の一例を示す概略図である。

【図 29】特定のマッチモードによって得られた検索結果の他の例を示す概略図である。

【図 30】特定のマッチモードによって得られた検索結果の他の例を示す概略図である。

【図 31】特定のマッチモードによって得られた検索結果の他の例を示す概略図である。

【発明を実施するための形態】

【0018】

以下、本発明の一実施形態を、図面を参照して説明する。図 1 は、本実施の形態に係る情報検索システムの一例を示すブロック図である。

20

図 1 に示す通り、情報検索システム 1 は、クライアント端末装置 100 と、WEB サーバ 300 と、日本語解析サーバ 500 と、データベースファイルサーバ 700 とを備える。

【0019】

クライアント端末装置 100 は、例えば、パーソナルコンピュータ等の情報計算処理装置であって、検索キー文が入力される入力部と、入力部を介してユーザから入力された検索キー文を、ネットワークを介して WEB サーバ 300 に送信する。なお、詳細については図 2 を用いて後述する。

【0020】

30

WEB サーバ 300 は、クライアント端末装置 100 と日本語解析サーバ 500 間の通信を行い、例えば、クライアント端末装置 100 から受信した検索キー文を日本語解析サーバ 500 に送信し、日本語解析サーバ 500 から受信した検索結果をクライアント端末装置 100 に送信する。なお、詳細については、図 3 を用いて後述する。

【0021】

日本語解析サーバ 500 は、検索処理部 501 と、マッチプロファイル記憶部 502 と、マッチ辞書記憶部 503 と、辞書作成部 504 と、文書解析部 505 と、メモリ領域 506 を備える。日本語解析サーバ 500 は、WEB サーバ 300 から検索キー文を受信すると、この検索キー文に基づく検索を行い、検索結果を WEB サーバ 300 を介してクライアント端末装置 100 に送信する。なお、詳細については図 4 ~ 8 を用いて後述する。

40

【0022】

データベースファイルサーバ 700 は、日本語解析サーバ 500 の検索対象となる情報（以下、検索対象情報）を記憶する記憶部であって、例えば、問い合わせの対応履歴や、修理のマニュアルに関する情報、あるいは、電子メールによってユーザ等から取得された意見や苦情等の、企業や取扱い製品やサービスに関する情報を、テキストデータとして記憶する記憶装置であるデータソース 701 を備える。

また、データベースファイルサーバ 700 は、例えば、企業においてデータウェアハウスとして利用される蓄積装置が利用可能である。

【0023】

次に、図 2 を用いて、クライアント端末装置 100 について詳細に説明する。図 2 は、

50

本実施の形態に係るクライアント端末装置１００の一例を示すブロック図である。

図２に示す通り、クライアント端末装置１００は、ブラウザ（表示制御部）１０１と、表示部１０２と、入力部１０３と、通信部１０４とを備える。

表示部１０２、例えば、液晶表示装置等であって、操作画面や検索結果画面等の表示データを表示する。

入力部１０３は、例えば、キーボードやマウスからなる入力インターフェースであって、ユーザからの操作指示や検索キー文の入力を受け付ける。

通信部１０４は、入力部１０２を介して、ユーザから検索サービスの種類が指定された場合、指定された検索サービスによる検索の実行を要求するリクエスト制御信号を、WEBサーバ３００を介して日本語解析サーバ５００に送信する。また、通信部１０４は、入力部１０２を介してユーザから入力された検索キー文を、ネットワークを介してWEBサーバ３００に送信する。

【００２４】

ブラウザ１０１は、例えば、WEBサーバ３００から受信したウェブページを表示部１０２に表示させるためのプログラム（例えば、javascript（登録商標）等）をWEBサーバ３００から受信して、このプログラムを実行する表示制御部である。また、ブラウザ１０１は、このプログラムとして機能し、表示部１０２によって表示される表示データを生成し、表示部１０２に出力する。

ブラウザ１０１は、記憶部１１１と、データ処理部１１２と、表示処理部１１３とを含み、各構成について以下説明する。

【００２５】

記憶部１１１は、データ処理部１１２や表示処理部１１３によって処理される際のプログラムや所定の設定値等を記憶する。また、記憶部１１１は、日本語解析サーバ５００によって得られた検索結果（例えば、マッチした文章、マッチした単文、マッチした単語、これらのマッチングに用いられたマッチング条件、あるいはマッチ位置情報を含むマッチ情報等）や、絞り込み検索を実行する際の検索ルール情報（例えば、入力部１０３を介して指定されたマッチした単語を検索キーとして、検索結果から、絞り込み対象を検索するためのプログラムや設定値等）を記憶する。

【００２６】

データ処理部１１２は、WEBサーバ３００から受信されるブラウザ１０１上のプログラムとして動作し、WEBサーバ３００から受信される表示データを表示部１０２の画面に表示するための表示データに変換し、表示処理部１１３を制御して表示部１０２に表示データを表示させる。また、データ処理部１１２は、記憶部１１１に記憶されている検索結果に基づき、マッチング条件を満たす単語に対して強調表示を表すタグ情報を付与した結果表示データを作成する。

表示処理部１１３は、データ処理部１１２によって制御され、データ処理部１１２が変換した表示データを表示部１０２に表示させる。

【００２７】

次に、図３を用いて、WEBサーバ３００について詳細に説明する。図３は、本実施の形態に係るWEBサーバ３００の一例を示すブロック図である。

図３に示す通り、WEBサーバ３００は、通信部３０１と、リクエスト処理部３０２と、データ変換部３０３と、記憶部３０４とを備える。

通信部３０１は、例えば、ネットワークを介してクライアント端末装置１００および日本語解析サーバ５００と通信する。

リクエスト処理部３０２は、通信部３０１を介してクライアント端末装置１００から受信したリクエスト制御信号に基づき、クライアント端末装置１００の表示部１０２によって表示される表示データのウェブページのデータを作成するようデータ変換部３０３を制御する。また、リクエスト処理部３０２は、クライアント端末装置１００からリクエスト制御信号を受信して、クライアント端末装置１００によって実行されるコードファイルや、表示データの表示に関する設定データを作成し、クライアント端末装置１００に送信す

10

20

30

40

50

る。

【 0 0 2 8 】

データ変換部 3 0 3 は、リクエスト処理部 3 0 2 によって制御され、日本語解析サーバ 5 0 0 から受信した検索結果等に基づき、クライアント端末装置 1 0 0 に対して送信するウェブページのデータを作成する。

記憶部 3 0 4 は、リクエスト処理部 3 0 2 やデータ変換部 3 0 3 によって利用される設定データ等や、日本語解析サーバ 5 0 0 の検索によって得られた検索結果を一時的に記憶する。

【 0 0 2 9 】

なお、WEBサーバ 3 0 0 は、例えば、社内 LAN (Local area network) に接続され、コールセンターのオペレーターによってなされる、ユーザからの問い合わせの対応に適した情報を提供するWEBサーバであってよく、社内における知識共有用のデータを提供するためのWEBサーバであってもよい。このように、WEBサーバ 3 0 0 は、情報検索システム 1 の検索の目的に応じて、複数あってもよい。

【 0 0 3 0 】

次に、図 4 を用いて、日本語解析サーバ 5 0 0 について詳細に説明する。図 4 は、本実施の形態に係る日本語解析サーバ 5 0 0 の一例を示すブロック図である。

図 4 に示す通り、検索処理部 5 0 1 は、異なる検索サービス α 、 β をクライアント端末装置 1 0 0 に提供するためのプログラムを実行する機能を提供し、ユーザによって指定された検索サービス α による検索を実行するための制御信号をWEBサーバ 3 0 0 から受信した場合、検索サービス α のプログラムを起動させ、検索サービス α と関連付けられているマッチプロファイルをマッチプロファイル記憶部 5 0 2 から読み出す。また、検索処理部 5 0 1 は、マッチ辞書記憶部 5 0 3 に記憶されているマッチ辞書データを読み出す。さらに、検索処理部 5 0 1 は、読み出したマッチプロファイルおよびマッチ辞書データを、メモリ領域 5 0 6 に展開して辞書オブジェクトを作成する。

【 0 0 3 1 】

また、検索処理部 5 0 1 は、WEBサーバ 3 0 0 を介してクライアント端末装置 1 0 0 から検索キー文を受信すると、受信した検索キー文を文書解析部 5 0 5 に出力し、検索結果を書き込むための検索結果オブジェクト（空の状態）をメモリ領域 5 0 6 に生成する。これにより、検索結果を記録するためのメモリ領域を確保することができる。

さらに、検索処理部 5 0 1 は、マッチプロファイル記憶部 5 0 2 から読み出したマッチプロファイルにおいて予め決められているマッチモードに従って、解析された検索キー文と、メモリ領域 5 0 6 の辞書オブジェクトに展開されているマッチ辞書データとの照合（以下、マッチングという）を行い、マッチモードの条件を満たす文章等の検索を行う（以下、マッチング処理という）。

【 0 0 3 2 】

また、検索処理部 5 0 1 は、マッチプロファイルにおいて予め決められているスコアモードに従って、マッチングによって得られた文章（以下、マッチした文章という）における、検索キー文とのマッチングの程度を評価するスコアを算出する（以下、スコアリング処理という）。なお、詳細は後述するが、検索処理部 5 0 1 は、例えば、マッチングによって得られた単文（以下、マッチした単文という）に含まれる単語のスコアを算出し、この単語のスコアの和を算出することによって、マッチした文章のスコアやマッチした単文のスコアを算出する。

さらに、検索処理部 5 0 1 は、マッチングによって得られた文章等の文章 ID 等と、当該文章等のスコアを関連付けて、メモリ領域 5 0 6 の検索結果オブジェクトに検索結果として格納する。また、検索処理部 5 0 1 は、この検索結果オブジェクトの検索結果を、WEBサーバ 3 0 0 を介してクライアント端末装置 1 0 0 に送信する。

【 0 0 3 3 】

メモリ領域 5 0 6 は、一時的に情報を記憶する記憶領域であって、例えば、検索処理部 5 0 1 によって作成される辞書オブジェクトや検索結果オブジェクトが作成される領域で

10

20

30

40

50

ある。

【 0 0 3 4 】

マッチプロファイル記憶部 5 0 2 には、例えば、各検索サービス α 、 β ・・・に応じたマッチプロファイル A、B・・・が格納されている。ここで、マッチプロファイルとは、予め決定されているマッチモードを定義するマッチモード情報と、当該マッチモードで抽出された結果に適用されるスコアの算出手法を定義するスコアモード情報とを含む。例えば、マッチプロファイル A のマッチモード定義においては、検索サービス α とマッチプロファイル A とが予め関連づけられている。よって、ユーザによって検索サービスが指定されると、この検索サービスに応じて予め決定されているマッチプロファイルと、このマッチプロファイルにおいて予め決められているマッチモードやスコアモードが、決定される。

10

なお、本発明はこれに限られず、クライアント端末装置 1 0 0 が、これら検索サービスの種類、マッチプロファイルの種類、マッチモードの種類の組み合わせを表す情報をリクエスト制御信号とともに W E B サーバ 3 0 0 を介して日本語解析サーバ 5 0 0 に送信する場合、ユーザによって、これらの組み合わせが決定される構成であってもよい。

【 0 0 3 5 】

ここで、マッチプロファイルについて、図 5 を用いて詳細に説明する。図 5 は、本実施の形態に係るマッチプロファイル記憶部 5 0 2 に記憶されているマッチプロファイルの一例を示す概略図である。

図 5 に示す通り、マッチプロファイル A は、マッチモード情報として、マッチモード定義 P A 1 を含み、スコアモード情報として、相対出現頻度フラグ P A 2 と、相対出現頻度重視係数 P A 3 と、文章出現位置 P A 4 と、検索キー出現位置 P A 5 と、述語属性マッチ係数 P A 6 と、係りうけマッチ係数 P A 7 と、品詞カテゴリ P A 8 と、接続詞評価 P A 9 と、同義語マッチ係数 P A 1 0 と、を含む。

20

【 0 0 3 6 】

マッチモード定義 P A 1 とは、検索サービス α 、 β ・・・に応じて予め決定されているマッチモードの組み合わせに関する情報である。ここで、マッチモードとは、検索キー文とマッチ辞書データのマッチングを行う際の手法を表すものであって、例えば、図 9 ~ 1 1 において後述するような、単語要素マッチング、係りうけマッチング、属性マッチングがあり、その組み合わせとして定義される。

30

【 0 0 3 7 】

スコアモード情報は、マッチモードに応じて得られた結果に対する、スコア条件や重みの付け方に関する情報を含み、マッチした文章や単文のスコアを算出する際に参照される情報である。

スコアモード情報は、相対出現頻度フラグ P A 2、相対出現頻度重視係数 P A 3、文章出現位置 P A 4、および検索キー出現位置 P A 5、述語属性マッチ係数 P A 6、係りうけマッチ係数 P A 7、品詞カテゴリ P A 8、接続詞評価 P A 9、および同義語マッチ係数 P A 1 0 を含み、いずれのマッチモードにおいても利用可能なスコアモードに関する情報である。ここで、マッチモード情報とスコアモード情報は、マッチモード定義 P A 1 において予め決められているマッチングモードに関わらず、任意に組み合わせが可能である。

40

【 0 0 3 8 】

また、スコアモード情報は、マッチした単語に対して付与されるスコアを算出するか否かを表す情報であって、算出することを表す情報が設定されている場合、それぞれにおいて付与される係数や得点が設定値として決められている。つまり、検索処理部 5 0 1 は、マッチングモード情報を利用して、マッチした文章や単文に対して、スコアモード情報を利用して、さらにマッチングした文章や単文と、検索キー文とのマッチングの度合いを評価するためのスコアを算出することができる。ここで、スコアは、マッチングした文章や単文と、検索キー文とのマッチングの度合いを示すものであって、例えば、マッチした文章等と検索キー文との関係において、文章構成や係りうけ関係が一致している、あるいは、述語属性が一致している等によって、文としての意味合いがどのくらい類似しているか

50

を評価するための点数である。

【 0 0 3 9 】

相対出現頻度フラグ P A 2 は、相対出現頻度 ($t f \times i d f$) に基づく重み付けの利用を行うか否かを表す。フラグがオンとなっている状態では、この重み付けを行うことを表し、フラグがオフとなっている状態では行わないことを表す。

ここで、相対出現頻度 ($t f \times i d f$) は、キーワード(重要語)抽出の手がかりとして一般的に用いられる相対値であって、以下の各係数 ($t f$, $i d f$) 係数を乗じたものである。

なお、 $t f$ (term frequency) は、ある文章における、特定の単語の相対出現頻度であって、 $i d f$ (inversed document frequency) は、特定の単語が含まれる文章数の逆数である。つまり、どの文章にも含まれるようなありふれた単語ほど、相対出現頻度は小さくなる。したがって、ある文章は、それに含まれ相対出現頻度 ($t f \times i d f$) の高い単語によって特徴づけられる。

【 0 0 4 0 】

相対出現頻度重視係数 P A 3 は、相対出現頻度フラグ P A 2 がオンの状態で行われる、 $t f$ 値のスコアに対する重み付けの係数である。

【 0 0 4 1 】

文章出現位置 P A 4 は、文章や単文における文の出現位置に応じた重み付けを行うか否かを表す情報である。例えば、文章出現位置 P A 4 が出現位置に応じた重み付けを行うことを表している状態で、マッチ辞書記憶部 5 0 3 に記憶されている文章の先頭に近い単語から傾斜的な係数を付与することを表している。ここでは、先頭に近い位置に出現する単語に対しては、重み付けを重くし、文末に近い位置に出現する単語ほど重み付けが軽くなるような係数が設定されている。なお、この係数は、任意に設定可能である。

【 0 0 4 2 】

検索キー出現位置 P A 5 は、検索キー文における文の出現位置に応じた重み付けを行うか否かを表す情報である。例えば、検索キー出現位置 P A 5 は、出現位置に応じた重み付けを行うことを表している状態で、検索キー文とマッチした単語が、検索キー文において出現する位置に応じて、先頭に近い位置に出現した単語等の文字列から傾斜的な係数を付与することを表している。ここでは、検索キー文の先頭に近い位置に出現する単語に対しては、重み付けを重くし、検索キー文の文末に近い位置に出現する単語ほど重み付けが軽くなるような係数が設定されている。なお、この係数は、任意に設定可能である。

【 0 0 4 3 】

述語属性マッチ係数 P A 6 は、述語属性一致の際のノードのスコア係数を表す情報である。例えば、マッチした単語や係りうけ関係にある単語に対して、これらマッチした単語と検索キー文の対応する単語との述語属性がさらに一致した場合に、どれだけ重み付けを加算するかが規定されている。なお、単純に名詞など属性が無い場合は、重み付けの加算は行わず、否定などの属性が付与され、一致した場合の重み付けの加算は、乗数としてのスコア計数を任意に設定可能である。なお、単純な名詞文節などのように属性が無い場合は、重み付けの加算は行わない。属性が付与され一致した場合の重み付けの加算は、乗数としてのスコア計数を任意に設定可能である。

【 0 0 4 4 】

係りうけマッチ係数 P A 7 は、係りうけ単位で一致した場合のスコア係数を表す情報である。例えば、マッチした単語同士が係りうけ関係にある単語 (以下、係りうけペアという) に対して重み付けを行うか否かが規定されている。なお、スコア計数は、乗数として任意に設定可能である。

【 0 0 4 5 】

品詞カテゴリ P A 8 は、品詞カテゴリごとの重み付けを表す情報であって、例えば、ユーザー単語 > 固有名詞 > 一般名詞 > 形容詞・形容動調 > 動詞といった優先順位に応じて、品詞ごとに傾斜的な係数を与えるか否かを表している。なお、品詞ごとに傾斜的な係数を任意に設定可能である。

10

20

30

40

50

【 0 0 4 6 】

接続詞評価 P A 9 は、検索キー文とマッチした文章等において、特定の接続詞が含まれている場合、当該マッチした文全体を重視(または軽視)するか否かを表す情報である。例えば、接続詞評価 P A 9 は、特定の接続詞がある場合に文全体を重視することを表す状態で、特定の接続詞が含まれているマッチした文章の各単語に乘じる係数や、特定の接続詞が含まれているマッチした文章のうちマッチした単語にのみ乗じられる係数等が規定されている。

【 0 0 4 7 】

同義語マッチ係数 P A 1 0 は、後述する文書解析部 5 0 5 の類義語処理部による文字列置換が行われた文字列のスコアを何倍にするかを表す情報である。また、同義語マッチ係数 P A 1 0 は、単語そのものが一致した場合に比較して、同義語や類義語との一致のスコア順位を落とすために用いる情報である。

【 0 0 4 8 】

マッチ辞書記憶部 5 0 3 は、マッチ辞書データを記憶し、このマッチ辞書データは、例えば、シンボル I D に置き換えられた単語情報との対応関係を表すシンボルマップ M D 1 と、各文章に関することを表す文章情報 M D 2 と、この文章に含まれる単文情報 M D 3 とを含む。

ここで、マッチ辞書データについて、図 6 を用いて詳細に説明する。図 6 は、本実施の形態に係るマッチ辞書記憶部 5 0 3 に記憶されているマッチ辞書データの一例を示す概略図である。

【 0 0 4 9 】

シンボルマップ M D 1 は、シンボル I D によって識別される単語情報のテキストデータと、単語情報を識別するためのシンボル I D とを対応付ける情報である。これにより、マッチ辞書記憶部 5 0 3 は、文章情報 M D 2 および単文情報 M D 3 において、単語情報をテキストデータで記憶することなく、シンボルマップ M D 1 において対応付けられているシンボル I D に置き換えて格納することができる。

【 0 0 5 0 】

文章情報 M D 2 は、解析された文章の構造木(詳細については後述する)をマッチ辞書記憶部 5 0 3 に登録するために必要な登録情報であって、文章 I D 2 1 と、この文章 I D 2 1 によって識別される文章の文章テキストデータ 2 2 と、当該文章が検索対象情報としてデータベースファイルサーバ 7 0 0 に格納された日時やデータソース 7 0 1 での格納場所を表すアドレス等の情報を含む文章付加情報 2 3 と、用語マップ 2 4 とを含む。ここで、用語マップ 2 4 は、文章内に含まれる単語情報の出現回数を表す情報であって、単語情報毎の出現回数とシンボル I D とが対応付けられている。

【 0 0 5 1 】

一方、単文情報 M D 3 は、単文に含まれる文節毎に、構造木(図 8 参照にて、詳細は後述する)における部分木ノードの情報を表すルール(ルール情報) 3 2 を含み、各単文を識別するための単文 I D 3 1 が付与されている。

単文情報 M D 3 に含まれるルール 3 2 は、例えば、語情報 3 2 1、述語属性 3 2 2、親ルール I D 3 2 3、重み値 3 2 4、接続詞種別 3 2 5、カテゴリ 3 2 6、子ノード有フラグ 3 2 7 等を含む。

【 0 0 5 2 】

語情報 3 2 1 は、例えば、シンボル I D や、単文内における単語情報の位置を表す位置情報等を含む。なお、語情報 3 2 1 は、単文内に含まれる単語の数に応じた数の語情報 1、語情報 2、・・・、語情報 n を含み、例えば、単語情報のシンボル I D、単文内における単語の位置情報(開始位置と終了位置)を含む。

述語情報 3 2 2 は、例えば、語 I D や、動詞形容詞等の句の属性、および句の意味(否定、否定の傾向、願望、肯定・・・等)を表す属性シンボル I D 等を含む。

【 0 0 5 3 】

親ルール I D 3 2 3 は、親子の係りうけ関係にある親の部分木ノードの文節を表す情報

10

20

30

40

50

である。

重み値 3 2 4 は、例えば、文章内での主語や述語に応じた重みを付与する係数等である。また、重み値 3 2 4 は、後述するスコアリングにおいて、当該ルール（ノード）がマッチした際に基準となる点数を規定する係数等である。通常は辞書全体で任意の一つの値に設定されるが、辞書の作成時に、文章内での主語や述語に応じた重みを付与することができる。

接続詞種別 3 2 5 は、ルール 3 2 に対応する文節（句）が、「したがって、だから、すなわち、・・・」等の接続詞である場合、その接続詞を表す情報である。

カテゴリ 3 2 6 は、動詞、名詞、副詞、接続詞・・・等の品詞の種類を表す情報である。

10

子ノード有フラグ 3 2 7 は、親子の係りうけ関係にある子の部分木ノードの文節の有無を表す情報であって、フラグがオンされている状態で、当該部分木ノードが親の部分木ノードであることを表す情報である。

【 0 0 5 4 】

次に、図 7 を用いて、検索処理部 5 0 1、辞書作成部 5 0 4、および文書解析部 5 0 5 について詳細に説明する。図 7 は、本実施の形態に係る日本語解析サーバ 5 0 0 の一例を示すブロック図である。

辞書作成部 5 0 4 は、データベースファイルサーバ 7 0 0 から、検索対象となる文章（検索対象情報）を読み出し、検索対象情報に含まれる文章を、例えば句点ごとに区切って単文の単位に変換し、文書解析部 5 0 5 に出力する。例えば、検索対象情報に含まれる文章 A のテキストデータが「P C 内にある画像を送信したところ受信できませんでした。また、携帯内にある画像を送信したところ受信できました。」である場合、句点「。」で区切って 2 つの単文に分割し、単文 A 1 「P C 内にある画像を送信したところ受信できませんでした。」と、単文 A 2 「また、携帯内にある画像を送信したところ受信できました。」を文書解析部 5 0 5 に出力する。

20

また、辞書作成部 5 0 4 は、文書解析部 5 0 5 によって解析された結果を、文書解析部 5 0 5 から受け取った場合、マッチ辞書記憶部 5 0 3 に格納する。

なお、辞書作成部 5 0 4 は、データベースファイルサーバ 7 0 0 から読み出した検索対象情報を、適当な長さに区切るものであればよく、例えば、ピリオド、箇条書きの一文、スペースや改行によって一文と判断される一文で、区切るものであってもよい。

30

【 0 0 5 5 】

文書解析部 5 0 5 は、辞書作成部 5 0 4 によってマッチ辞書データが作成される際に、例えば、単文単位に区切られた検索対象情報を、辞書作成部 5 0 4 から受け取り、文書解析を行って、解析結果を辞書作成部 5 0 4 に出力する。

また、文書解析部 5 0 5 は、クライアント端末装置 1 0 0 から特定の検索サービスによる検索を実行するようリクエスト制御信号を受信した場合、検索処理部 5 0 1 から受け取った検索キー文に対して文書解析を行い、解析結果を検索処理部 5 0 1 に出力する。以下、文書解析部 5 0 5 について詳細に説明する。

【 0 0 5 6 】

文書解析部 5 0 5 は、辞書部 5 5 0 と、解析部 5 5 1 とを備える。

40

辞書部 5 5 0 は、システム辞書 5 5 0 1 と、ユーザ辞書 5 5 0 2 と、類義語辞書 5 5 0 3 とを備える。一方、解析部 5 5 1 は、形態素解析部 5 5 1 1 と、構文解析部 5 5 1 2 と、類義語処理部 5 5 1 3 とを備える。

【 0 0 5 7 】

システム辞書 5 5 0 1 は、例えば、形態素として意味を持つ最小単位の単語と、この単語の意味や品詞、属性情報等が対応付けられている辞書データである。

ユーザ辞書 5 5 0 2 は、例えば、日本語解析サーバ 5 0 0 を利用する管理者等によって、システム辞書 5 5 0 1 に追加される辞書データである。

類義語辞書は、複数の類義語や同義語を置き換えるため、単語とその類義語等とを対応付けている辞書データであって、例えば、マッチ辞書記憶部 5 0 3 のマッチ辞書データの

50

単語情報と、その類義語が対応付けられている。

【 0 0 5 8 】

形態素解析部 5 5 1 1 は、例えば、辞書作成部 5 0 4 によって、単文ごとに区切られた検索対象情報を受け取り、検索対象となる文章を複数の形態素（単語情報）に分解する。例えば、形態素解析部 5 5 1 1 は、文章 A が入力されると、文章 A に含まれる単文 A 1 を、複数の形態素（「 P C 」 「 内 」 「 に 」 「 ある 」 「 画像 」 「 を 」 「 送信 」 「 した 」 「 と ころ 」 「 受信 」 「 できませんでした 」 ）に分解する。

このようにして、形態素解析部 5 5 1 1 は、マッチ辞書データ作成の際に、検索対象情報を形態素に分解することができるが、これに限られず、検索キー文を入力して検索を行う際には、検索処理部 5 0 1 によって単文ごとに区切られた検索キー文を受け取り、形態素に分解する。

10

【 0 0 5 9 】

また、形態素解析部 5 5 1 1 は、システム辞書 5 5 0 1 やユーザ辞書 5 5 0 2 を参照して、分解した形態素の品詞を検索し、得られた品詞情報に基づき、文章中の係りうけ関係や形態素の意味に応じた文節を作成する。例えば、構文解析部 5 5 1 2 は、単文 A 1 から分解された形態素に基づき、文節 A 1 0 1 「 P C 内に 」 と、文節 A 1 0 2 「 ある 」 と、文節 A 1 0 3 「 画像を 」 と、文節 A 1 0 4 「 送信したところ 」 と、文節 A 1 0 5 「 受信できませんでした 」 とを作成する。なお、ここで、文節とは、少なくとも 1 つの単語を含む文字列の単位である。また、単文とは、少なくとも 1 つの文節を含み 1 つの文からなる文字列の単位であって、例えば、句点等で区切られる。さらに、文章とは、複数の単文を含む文の単位である。なお、文は、単文と文章の両方を含むものとする。

20

【 0 0 6 0 】

さらに、形態素解析部 5 5 1 1 は、システム辞書 5 5 0 1 およびユーザ辞書 5 5 0 2 に記憶されている辞書データを参照して、それぞれ品詞のカテゴリ（例えば、動詞、名詞、副詞、接続詞・・・等）や、接続詞の種類（例えば、したがって、だから、といった・・・等）、動詞や形容詞等の句の意味を表す属性（例えば、否定、否定の傾向、願望、肯定・・・等）を検索し、検索によって得られた情報を形態素や文節に付与する。

例えば、形態素解析部 5 5 1 1 は、文節 A 1 0 5 「 受信できません 」 を解析して、品詞は “ 名詞（サ変接続） ” であって、句の意味は “ 否定 ” であることを解析結果として得て、この解析結果を文節 A 1 0 5 に付与する。

30

【 0 0 6 1 】

構文解析部 5 5 1 2 は、形態素解析部 5 5 1 1 によって解析された情報に基づき、文章を構成する文節の品詞や意味、属性情報、文章内での位置、並び等を評価し、文章における文節どうしの係りうけ関係を解析し、解析結果を類義語処理部 5 5 1 3 に出力する。

【 0 0 6 2 】

さらに、構文解析部 5 5 1 2 は、検索対象情報を解析した際に、各文章を識別するための文章 ID を付与し、この文章をマッチ辞書記憶部 5 0 3 に登録（記憶）する際に必要な登録情報を生成する。また、構文解析部 5 5 1 2 は、単語情報や文節等の解析結果を利用して、文節同士の係りうけ関係に基づき、図 8 に示すような構造木を作成し、部分木ノード毎のルールを表す情報を生成して、類義語処理部 5 5 1 3 に出力する。

40

なお、ルールとは、図 8 に示す構造木を構成する部分木ノード毎に対応付けられている情報であって、図 6 に示したように、語情報 3 2 1、述語属性 3 2 2、親ルール ID 3 2 3、重み値 3 2 4、接続詞種別 3 2 5、カテゴリ 3 2 6、子ノード有フラグ 3 2 7 等を含む。

【 0 0 6 3 】

ここで、図 8 を用いて構文解析部 5 5 1 2 によって作成される構造木について説明する。図 8 は、構文解析部 5 5 1 2 によって作成される構造木の一例を示す概略図である。

図 8 に示す通り、部分木ノードに対応するルールは、形態素解析部 5 5 1 1 によって区分された文節毎に作成されている。また、部分木ノードによって構成される構造木は、文章の前後関係に応じた係りうけ関係によって作成されている。

50

【 0 0 6 4 】

類義語処理部 5 5 1 3 は、類義語辞書 5 5 0 3 を参照して、分解された形態素や文節に対して、統一すべき類義語があるか否かを検索し、該当する類義語があれば、類義語辞書 5 5 0 3 から検索によって得られた類義語への置換えを行う。

【 0 0 6 5 】

ここで、辞書作成部 5 0 4 が、データベースファイルサーバ 7 0 0 から検索対象情報を読み出して、単文に区切られた検索対象情報を文書解析部 5 0 5 に出力した場合、文書解析部 5 0 5 は、上述のような文書解析を行い、解析結果を辞書作成部 5 0 4 に出力する。

辞書作成部 5 0 4 は、文書解析部 5 0 5 から解析結果を受け取り、単文を構成する文節のルールとしての情報、例えば、単語情報、シンボル ID、各単語語情報の述語属性、ルール（部分木ノード）間のつながりを表す親ルール ID や子ノード有フラグ、重み値、接続詞種別、カテゴリ、等のマッチ辞書記憶部 5 0 3 の登録に必要な情報を得る。

10

【 0 0 6 6 】

また、辞書作成部 5 0 4 は、マッチ辞書記憶部 5 0 3 から読み出したシンボルマップ MD 1 を確認して、マッチ辞書記憶部 5 0 3 内で統一的に利用されている単語情報を、シンボル ID に置き換える。さらに辞書作成部 5 0 4 は、文章 ID や、文章テキスト、文章付加情報や用語マップ等を含む文章情報 MD 2 と、単語 ID 3 1 とルール 3 2 とを含む単語情報 MD 3 とを作成し、マッチ辞書登録用のデータを作成する。また、辞書作成部 5 0 4 は、このマッチ辞書登録用のデータを、マッチ辞書記憶部 5 0 3 のマッチ辞書データに追加する。

20

なお、辞書作成部 5 0 4 は、文書解析部 5 0 5 から受け取った解析結果の中に、シンボルマップ MD 1 に対応するシンボル ID がない単語情報が存在した場合、当該単語情報に対して、新たなシンボル ID を付与して、当該単語情報と新たなシンボル ID との対応関係をシンボルマップ MD 1 に追加する。

【 0 0 6 7 】

次に、図 9 ～ 1 1 を用いて、検索処理部 5 0 1 によって実行されるマッチモードについて詳細に説明する。図 9 は、単語要素マッチングについて説明するための概略図である。図 1 0 は、係りうけマッチングについて説明するための概略図である。図 1 1 は、属性マッチングについて説明するための概略図である。

【 0 0 6 8 】

30

図 9 (a) ～ (c) に示す通り、単語要素マッチングは、積集合タイプと、全集合タイプと、部分集合タイプの 3 タイプがある。単語要素マッチングでは、検索処理部 5 0 1 が、マッチ辞書記憶部 5 0 3 に記憶されているマッチ辞書データのルール 3 2 と、検索キー文においてルール 3 2 と対応する文字列（例えば、文節）との照合を行う。なお、検索処理部 5 0 1 は、単語要素マッチングがマッチプロファイルのマッチモード定義 PA 1 において、積集合タイプと、全集合タイプと、部分集合タイプのうち、いずれか 1 つのタイプを実行する。

【 0 0 6 9 】

ここで、積集合タイプとは、マッチ辞書記憶部 5 0 3 のマッチ辞書データのルールの文字列（部分木ノードに対応する文節）の少なくとも一部と、検索キー文に含まれる文字列（部分木ノードに対応する文節）の少なくとも一部とが一致する場合、この一致する文字列をマッチモードの条件を満たす単語として得るマッチモードである。なお、一部でも一致する単語がなかった場合、マッチモードの条件を満たす単文が得られなかったという結果となる。

40

【 0 0 7 0 】

例えば、図 9 (a) に示す通り、マッチ辞書記憶部 5 0 3 のマッチ辞書データのルールに対応する文節「サッカー 観戦」と、検索キー文に含まれる文節「サッカー 少年」、「観戦 ツアー」および「戦」とがそれぞれ照合された場合、検索キー文に含まれる文節「サッカー 少年」は、その一部である単語「サッカー」において、ルールに対応する文節「サッカー 観戦」の一部と一致している。この場合、ルールに対応する文節「サッカ

50

ー 観戦」は、マッチモードの条件を満たす。

また、検索キー文に含まれる文節「観戦 ツアー」は、その一部である単語「観戦」が、ルールに対応する文節「サッカー 観戦」の一部と一致しており、マッチモードの条件を満たす。

このように、マッチモードの条件を満たす場合、マッチング処理の結果として、ルールに対応する文節「サッカー 観戦」が得られる。また、検索処理部 501 は、このマッチモードの条件を満たす単語をマッチした単語として検出する。

【0071】

一方、検索キー文に含まれる文節「戦」は、ルールに対応する文節「サッカー 観戦」に含まれる単語の単位で、一致する部分がないため、マッチモードの条件を満たさない。このため、マッチング処理の結果としては、マッチモードの条件を満たす単文が得られなかったという結果となる。なお、単語「戦」は、文節「サッカー 観戦」の単語「観戦」の一部に含まれる文字であるが、単語単位で比較した場合、「戦」と「観戦」とは異なる文字列（単語）となるため、ここでは、一致しないと判断される。

【0072】

次に、全集合タイプについて図9(b)を用いて説明する。

全集合タイプは、マッチ辞書記憶部 503 のマッチ辞書データのルールの文字列（部分木ノードに対応する文節）と、検索キー文に含まれる文字列（部分木ノードに対応する文節）の全てが一致する場合、一致した文字列がマッチモードの条件を満たす単語（マッチした単語）として得られるマッチモードである。この場合、上述の積集合と異なり、一部でも（単語1つでも）一致する文節があった場合であって、文節の全ての文字列が一致しなければ、マッチモードの条件を満たす単文が得られなかったという結果となる。

例えば、図9(b)に示す通り、マッチ辞書記憶部 503 のマッチ辞書データのルールに対応する文節「サッカー 観戦」と、検索キー文に含まれる文節「サッカー 観戦」「観戦」とが照合された場合、文節「サッカー 観戦」は、ルールに対応する文節と、検索キー文の両方ともが完全に一致しており、マッチモードの条件を満たすため、マッチング処理の結果として、文節「サッカー 観戦」（マッチした文節）が得られる。

一方、検索キー文の文節「観戦」は、ルールに対応する文節「サッカー 観戦」の一部とは一致するものの、全ての文字列が一致していないため、マッチモードの条件を満たさず、マッチング処理の結果としては、マッチモードの条件を満たす単文が得られなかったという結果となる。

【0073】

次に、部分集合タイプについて図9(c)を用いて説明する。

部分集合タイプは、マッチ辞書記憶部 503 のマッチ辞書データのルールの文字列（部分木ノードに対応する文節）が、検索キー文に含まれる文字列（部分木ノードに対応する文節）の一部と完全に一致する場合、一致している文字列をマッチモードの条件を満たす単語として得るマッチモードである。この場合、上述の積集合タイプと異なり、少なくともルールに対応する文節の全てを含んでいれば、マッチモードの条件を満たし、一致する単語や文節をマッチした単語あるいはマッチした文節して得る。一方、検索キー文に含まれる文節が、ルールに対応する文節の全てを含むものでなければ、マッチモードの条件を満たす語情報が得られなかったという結果となる。

【0074】

例えば、図9(c)に示す通り、マッチ辞書記憶部 503 のマッチ辞書データのルールに対応する文節「サッカー 観戦」と、検索キー文に含まれる文節「サッカー 観戦 ツアー」「観戦 ツアー」とが照合された場合、ルールに対応する文節「サッカー 観戦」は、検索キー文に含まれる文節「サッカー 観戦 ツアー」の一部であって、その文節を構成する全ての単語が、検索キー文に含まれる文節に含まれる単語として一致しているため、マッチモードの条件を満たす。このように、マッチモードの条件を満たす場合、マッチング処理の結果として、ルールに対応する文節「サッカー 観戦」（マッチした文節）が得られる。

一方、検索キー文の文節「観戦 ツアー」は、ルールに対応する文節「サッカー 観戦」の一部の単語「観戦」が一致するものの、ルールに対応する文節の全てが検索キー文の一部と一致していなため、マッチモードの条件を満たさず、マッチング処理の結果としては、マッチモードの条件を満たす単文が得られなかったという結果となる。

【 0 0 7 5 】

次に、係りうけマッチングの例について図 1 0 (a) ~ (b) を用いて説明する。

図 1 0 (a) ~ (b) に示す通り、係りうけマッチングは、係りうけの関係のあるものを抽出するノード親子関係タイプと、係り受けの評価を行わないノード単独タイプの 2 タイプがある。係りうけマッチングでは、検索処理部 5 0 1 によって、マッチ辞書記憶部 5 0 3 に記憶されているマッチ辞書データのルール 3 2 の文字列（部分木ノードに対応する文節）の係りうけ関係と、検索キー文に含まれる文字列（部分木ノードに対応する文節）の係りうけ関係との照合が行われる。なお、検索処理部 5 0 1 は、係りうけマッチングがマッチプロファイルのマッチモード定義 P A 1 おいて、ノード親子関係タイプと、ノード単独タイプのうち、いずれか 1 つのタイプを実行する。

【 0 0 7 6 】

ここで、ノード親子関係タイプは、単語要素マッチングの集積合タイプによってマッチングした単語同士の係りうけ関係についての条件であって、マッチ辞書記憶部 5 0 3 のマッチ辞書データのルールの文字列（部分木ノードに対応する文節）の親子関係と、検索キー文に含まれる文字列のうち単語要素マッチングの集積合タイプによって得られたマッチした単語（部分木ノードに対応する文節）の親子関係が一致する場合、この一致する文字列をマッチモードの条件を満たす単語として得るマッチモードである。なお、親子関係が一致する文字列がなかった場合、マッチモードの条件を満たす単文が得られなかったという結果となる。

【 0 0 7 7 】

例えば、図 1 0 (a) に示す通り、マッチ辞書記憶部 5 0 3 のマッチ辞書データの親子関係において、親ルールに対応する文節が「行く」、子ルールに対応する文節が「サッカー 観戦」である場合、この親子関係と一致するパターンは、以下の 2 つである。つまり、親ルールに対応する文節が「行く」であって、子ルールに対応する文節が「サッカー」であるパターンと、親ルールに対応する文節が「行く」であって、子ルールに対応する文節が「観戦」であるパターンである。

よって、検索キー文に含まれる文字列として、親ルールに対応する文節が「行く」、子ルールに対応する文節が「サッカー」であるものの親子関係「(サッカー) - (行く)」は、マッチモードの条件を満たす。このように、マッチモードの条件を満たす場合、マッチング処理の結果として、ルールに対応する文節「サッカー 観戦」—「行く」の親子関係が得られる。

一方、検索キー文に含まれる文字列として、子ルールに対応する文節が「サッカー 観戦」であって、親ルールに対応する文節がないものや、子ルールに対応する文節が「行く」であって、親ルールに対応する文節がないものは、マッチモードの条件を満たさない。

【 0 0 7 8 】

次に、図 1 0 (b) を用いて、ノード単独タイプについて説明する。

ノード単独タイプは、マッチ辞書記憶部 5 0 3 のマッチ辞書データのルールの文字列（部分木ノードに対応する文節）の親子関係と、検索キー文に含まれる文字列（部分木ノードに対応する文節）の親子関係において、親ノードあるいは子ノードの少なくともいずれか一方が一致する場合、この一致する文字列をマッチモードの条件を満たす単語として得るマッチモードである。すなわち、実際には係り受けの評価は行われないことになる。なお、ノード内の文字列の比較は、単語要素マッチングに応じたタイプで行われるが、親ノードあるいは子ノードのいずれか一方でも一致する文字列がなかった場合、マッチモードの条件を満たす単文が得られなかったという結果となる。

【 0 0 7 9 】

例えば、図 1 0 (b) に示す通り、マッチ辞書記憶部 5 0 3 のマッチ辞書データの親子

関係が、親ノードに対応する文節「サッカー 観戦」、子ノードに対応する文節が「行く」である場合、親ノードに対応する文節として「サッカー」を含む検索キー文は、マッチモードの条件を満たし、マッチング処理の結果として、ルールに対応する文節「サッカー 観戦」―「行く」の親子関係が得られる。

また、子ノードに対応する文節として「行く」を含む検索キー文は、マッチング条件を満たす。一方、子ノードに対応する文節として「する」を含む検索キー文は、親ノードあるいは子ノードのいずれか一方でも一致する文字列がないため、マッチモードの条件を満たさない。

例えば、マッチプロファイルのマッチモード定義 P A 1 においては、係り受けマッチングを行うか否かが予め決められており、上述した複数のタイプのうち、いずれか一方が予め決められている。

【 0 0 8 0 】

次に、属性マッチングについて説明する。属性マッチングは、文属性一致タイプと、係り受けマッチングと同様に、実質的には属性のマッチを評価しない単語一致タイプがある。ここでは、文属性一致タイプを図 1 1 に示す。

属性マッチングでは、検索処理部 5 0 1 によって、マッチ辞書記憶部 5 0 3 に記憶されているマッチ辞書データのルールの文字列（部分木ノードに対応する文節）の属性と、検索キー文における対応する文字列（部分木ノードに対応する文節）の属性との照合が行われる。

ここで、文属性の一致のタイプは、マッチ辞書記憶部 5 0 3 のマッチ辞書データのルールの文節の少なくとも一部と、検索キー文に含まれる文節のうち少なくとも一部（単語）が一致しており、この一致している部分の属性もそれぞれ一致する場合、この一致する文字列をマッチモードの条件を満たす単語として得られるマッチモードである。なお、文字列が一致していても、属性が異なる場合は、マッチモードの条件を満たす単文が得られなかったという結果となる。

【 0 0 8 1 】

例えば、マッチ辞書記憶部 5 0 3 のマッチ辞書データのルール 3 2 において、ルールと対応する文節「サッカー 観戦」の述語属性が「否定」であって、検索キー文に含まれる文節「観戦」の述語属性が「否定」である場合、マッチモードの条件を満たす。つまり、「サッカー 観戦しない」の場合、「サッカー（名詞）」+「観戦（名詞）」+「しない（助動詞）」に分解されるが、「しない（助動詞）」の属性が「否定」である。このため、「サッカー 観戦」の述語属性が「否定」となり、マッチモードの条件を満たす。

一方、検索キー文に含まれる文節「サッカー 観戦」の述語属性が「可能」である場合、文節「サッカー 観戦」の文字列は一致しているものの、属性が異なるため、マッチモードの条件を満たさない。つまり、「サッカー観戦できる」の場合、「サッカー（名詞）」+「観戦（名詞）」+「できる（助動詞）」に分解されるが、「できる（助動詞）」の属性が「可能」である。このため、「サッカー 観戦」の述語属性は「可能」となり、マッチモードの条件を満たさない。

ここでいう「属性」とは、助動詞の意味情報のことで、否定、疑問、可能などがある。例えば、「使えない」という語に対しては、「使う（動詞）」+「ない（助動詞）」に分解されるが、「ない（助動詞）」の属性は「否定」である。

なお、属性マッチングにおいては、名詞など「属性なし」同士でマッチしてもマッチモードの条件を満たすと判定する。

【 0 0 8 2 】

次に、図 1 2 を用いて、本実施の形態に係る情報検索システム 1 におけるマッチ辞書データの作成方法の一例について説明する。図 1 2 は、本実施の形態に係る情報検索システム 1 におけるマッチ辞書データの作成方法の一例を示すフローチャートである。

図 1 2 に示す通り、マッチ辞書データの作成が、例えば日本語解析サーバ 5 0 0 の操作部（図示せず）から指示された場合、辞書作成部 5 0 4 は、データベースファイルサーバ

10

20

30

40

50

700のデータソース701から検索対象となる文章を読み出し、文章を句点等ごとに区切って単文の単位として、文書解析部505に出力する。例えば、検索対象となる文章Aのテキストデータが「PC内にある画像を送信したところ受信できませんでした。また、携帯内にある画像を送信したところ受信できました。」である場合、句点「。」で区切られている2つの単文に区切って、単文A1「PC内にある画像を送信したところ受信できませんでした。」と、単文A2「また、携帯内にある画像を送信したところ受信できました。」を文書解析部505に出力する(ステップST1)。

【0083】

単文A1を受け取った文書解析部505の形態素解析部5511は、単文Aを形態素(例えば単語単位)に分離して、検索対象となる文章を複数の形態素に分解する。例えば、形態素解析部5511は、文章Aが入力されると、文章Aに含まれる単文A1を複数の単語「PC」「内」「に」「ある」「画像」「を」「送信」「した」「ところ」「受信」「できませんでした」に分解する。

10

そして、形態素解析部5511は、システム辞書5501やユーザ辞書5502を参照して、分解した形態素の品詞や属性、意味等を解析し、解析結果として得る(ステップST2)。

【0084】

次いで、構文解析部5512は、少なくとも1つ以上の形態素を組み合わせて、構造木の部分木ノードに対応する文節を作成する。なお、ここでは、部分木ノードに対応する文字列として、文章を構成する構成要素の1つであって、文章を実際の言葉として不自然にならない程度に区切ったときに得られる最小のひとまとまりである文節を1つの単位とした例について説明する。しかし、本発明はこれに限られない。

20

【0085】

そして、構文解析部5512は、形態素解析部5511によって解析された解析結果に基づき、文章を構成する単語の品詞や意味、属性情報、文章内での位置、並び等を評価し、文章における文節同士の係り受け関係を解析し、文節どうしの係り受け関係や、単語の出現位置、文章中での文の成分(主語や述語等)等を解析結果として得る。また、構文解析部5512は、各単文を識別するための単文IDを単文毎に付与する。

次いで、構文解析部5512は、解析結果に基づき、文節を部分木ノードとする構造木を作成し、解析結果を類義語処理部5513に出力する(ステップST3)。

30

【0086】

類義語処理部5513は、類義語辞書5503を参照して、分解された単語に対して、統一すべき類義語や同意語があるか否かを検索し、該当する類義語等があれば、該当する単語を、類義語辞書5503から検索によって得られた類義語等に置換える(ステップST4)。そして、類義語処理部5513は、解析部551による解析結果を辞書作成部504に出力する。

【0087】

解析結果を受け取った辞書作成部504は、この解析結果から、文節毎のルールとして、例えば、語情報テキストデータ、各語情報の述語属性、ルール(部分木ノード)間のつながりを表す親ルールIDや子ノード有フラグ、重み値、接続詞種別、カテゴリ、等のマッチ辞書記憶部503の登録に必要な情報を得る。そして、辞書作成部504は、これら文節によって構成される単文毎に、各文節のルールと単文IDとをまとめて、マッチ辞書データの単文情報MD3として登録可能な登録用データを作成する。

40

【0088】

また、辞書作成部504は、マッチ辞書記憶部503からシンボルマップMD1を読み出して、検索対象である文章に、マッチ辞書記憶部503内で統一的に利用されている語情報があるか否かを確認し、単語情報と同一の単語があった場合は、この単語を、シンボルIDに置き換える。なお、シンボルマップMD1内に置き換えるべき同一の単語情報がない場合、辞書作成部504は、新たなシンボルIDを当該単語情報に与える。

そして、辞書作成部504は、文章IDや、文章テキスト、文章付加情報や用語マップ

50

等を含む文章情報（文章の登録のために必要な情報）、および単文情報に基づくマッチ辞書の登録用データを作成する（ステップS T 5）。

【0089】

次いで、辞書作成部504は、作成したマッチ辞書の登録用データをマッチ辞書記憶部503に書込み、解析結果をマッチ辞書データとして登録する（ステップS T 6）。

【0090】

次に、図13を用いて、本実施の形態に係る情報検索システム1におけるマッチ辞書データの基づく検索方法の一例について説明する。図13は、本実施の形態に係る情報検索システム1における検索方法の一例を示すフローチャートである。

図13に示す通り、例えば、ユーザによってクライアント端末装置100の入力部103から検索サービスαが指定された場合、クライアント端末装置100は、指定された検索サービスαによる検索を実行するためのリクエスト制御信号を、WEBサーバ300を介して日本語解析サーバ500に送信する。

【0091】

日本語解析サーバ500は、この検索リクエスト制御信号を受信すると、検索処理部501が検索サービスαのプログラムを起動させ、検索サービスαと関連付けられているマッチプロファイルをマッチプロファイル記憶部502から読み出す。ここでは、マッチプロファイルAのマッチモード定義において、予め検索サービスαとマッチプロファイルAとが関連づけられているため、検索プログラムαが起動されることにより、検索処理部501がマッチプロファイルAを読み出す。また、検索処理部501は、マッチ辞書記憶部503に記憶されているマッチ辞書データを読み出す（ステップS T 10）。

【0092】

そして、検索処理部501は、例えば、読み出したマッチプロファイルAおよびマッチ辞書データとを、メモリ領域506に展開して辞書オブジェクトを作成する（ステップS T 11）。なお、検索処理部501は、マッチプロファイルAから、各文章に付与された用語マップ24を読み出し、単語情報の出現頻度情報を計算し、メモリ領域506に一時的に記憶させておいてもよい。このように、辞書オブジェクトを作成するメモリ展開時に、予め各単語情報の出現頻度情報を得て置くことにより、マッチング処理の際に、単語情報の出現頻度情報を計算する処理負荷が軽減される。

【0093】

ここで、クライアント端末装置100の入力部103から、ユーザによって検索キー文が入力されると、クライアント端末装置100は、検索キー文を、WEBサーバ300を介して日本語解析サーバ500に送信する（ステップS T 12）。

そして、日本語解析サーバ500は、この検索キー文を受信すると（ステップS T 13）、以下に示す通り、この検索キー文に基づく検索を行う。

【0094】

まず、検索処理部501は、検索サービスαに対する検索結果をクライアント端末装置100に返信するため、メモリ領域506に、空のオブジェクトである検索結果オブジェクトを生成し、結果記録用の記憶領域を確保する（ステップS T 14）。

そして、検索処理部501は、検索サービスαのプログラムに従って、検索キー文の解析を行う。すなわち、形態素解析部5511は、検索処理部501によって分割された単文の形態素解析を行い形態素に分割し、システム辞書5501およびユーザ辞書5502を参照して、品詞や属性等を検索する。そして、形態素解析部5511は、得られた品詞や属性等を表す情報に基づき、文章中の係りうけ関係や形態素の意味に応じた文節を作成する。

【0095】

次いで、構文解析部5512は、形態素解析部5511による解析結果に基づき、文章を構成する文節の品詞や意味、属性情報、文章内での位置、並び等を評価し、文章における文節同士の係りうけ関係を解析し、解析結果を類義語処理部5513に出力する。

類義語処理部5513は、類義語辞書5503を参照して、分解された単語や文節に対

10

20

30

40

50

して、統一すべき類義語や同義語があるか否かを検索し、該当する類義語等があれば、該当する単語や文節を、類義語辞書 5503 から検索によって得られた類義語等に置換える。そして、類義語処理部 5513 は、解析部 551 による解析結果を検索処理部 501 に出力する（ステップ S T 15）。

【0096】

そして、検索処理部 501 は、マッチプロファイル A において予め決められているマッチモードに従って、解析された検索キー文と、メモリ領域 506 の辞書オブジェクトに展開されているマッチ辞書データとのマッチングを行い、各マッチングのどのタイプに属するかを判定し、マッチモード定義の条件を満たす文章等の検索（マッチング処理）を行う（ステップ S T 16）。なお、詳細については、後述する。

10

さらに、検索処理部 501 は、ステップ S T 16 において、マッチプロファイル A において予め決められているスコアモードに従って、マッチングによって得られた文章等における、検索キー文とのマッチングの程度を評価するスコアを算出する（スコアリング処理）。

【0097】

そして、検索処理部 501 は、マッチング処理における検索によって得られたマッチモードの条件を満たす文章と、スコアリング処理によって得られたスコアとを、メモリ領域 506 の検索結果オブジェクトに書き込む（ステップ S T 17）。

そして、検索処理部 501 は、WEBサーバ 300 を介してクライアント端末装置 100 に、検索結果オブジェクトの内容を送信する（ステップ S T 18）。

20

【0098】

次に、図 14 を用いて、本実施の形態に係る情報検索システム 1 におけるマッチング処理とスコアリング処理の一例について詳細に説明する。図 14 は、本実施の形態に係る情報検索システム 1 におけるマッチング処理とスコアリング処理の一例について詳細に説明するフローチャートである。なお、図 14 に示す処理は、図 13 のステップ S T 16 に対応する処理を詳細に記載したものである。

図 14 に示す通り、検索処理部 501 は、マッチ辞書データのシンボルマップ MD1 を参照して、図 13 のステップ S T 15 において文書解析部 505 によって解析された検索キー文の単語情報をシンボル ID に置き換える（ステップ S T 20）。

そして、検索処理部 501 は、マッチプロファイル A において予め決められているマッチモードに従ってマッチング処理を行う。本実施の形態において、検索処理部 501 は、単語要素マッチング、係りうけマッチング、属性マッチングについて、それぞれどのタイプに合致するかを判定し、マッチモード定義 PA1 によって定義された条件に合致する単文を抽出する（ステップ S T 21）。これにより、検索処理部 501 は、マッチング処理によって検索キー文とマッチした文章を、マッチ辞書データから検索によって得ることができる。

30

【0099】

検索処理部 501 はマッチング処理によって得られた結果に対し、各マッチングモードで判定されたマッチタイプの情報と、マッチプロファイル A によって定義されているスコアモードを利用してスコアを算出し、これらの和をメモリ領域 506 に一時的に記憶させる（ステップ S T 22）。ここで、スコアの算出にあたっては、実際に結果の抽出に用いられたマッチモード定義とは関連なく行われる。これは、マッチモード定義を条件としたマッチングは、検索結果そのものの抽出である一方、スコア算出の処理は、抽出された結果の中で、より検索の意図に適した結果を得やすくするための評価を行う処理であり、抽出した結果に対して、マッチモード定義の条件にかかわらず再度、単語要素、係りうけ、属性の観点で結果を評価することが有用だからである。

40

【0100】

例えば、係りうけマッチングにおいて、図 10 (b) のように、ノード単独タイプが行われていた場合、ルールに対応する文節が「(サッカー 観戦) - (行く)」であり、検索キー文に含まれる文節が「(サッカー) - (行く)」、「(サッカー) - (する)」で

50

あった場合、いずれもマッチング結果としては、一致すると判断されるが、マッチプロファイルにおける係りうけマッチ係数 $P A 7$ は、検索キー文の係りうけとして一致する「(サッカー) - (行く)」に対して、係数を適用してスコアが算出される。

つまり、検索処理部 501 は、このような係りうけマッチングの条件を満たす文章や単文が得られた場合、このマッチした文章や単文と、検索キー文とを比較して、スコアモード情報が示すような関係であった場合、係数を適用してスコアを算出することができる。

【0101】

そして、検索処理部 501 は、マッチした文章と、この文章のスコアに基づき、マッチした文章のマッチ情報（例えば、マッチング処理において利用したマッチモードの種類、マッチした単語やマッチした文節の文章内における出現位置（以下、マッチ位置情報という）、スコア）を生成する（ステップ S T 2 3）。

10

次いで、検索処理部 501 は、例えば、ステップ S T 2 2 によって算出されたスコアの点数が高い順にマッチした文章を並び替える（ステップ S T 2 4）。そして、検索処理部 501 は、検索キー文、マッチした文章、マッチした単文およびマッチ情報を関連付けて、メモリ領域 506 の検索結果オブジェクトに書き込む（ステップ S T 2 5）。

【0102】

そして、検索処理部 501 は、WEBサーバ 300 に検索結果オブジェクトの内容を送信する（ステップ S T 2 6）。WEBサーバ 300 は、受信した検索結果オブジェクトを記憶部 304 に一時的に記憶させ、クライアント端末装置 100 の表示部 102 によって表示可能な検索結果の表示データ（ウェブページ）を作成しクライアント端末装置 100 に送信する。クライアント端末装置 100 は、この表示データに基づき、検索結果の表示データを表示部 102 に表示する。

20

なお、マッチ位置情報とは、マッチング条件を満たす単語（マッチした単語）が、このマッチした単語を含むマッチした単文や文章の文中に出現する文字位置を表す情報である。

【0103】

ここで、図 15 を用いて、検索結果オブジェクトの内容である検索結果データの一例について説明する。図 15 は、検索結果データの一例について説明する概略図である。

図 15 に示す通り、検索結果データは、クライアント端末装置 100 においてユーザによって入力された“検索キー文”と、日本語解析サーバ 500 によるマッチング処理によって得られた検索キー文とマッチした“マッチした文章”と、検索キー文とマッチした単文であって、マッチした文章に含まれる単文であることを表す“マッチした単文”と、このマッチした単文に含まれる文節毎に生成される“マッチ情報”とが、それぞれ関連付けられている。

30

【0104】

次に、図 16 を用いて、検索結果オブジェクトの内容として、クライアント端末装置 100 に送信される検索結果データの一例について、より詳細に説明する。

図 16 (a) に示す通り、検索キー文が「インターネットが繋がらない」であって、日本語解析サーバ 500 によって、例えば、検索結果 1 として、単文 I D 「001 - 1」、テキスト「インターネットが繋がらない場合でも操作は可能ですか?」と、検索結果 2 として、単文 I D 「002 - 3」、テキスト「突然、インターネットができなくなりました」が得られた場合について以下説明する。

40

【0105】

図 16 (b) は、検索キー文や検索結果で得られた文章のマッチ位置情報を表す文字位置について説明する図である。図 16 (b) に示す通り、例えば、検索キー文は、文章の先頭から順番に、一文字ずつ「1, 2, 3, ..., 12」という文字位置を表す番号が付与されている。この文字位置を表す番号によって、マッチした単文や文章の文中に出現するマッチした単語の文字位置を表すことができる。

【0106】

図 16 (c) は、検索結果の一例を示す。図 16 (c) に示す通り、検索結果 1 は、文

50

章IDが「001」、単文IDが「001-1」であって、マッチモードが「係り受けマッチング」の場合では、スコアが「8.9」であって、マッチ位置情報が「key1:7, res1:7」「key9:12, res9:12」であることが示されている。ここで、マッチ位置は、当該マッチモードにおいてマッチングした単語を示しており、マッチ位置情報「key1:7, res1:7」は、検索キー文の「インターネット」を、マッチ位置情報「key9:12, res9:12」は、検索キー文の「つながらない」を、意味している。つまり、この「インターネット」と「つながらない」は、係り受けマッチングにおいて、検索キー文における親子関係と、検索結果1の文章における親子関係のマッチングの程度がスコア「8.9」と評価されていることを意味している。

【0107】

10

ここで、マッチ位置情報「key1:7, res1:7」とは、マッチした単語の文中における位置を表す情報であって、「key1:7」は、検索キー文の先頭から数えて、1番目から7番目までの文字列が、マッチした単語に該当することを表している。また、「res1:7」は、マッチした単文（あるいはマッチした文章）の先頭から数えて、1番目から7番目までの文字列が、マッチした単語に該当することを表している。なお、この数字は、文の先頭を基点として数えられる文字の数であって、文中における文字の位置を表す情報である。

このように、検索結果は、マッチング条件を満たす単語を含む文と、当該文に含まれるマッチした単語のマッチ位置情報とが関連付けられている情報を含む。

【0108】

20

次に、図17を用いて、本実施の形態に係る情報検索システム1における検索開始処理の一例について説明する。図17は、本実施の形態に係る情報検索システム1における検索開始処理の一例を示すフローチャートである。

図17に示す通り、例えば、クライアント端末装置100の入力部103が日本語解析サーバ500による検索サービスを利用するリクエストをユーザから受け付けると、クライアント端末装置100は、日本語解析サーバ500による検索用表示データを送信するよう、WEBサーバ300に対してリクエスト制御信号を送信する。

【0109】

WEBサーバ300は、通信部301を介して、クライアント端末装置100からリクエスト制御信号を受信すると、リクエスト処理部302が、このリクエスト制御信号に基づき、クライアント端末装置100の表示部102によって表示される表示データのウェブページのデータを作成するようデータ変換部303を制御する。次いで、データ変換部303は、記憶部304から必要な設定データ等を読み出し、ユーザによって検索キー文が入力されるテキストボックスを表示する検索用表示データを作成する。そして、通信部301は、この検索用表示データを、クライアント端末装置100に送信する（ステップST30）。

30

例えば、データ変換部303は、クライアント端末装置100に対して検索用表示データを表示させるためのHTML文章等で構成された表示データを作成する。そして、リクエスト処理部302は、この表示データと、検索結果を表示するためのルールが記載されたルール情報（例えば、CSSファイルで構成されたもの）、あるいは、検索結果をクライアント端末装置100の表示部102に表示されるために利用されるプログラム（例えば、javascript等）であって、ブラウザ101上で動作するプログラムコードを、通信部301を介して、クライアント端末装置100に送信する。

40

【0110】

クライアント端末装置100は、WEBサーバ300から表示データやプログラム等を受信すると、このプログラムを起動させる。そして、データ処理部112は、このプログラムに従って、WEBサーバ300から受信される検索用表示データを、表示部102によって表示される表示データを生成し、表示処理部113を制御する。そして、表示処理部113は、データ処理112によって生成された表示データを、表示部102に表示させる。

50

クライアント端末装置 100 の入力部 103 を介して、ユーザから特定の検索サービスが指定されると、クライアント端末装置 100 は、指定された検索サービスによる検索を実行するためのリクエスト制御信号を生成する。

また、ユーザによって検索キー文が入力されると、入力部 103 はこれを受け付ける（ステップ S T 3 1 ）。

【 0 1 1 1 】

次いで、クライアント端末装置 100 は、ユーザによって指定された検索サービスの種類や、入力された検索キー文を、検索リクエストメッセージとともに、通信部 104 を介して W E B サーバ 300 に送信する。

W E B サーバ 300 は、クライアント端末装置 100 から検索リクエストメッセージを受信すると、この検索リクエストメッセージから検索キー文を取り出し、日本語解析サーバ 500 に、例えば、検索サービス α による検索を要求する（ステップ S T 3 2 ）。

【 0 1 1 2 】

次に、図 18 を用いて、本実施の形態に係る情報検索システム 1 における検索結果の表示方法の一例について説明する。図 18 は、本実施の形態に係る情報検索システム 1 における検索結果の表示方法の一例を示すフローチャートである。

図 18 に示す通り、W E B サーバ 300 は、日本語解析サーバ 500 から検索結果を受信すると、記憶部 304 に一時的に記憶させる。そして、データ変換部 303 が、クライアント端末装置 100 に表示させるためのルール情報を記憶部 304 から読み出す。次いで、データ変換部 303 は、このルール情報に基づき、検索結果をクライアント端末装置 100 の表示装置 102 において表示するためのウェブページの表示データを作成し、検索結果のメッセージとして、通信部 301 を介して、クライアント端末装置 100 に送信する（ステップ S T 4 0 ）。

例えば、データ変換部 303 は、検索キー文ごとに、マッチした文章へのリンク情報、マッチした単文、適用したマッチモード、マッチ位置情報、スコア等のそれぞれに所定のタグを付与して、クライアント端末装置 100 側のデータ処理部 112 によって取り扱い可能なデータ（X M L ファイル）を作成し、検索結果として送信する。

【 0 1 1 3 】

クライアント端末装置 100 のデータ処理部 112 は、この検索結果を受信すると、この検索結果（X M L ファイル）を記憶部 111 に一時的に記憶させる。そして、データ処理部 112 は、記憶部 111 に記憶されているルール情報に基づき、この X M L ファイルの電文に含まれるマッチ位置に該当する単語に対して、適用したマッチモードに対応した表示用のタグを挿入する（ステップ S T 4 1 ）。

例えば、図 16（c）に示すような検索結果の場合、データ処理部 112 は、マッチ位置情報の「k e y 1 : 7、r e s 1 : 7」に該当する「インターネット」に対して、適用したマッチモードとして「係りうけマッチング」であることを表す表示用のタグを、検索結果を表示するためのルールが記載されたルール情報（C S S ファイル）を参照することで作成する。例えば、データ処理部 112 は、検索結果を表示するための情報として、特定の単語を強調して表示する強調表示設定情報をルール情報に基づき作成し、タグとしてマッチ位置情報に該当する単語に付与する。この強調表示設定情報としては、例えば、マッチした単語には下線を付加して表示するための設定情報や、あるいは、単語要素マッチングのマッチモードによってマッチした単語等を赤色で表示し、属性マッチングのマッチモードによってマッチした単語等を青色で表示することによって、マッチモードごとにユーザが視覚的に区別して認識することができるように表示するための設定情報が含まれている。

【 0 1 1 4 】

また、強調表示設定情報としては、マッチしたと判断されるマッチモードが複数ある場合、予めユーザによって設定されている優先順位に従って、優先的に表示させるように設定されているマッチモードとマッチした単語等を強調表示するようなものであってもよく、スコアの高い順に優先的に強調表示するようなものであってもよい。

10

20

30

40

50

さらに、検索結果として、上述のように予めユーザによってマッチモードの優先順位が決定されている場合、マッチした単文や文章が複数あれば、この優先順位に従って、マッチした単文等を表示させるものであってもよい。

また、検索キー文の文頭に近いマッチした単語から順番に、検索結果として、優先的に表示させるものであってもよい。

【0115】

このような強調表示設定情報に基づき、検索結果をクライアント端末装置100の表示部102に表示させる。このとき、ユーザによって、検索結果として表示されているマッチした単文等の中から、強調して表示されているマッチした単語等が選択されると、入力部102は、これを受け付ける。そして、データ処理部112は、記憶部111に一時的に記憶されている検索結果のマッチ位置情報から、ユーザによって選択されたマッチした単語等の位置を特定し、このユーザによって選択されたマッチした単語等に応じた絞り込み検索をさらに行う(ステップS T 4 2)。

10

【0116】

ここで、絞り込み検索について、図19、20を用いて、詳細に説明する。図19は、検索結果表示データに基づき、クライアント端末装置100の表示部102に表示される検索結果を表す画像の一例を示す概略図である。

図19に示す通り、クライアント端末装置100の表示部102は、検索結果表示データに基づき画面102Aを表示し、この画面102Aの左側には、検索キー文を表示するテキストボックス102Bが、右側には検索結果を表示するサジェスト画面102Cが表示されている。

20

テキストボックス102Bには、検索キー文「クレジットカードの支払方法を登録したのですが料金サポート窓口から請求書が届きます。」が表示されている。

【0117】

サジェスト画面102Cには、この検索キー文102C1と、この検索キー文に基づき日本語解析サーバ500による検索が行われた検索結果であるマッチした単文102C2が表示されている。

例えば、サジェスト画面102Cでは、検索キー文102C1のマッチした単語が、それぞれマッチモードに応じた色ごとに強調表示されている。マッチした単文102C2では、複数の単文が表示されている場合、通常の実検結果として、マッチした単文が、スコアが高い順番で表示されている。また、マッチした単文102C2に含まれるマッチした単語(例えば、「クレジットカードで支払方法を登録したのですが、料金サポート窓口から請求書が届きます。」)は、検索キー文102C1に含まれるマッチした単語と同様な強調して表示されており、同じマッチモードによって検索された単語に対しては、同じ色で強調して表示されている。

30

なお、サジェスト画面102Cの検索キー文102C1において強調して表示されているマッチした単語は、クライアント端末装置100の入力部102からの選択指示を受け付け、ユーザによって選択可能である。

WEBサーバ300は、検索結果が得られた場合、図19に示すような検索結果の表示データを作成し、クライアント端末装置100に送信する。

40

【0118】

次に、図20を用いて、図19に示す画面から絞り込み検索を行う方法について説明する。図20は、図19に示す画面から、絞り込み検索を行った後に表示される画面の一例を示す概略図である。なお、クライアント端末装置100は、日本語検索サーバ500から受信した検索結果(例えば、マッチした文章、マッチした単文、マッチした単語、これらのマッチングに用いられたマッチング条件、あるいはマッチ位置情報を含むマッチ情報やスコア等)や、絞り込み検索を実行する際の検索ルール情報(例えば、入力部103を介して指定されたマッチした単語を検索キーとして、検索結果から、絞り込み対象を検索するためのプログラムや設定値等)を記憶部111に記憶しておく。図20に示す通り、例えば、サジェスト画面102Cの検索キー文102C1のうち、強調して表示されてい

50

るマッチした単語の「クレジットカード」が、ユーザによって選択された場合（例えば、マウスによって画面上のポインタが「クレジットカード」と表示されている部分を指示した状態で、ダブルクリック等の操作によって絞り込み検索の対象として指定された場合）、クライアント端末装置 100 のデータ処理部 112 が、入力部 103 を介して、ユーザからの選択指示を受け付け、「クレジットカード」による絞り込み検索を実行する。

【0119】

データ処理部 112 は、記憶部 111 に記憶されている検索結果や検索ルール情報を参照して、検索キー文の「クレジットカード」のマッチ位置情報を検出する。また、データ処理部 112 は、「クレジットカード」のマッチ位置情報に基づき、検索キー文の「クレジットカード」とマッチする単語としてマッチ位置情報において関連付けられているマッ

10

チした単語を含むマッチした単文や文章を検索する。

例えば、図 19 に示す場合、検索キー文内の「クレジットカード」は、マッチ位置が「1:8」であるため、検索結果において、マッチ位置情報が「key 1:8, res 1:8」であって、同じマッチモードによってマッチした単語を含むマッチした文章を検索する。

さらに、データ処理部 112 は、この検索によって得られたマッチした単文や文章を、検索結果表示データの検索結果の上位に表示する表示データ（画面 102A-1、図 20 の上方に示す図を参照）を作成する。これにより、表示処理部 112 は、絞り込み検索によって得られたマッチした文章を上位に表示するような表示データを、表示部 102 に表示させる。

20

【0120】

そして、表示部 102 は、図 20 の上部に示す検索結果の画面 102A-1 の通り、検索結果のマッチした単文 102C2 の最上位として「クレジットカード支払いへの変更について」（スコア：2.2）を、その次に、「クレジットカードで支払方法を登録したのですが、料金サポート窓口から請求書が届きます。」（スコア：1.4）を表示する。

これにより、クライアント端末装置 100 は、検索結果のマッチ位置情報を利用して、マッチした単語に基づく再検索を実行することができ、マッチした単語に関してマッチングの程度が高い順に、マッチした単文を表示することができる。

なお、ここでは、絞り込み検索として、検索キー文のマッチした単語が指定された場合、単語要素マッチングにおいてマッチした単語であって、かつ、マッチ位置情報が一致する単語を、データ処理部 112 が、記憶部 111 に記憶されている検索結果に基づき、再検索する一例について説明したが、本発明はこれに限られない。

30

【0121】

例えば、サジェスト画面 102C の検索キー文 102C1 のうち、強調して表示されているマッチした単語の「請求書」が、ユーザによって選択された場合、データ処理部 112 は、入力部 103 を介して、ユーザからの選択指示を受け付け、「請求書」による絞り込み検索を実行する。

データ処理部 112 は、記憶部 111 に記憶されている検索結果や検索ルール情報を参照して、検索キー文の「請求書」のマッチ位置情報を検出する。また、データ処理部 112 は、「請求書」のマッチ位置情報に基づき、検索キー文の「請求書」とマッチする単語としてマッチ位置情報において関連付けられているマッチした単語を含むマッチした単文や文章を検索する。

40

例えば、図 19 に示す場合、検索キー文内の「請求書」は、マッチ位置が「33:35」であるため、検索結果において、マッチ位置情報が「key 33:35, res 33:35」であって、同じマッチモードによってマッチした単語を含むマッチした文章を検索する。

【0122】

そして、データ処理部 112 は、この検索によって得られたマッチした文書を、検索結果表示データの検索結果の上位に表示する表示データ（画面 102A-2、図 20 の下方に示す図を参照）を作成する。表示処理部 113 は、絞り込み検索によって得られたマッ

50

チした文章を上位に表示するような表示データを、表示部 102 に表示させる。

表示部 102 は、この表示データを表示し、図 20 の下部に示す検索結果の画面 102A-2 の通り、検索結果のマッチした単文 102C2 の最上位として「料金センタから請求書が届くのですが、請求書明細の内訳について確認したいです。」(スコア: 4.0) を、その次に、「解約したにも関わらず、請求書が届いたのですが。」(スコア: 2.5) ……を表示する。

【0123】

このように、検索結果をクライアント端末装置 100 の記憶部 111 に記憶させておき、マッチング位置情報を利用することで、データ処理部 112 は、マッチした単語に基づく再検索を行うことができる。これにより、クライアント端末装置 100 は、絞り込み検索がユーザによって指示(リクエスト)された場合、検索結果に対して、形態素分解や構文解析等の解析を行うことなく、文章内におけるマッチした単語の位置を、マッチ位置情報を利用して得ることができる。また、マッチ位置情報は、マッチモードごとに作成されているため、マッチモードに応じて異なる強調表示をするためのタグをマッチした単語に付与することができる。従って、クライアント端末装置 100 は、再検索された検索結果を、表示部 102 に表示することで、絞り込み検索の検索結果を得ることができる。

【0124】

このように、マッチ位置や、マッチパターンを含む検索結果をクライアント端末装置 100 の記憶部 111 に記憶しておき、かつデータ処理部 112 による絞り込み検索のキーとなる単語が入力部 103 を介して指定されると、対応する単語の位置情報とマッチパターンのみから簡単に絞り込みの検索結果を得ることができる。このため、クライアント端末装置 100 は、絞り込み検索による再検索結果の表示データを再構築できる。

一方、本実施の形態と異なり、マッチ位置情報が検索結果として記憶部 111 に記憶されていない場合、クライアント端末装置 100 によって絞り込み検索が指定された単語が、マッチした文章やマッチした単文の何処に含まれているのかを、文章を解析しなければ検出することができない問題があったと考えられる。また、本実施の形態のように、マッチモードごとにマッチした単語のマッチ位置情報が検索結果としてクライアント端末装置 100 に記憶されていない場合、マッチモードごとに同じ色、あるいは同じフォント等で強調して表示できない問題があったと考えられる。本実施の形態に係る情報検索システム 1 は、上述のような構成とすることによって、上記問題を解決することができる。

【0125】

また、本実施の形態に係るクライアント端末装置 100 は、マッチ位置情報を用いて、マッチした単文や文章の文中におけるマッチした単語の位置が分かるため、マッチした単語を文中から抽出することができる。このため、形態素解析や構文解析等の文書解析を行うことなく、再検索ができる。また、マッチした単語の位置が分かることにより、クライアント端末装置 100 は、マッチした単語を強調表示する表示データを作成し、表示することができる。

一方、マッチ位置情報がない場合、クライアント端末装置 100 は、検索結果に対して文書解析を行わない限り、強調表示するための単語が文中のどこにあるのかわからないため、強調表示することができない。また、マッチ条件に応じて異なる色で強調表示することもできない。

なお、上述の処理はクライアント端末装置 100 側のデータ処理部 112 で絞り込み検索を実施し、再表示を行ったが、クライアント端末装置 100 から、ユーザによって選択された単語情報をWEBサーバ300側に送信し、WEBサーバ300側に絞り込みの処理を実施させてもよい。この場合、検索結果や検索ルール情報は、日本語解析サーバからWEBサーバ300に送信され、記憶部304に記憶しておく。

【0126】

次に、マッチング処理とスコアリング処理の一例について詳細に説明する。

なお、ここでは、検索サービスαがユーザによって指定されており、検索サービスαのマッチプロファイルとして予め決められているマッチプロファイルAがメモリ領域506

10

20

30

40

50

の辞書オブジェクトに展開されている例について説明する。

【 0 1 2 7 】

図 2 1 は、検索キー文の一例を示す図である。

図 2 1 に示すように、検索キー文「 E T C カードがお店で使えません。」が、日本語解析サーバ 5 0 0 に入力される。これにより、日本語解析サーバ 5 0 0 によって照合が行われると、図 2 2 に示すようなマッチした単文が得られる。図 2 2 には、マッチした単文が複数示されており、例えば「 E T C カードが使えない」「クレジットカードが使えない」「 E T C カードを使いたい」「クレジットカードを使いたい」「 E T C カードは使いやすい」「 E T C カードを無くした。」等のマッチした単文が検索によって得られている。

【 0 1 2 8 】

図 2 3 は、マッチプロファイル記憶部 5 0 2 に記憶されているマッチプロファイル A ~ C の設定を説明するための概略図である。図 2 3 に示す通り、マッチプロファイル A ~ C の設定は、図 5 に示したマッチプロファイル記憶部 5 0 2 に記憶されているマッチモードの組み合わせが決定されているマッチモード定義 P A 1 を含むマッチモード情報と、相対出現頻度フラグ P A 2 から同義語マッチ係数 P A 1 0 に相当する情報を含むスコアモード情報である。なお、マッチモードは、図 2 7 を用いて後述するとおり、(1) ~ (4) に示す組み合わせパターンが利用可能である。ここでは、(2) 単語要素マッチング + 属性マッチングがマッチプロファイルのマッチ情報として予め決められている例について説明する。

【 0 1 2 9 】

マッチプロファイル A は、マッチモード情報 (マッチモード定義 P A 1) として、(2) 単語要素マッチングと属性マッチングが行われることが予め決められており、マッチング処理の結果に対しての重み付けは実行されないこと (例えば、係数 1 . 0) が予め決められている。また、マッチプロファイル B は、マッチモード情報 (マッチモード定義 P A 1) として、単語要素マッチングと属性マッチングが行われることが予め決められており、マッチング処理の結果に対しての重み付けは、係りつけマッチ係数 P A 7 に基づく係りつけマッチングにおける重み付けを実行すること (例えば、係数 2 . 0) が予め決められている。さらに、マッチプロファイル C は、マッチモード情報 (マッチモード定義 P A 1) として、単語要素マッチングと属性マッチングが行われることが予め決められており、マッチング処理の結果に対しての重み付けは、述語属性マッチ係数 P A 6 に基づき、属性マッチングにおいてマッチした単文には述語属性マッチ係数に従った重み付けを実行すること (例えば、係数 2 . 0) が予め決められている。

【 0 1 3 0 】

次に、図 2 4 を用いて、図 2 1 ~ 2 3 に示した例において、マッチプロファイル A に基づき得られた検索結果について説明する。図 2 4 は、マッチプロファイル A に基づき得られた検索結果をサジェスト画面 1 0 2 C に表示されている画面の一例を示す概略図である。

ここで、マッチプロファイル A のマッチモード定義 P A 1 は単語要素 + 属性であるから、図 2 4 にはこれらの条件に一致した結果 (単文) が抽出されている。

図 2 4 に示すスコアリング処理において、スコアモード情報としては、特に重み付け等は設定されていない (例えば、係数 0 . 0 もしくは 1 . 0) 。また、各ノード (ルール) がマッチした場合は、基準点を付与し、さらに、スコアモード情報による重み付けの計算を行う。なお、本実施例では基準点は、マッチ辞書内に記憶された各ルール (ノード) の重み値が適用されるが、単純に装置全体で所定の値を設定してもよい。今回は基準点を 1 . 2 点として、説明する。

例えば、「 E T C カードが使えない」は、「 E T C 」「カード」「使え」という単語がマッチし、さらに、「 E T C 」「カード」は名詞なので、「属性なし」で、それぞれ 1 . 2 点、「使え」は「使う (否定) 」の属性がマッチしているので、1 . 2 点 $\times$ 1 . 0 (述語属性マッチ係数) で、1 . 2 点となり、3 . 6 点である。また「 E T C 」と「使え」、「カード」と「使え」のそれぞれに係り受けペアが含まれる事から、係り受けマッチ係

10

20

30

40

50

数なし(1.0)が適用され1.2点×1.0×2で2.4点が加算され、合計で6.0点である。

以上のように、マッチモードでは係り受けマッチングは行われていないが、マッチプロファイルのスコアモードで、係り受けマッチ係数を定義しておけば、得られた結果に対して、スコアの算出において、柔軟に重み付けをすることができる。

【0131】

一方、「ETCカードを使いたい」は、「ETC」「カード」という単語・属性なしがマッチしている。この場合、1.2点+1.2点により、2.4点である。また、述語属性マッチ係数・係り受けマッチ係数が適用されるものはないため、その分の加点はない。

また、「クレジットカードが使えない」は、「使え」という単語と属性「使う(否定)」がマッチしている。この場合、1.2点×1.0で1.2点である。また、係り受けマッチ係数が適用されるものはないため、その分の加点はない。よって、図24に示す通り、スコア順に並べられた検索結果は、最上位が「ETCカードが使えない」(スコア6.0点)、その次が、「ETCカードを使いたい」(スコア2.4点)となる。

なお、本実施例では、「クレジットカード」と「ETCカード」は、「カード」の部分で共通しているが、「クレジット」単独での形で形態素解析をおこなう際のシステム辞書5501に登録がなければ、単語要マッチングにおいて、全要素のタイプが実行されているため、両者はマッチした文節として得られない。

【0132】

次に、図25を用いて、図21~23に示した例において、マッチプロファイルBに基づき得られた検索結果について説明する。図25は、マッチプロファイルBに基づき得られた検索結果をサジェスト画面102Cに表示されている画面の一例を示す概略図である。

図25に示すスコアリング処理において、スコアモード情報としては、係り受けマッチ係数が、“あり”(点数に2.0を乗じる)に予め決定している。例えば、「ETCカードが使えない」は、「ETC」「カード」「使え」という単語・属性がマッチし、プロファイルAと同様に3.6点である。また「ETC」と「使え」、「カード」と「使え」のそれぞれに係り受けペアが含まれることから、係り受けマッチ係数あり(2.0)が適用され1.2点×2.0×2で4.8点が加算され、合計で7.2点である。なお、その他のマッチした単文に対してのスコア処理は図24での説明と同様であるために、詳細な説明は省略する。

【0133】

次に、図26を用いて、図21~23に示した例において、マッチプロファイルCに基づき得られた検索結果について説明する。図26は、マッチプロファイルCに基づき得られた検索結果をサジェスト画面102Cに表示されている画面の一例を示す概略図である。

図26に示すスコアリング処理において、スコアモード情報としては、述語属性マッチ係数があり(点数に2.0を乗じる)に予め決定している。例えば、「ETCカードが使えない」は、「ETC」「カード」という単語・属性なしがマッチし、1.2点+1.2点で2.4点、さらに「使え」が単語、「使う(否定)」の属性でマッチしているので述語属性マッチ係数あり(2.0)が適用され1.2点×2.0で2.4点である。

また、プロファイルAと同様に「ETC」と「使え」、「カード」と「使え」のそれぞれに係り受けペアが含まれることから、係り受けマッチ係数なし(1.0)が適用され1.2点×1.0×2で2.4点が加算され、合計で7.2点である。一方、「クレジットカードが使えない」は「使え」が単語、「使う(否定)」の属性でマッチしているだけなので、述語属性マッチ係数あり(2.0)が適用され1.2点×2.0で2.4点である。

【0134】

このように、異なるスコアモード情報が設定されているマッチプロファイルA~Cを用いて、異なる検索結果を得ることができる。

これにより、利用するマッチモードに応じて、マッチ辞書データを作成したり、マッチ辞書データをマッチモードに応じて作成したりする必要がなく、1つのマッチ辞書データを利用して、複数のマッチモードやその組み合わせによる検索を行うことができる。

【0135】

次に、マッチモードに応じた検索結果の特徴について、図27～31を用いて、説明する。

図27は、検索結果を示し、マッチした単文と、このマッチした単文が得られたマッチモードの組み合わせを示す図である。

例えば、「ETCカードが使えない」は、(1)単語要素マッチング、(2)単語要素マッチングと属性マッチングの組み合わせ、(3)係りうけマッチング、(4)係りうけマッチングと属性マッチングの組み合わせの全てにおいて、マッチした単文として得られたものであることを示している。なお、検索キー文は「ETCカードがお店で使えません。」である。

【0136】

図28は、マッチプロファイルのマッチモード定義PA1において、(1)単語要素マッチングによって得られた検索結果を示す概略図である。図28に示す通り、マッチした単文としては、「ETCカードが使えない」と「ETCカードを使いたい」のように、意味が正反対の文章もマッチしている。一方、「ETCカードを失くした。」等の「ETCカード」という単語だけマッチしている文章や、「クレジットカードが使えない」等の「使え」という単語だけマッチしている文章も、マッチした文章として検索によって得られる。

このように、このマッチプロファイルでは、広い範囲にわたって、類似文章も含めて集めてくる検索ができる。よって、複数のマッチした単文が得られる。ここで、複数のマッチした単文が膨大に得られる場合、却ってユーザの利便性が害される場合がある。しかし、日本語解析サーバ500は、マッチした単語の重み付けを行うことによって、スコアに基づき、評価されたマッチした単文を検索結果として表示することができる。よって、広い範囲にわたって複数の類似文章を検索した場合であっても、これら検索結果に優先順位を与え、スコアに基づき優先順位に応じた順番で表示することができる。

【0137】

図29は、マッチプロファイルのマッチモード定義PA1において、(2)単語要素マッチングと属性マッチングの組み合わせによって得られた検索結果を示す概略図である。図29に示す通り、マッチした単文としては、「ETCカードが使えない」と「ETCカードを使いたい」のように、意味が正反対の文章もマッチしている。ただし、ここでは「使え」は「使う(否定)」の属性としてマッチするため、両者のスコアの差分は、図28に示した例に比べて大きくできる。このように、マッチする単文の属性に対してもスコアの重み付けを行うことによって、検索キー文と意味の近いマッチした単文を、上位に表示することができる。

【0138】

図30は、マッチプロファイルのマッチモード定義PA1において、(3)係りうけマッチングによって得られた検索結果を示す概略図である。図30に示す通り、マッチした単文としては、「ETCカードが使えない」、「ETCカードを使いたい」、「ETCカードが使いやすい」といったように、意味は異なるが、“ETCカードを使うこと”に関連する単文がマッチしている。このようなマッチングモードを使用した場合、属性に係らず、係りうけ関係を重視した検索を行うことができる。

【0139】

また、図31は、マッチプロファイルのマッチモード定義PA1において、(4)係りうけマッチングと属性マッチングの組み合わせによって得られた検索結果を示す概略図である。図31に示す通り、マッチした単文としては、「ETCカードが使えない」というような、検索キー結果に対してかなり近い意味の単文を検索により得ることができる。これは、係りうけマッチングのノードの親子関係のタイプで、係りうけ関係の親ノードと子

ノードがそれぞれマッチし、かつ、係りうけ関係の親ノードあるいは子ノードに対応する単語に属性がマッチする場合にのみ検索結果が得られるものである。

【0140】

よって、例えば、コールセンターにおけるオペレーターのように、お客様の問い合わせに対して的確な回答を短時間で検索し、回答する必要がある場合複数のマッチモードの組み合わせによって、係りうけ関係や単語に含まれる属性に応じた、より少ない数の検索結果に絞り込むことができ、例えば、どのような商品に対するどのような要望あるいはクレームなのか、どのようなことが分かり難いための問い合わせなのか等の、細かい意味の違いに応じた検索を行うことができる。

【0141】

上述の通り、マッチードは、それぞれ異なる性格を有しているため、検索の目的に応じて使用するマッチモードやその組み合わせを選択することによって、わざわざ辞書データを作りかえることなく、ユーザに応じた検索を実現することができる。

また、従来技術に比べて、一つの辞書データに複数のプロファイルを切替えて使用することで、全体として辞書データのデータ量を大幅に抑制することができる。

本実施の形態に係る日本語解析サーバ500は、上述の通り、マッチプロファイル毎に、利用するマッチモードを登録するマッチモード定義PA1を備え、また、マッチプロファイル毎にスコアモードの条件や係数を調整できるように、相対出現頻度フラグPA2～同義語マッチ係数PA10を備えるようにした。これにより、本実施の形態に係る日本語解析サーバ500は、マッチプロファイルに応じて決定されているマッチモードを利用することができる。よって、マッチ辞書記憶部503は、マッチモードごとに作成する必要がなく、1つのマッチ辞書データを利用して、異なるマッチモードでのマッチングを行うことができる。従って、マッチモードごとに、マッチ辞書データを作成する必要がなく、複数のマッチ辞書データを記憶するための記憶容量を削減することができる。また、マッチモード毎にマッチ辞書データを作成する労力が削減される。さらに、マッチモードは、ユーザによって検索サービスが選択されると、関連付けられているマッチモードが一義的に決定されるため、ユーザによって検索サービスが変更されることによって、簡単にマッチモードを変更することができる。

【0142】

また、本実施の形態に係る日本語解析サーバ500の検索処理部501は、マッチングモード情報を利用してマッチング条件を満たす単文や文章を得たうえで、さらに、得られたマッチした文章や単文に対して、スコアを算出するようにした。これにより、マッチした文章や単文の中から、検索キー文との関係において、意味合いがより近いものを、スコアによって表すことができる。

よって、マッチモード情報を利用して、マッチング条件に応じた検索結果に絞り込み、その上で、マッチした検索結果に対して、検索キー文の属性や文の構成（係りうけ関係）に応じたスコアを算出し、得られたスコアに基づいて、マッチした文章や単文の優先順位をつけることができる。これにより、検索キー文に、より近い意味合いのマッチした文章や単文を得ることができる。

さらに、本実施の形態に係る日本語解析サーバ500の検索処理部501は、マッチングモード情報とスコアモード情報とは、関連付けて決められることなく、独立して設定するようにした。これにより、検索キー文の特性や検索の目的や条件に応じた、より多面的な検索を実現することができる。

【0143】

なお、本実施形態に係るクライアント端末装置100は、例えば、コールセンターなどでの入力業務に用いる入力端末装置であって、ワークステーションやパーソナルコンピュータなどの情報処理装置から構成されるものが好ましい。また、本実施形態に係るクライアント端末装置100にあつては、携帯電話のユーザサポート業務を行うコールセンターにおいて、ユーザに対する電話応答時に頻出する語彙を基に、マッチ辞書データが構成されているものであってもよい。

10

20

30

40

50

これにより、電話による対応の内容をリアルタイムに入力する必要のあるコールセンターの作業などにおいて、より効率的な文字入力を実現することができる。

【 0 1 4 4 】

また、上述の情報検索システム 1 における動作の過程は、コンピュータに実行させるためのプログラムや、このプログラムとしてコンピュータ読み取り可能な記録媒体として利用可能であり、コンピュータシステムが読み出して実行することによって、上記処理が行われる。なお、ここでいう「コンピュータシステム」とは、CPU 及び各種メモリや OS、周辺機器等のハードウェアを含むものである。

また、「コンピュータシステム」は、WWWシステムを利用している場合であれば、ホームページ提供環境（あるいは表示環境）も含むものとする。

10

また、「コンピュータ読み取り可能な記録媒体」とは、フレキシブルディスク、光磁気ディスク、ROM、フラッシュメモリ等の書き込み可能な不揮発性メモリ、CD-ROM等の可搬媒体、コンピュータシステムに内蔵されるハードディスク等の記憶装置のことをいう。

【 0 1 4 5 】

さらに「コンピュータ読み取り可能な記録媒体」とは、インターネット等のネットワークや電話回線等の通信回線を介してプログラムが送信された場合のサーバやクライアントとなるコンピュータシステム内部の揮発性メモリ（例えばDRAM（Dynamic Random Access Memory））のように、一定時間プログラムを保持しているものも含むものとする。

また、上記プログラムは、このプログラムを記憶装置等に記憶したコンピュータシステムから、伝送媒体を介して、あるいは、伝送媒体中の伝送波により他のコンピュータシステムに伝送されてもよい。ここで、プログラムを伝送する「伝送媒体」は、インターネット等のネットワーク（通信網）や電話回線等の通信回線（通信線）のように情報を伝送する機能を有する媒体のことをいう。

20

また、上記プログラムは、前述した機能の一部を実現するためのものであっても良い。さらに、前述した機能をコンピュータシステムに既に記録されているプログラムとの組合せで実現できるもの、いわゆる差分ファイル（差分プログラム）であっても良い。

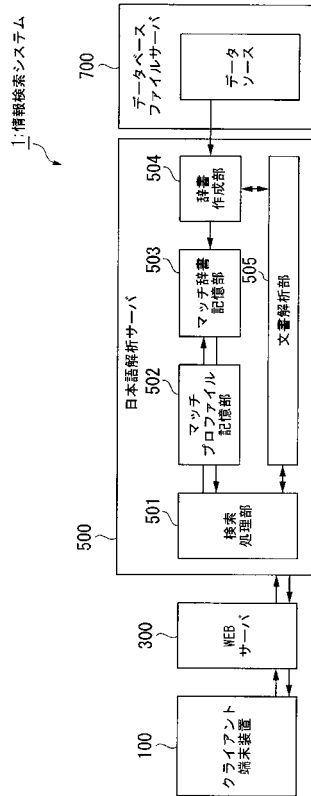
【符号の説明】

【 0 1 4 6 】

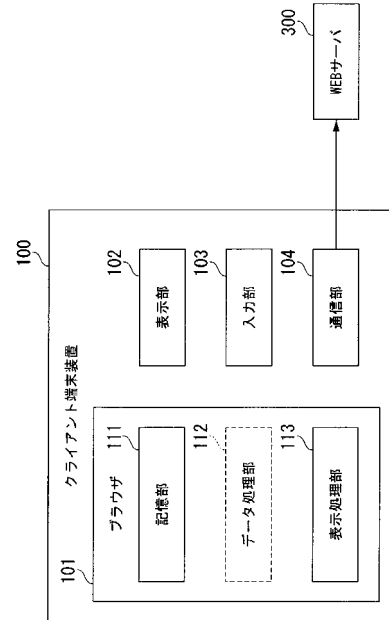
- 1 情報検索システム
- 1 0 0 クライアント端末装置
- 3 0 0 WEBサーバ
- 5 0 0 日本語解析サーバ
- 7 0 0 データベースファイルサーバ

30

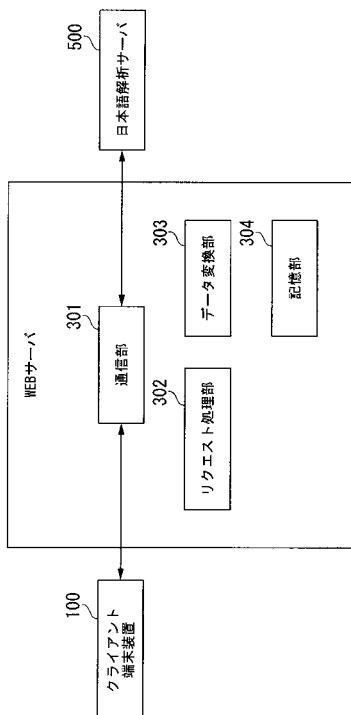
【図 1】



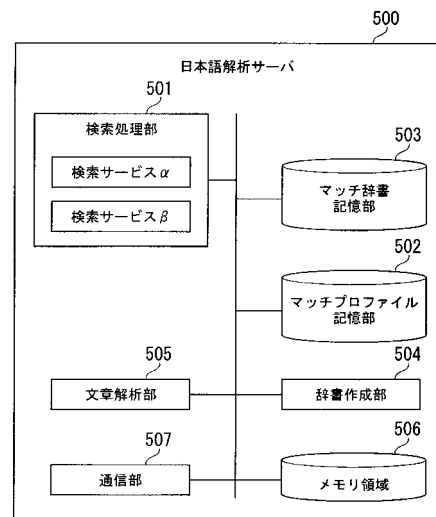
【図 2】



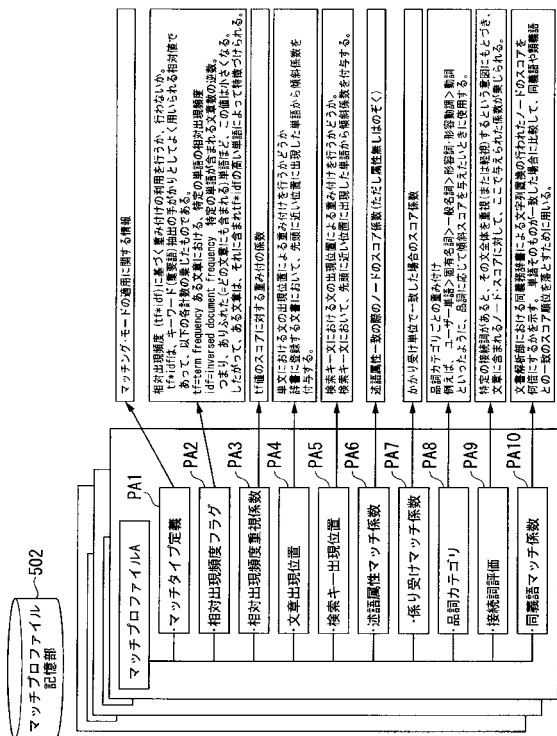
【図 3】



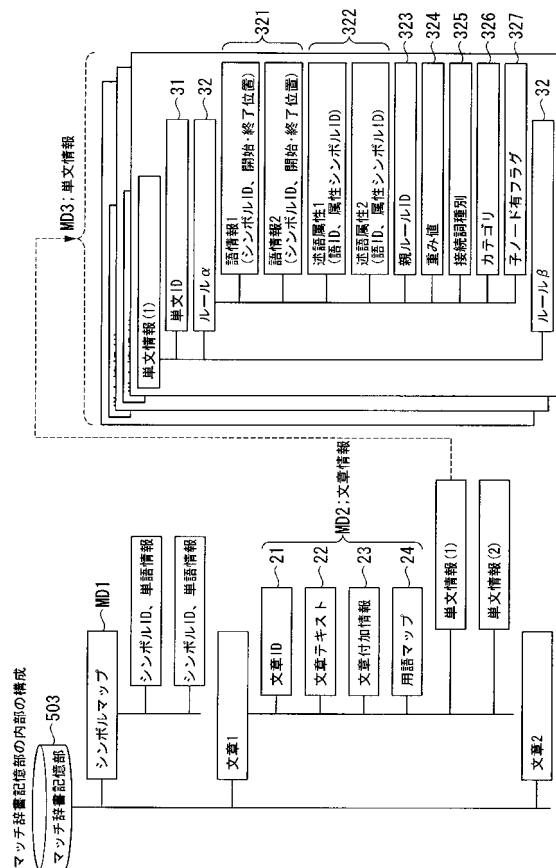
【図 4】



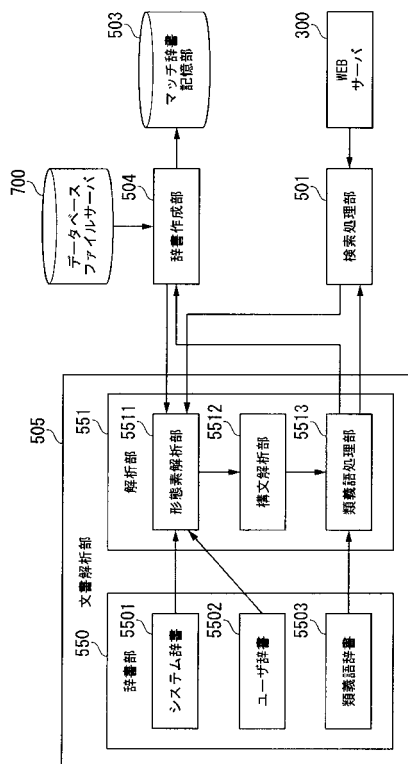
【 図 5 】



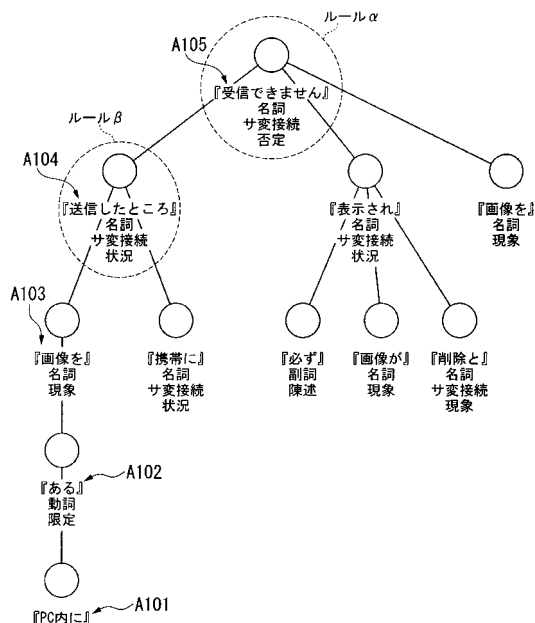
【 図 6 】



【圖 7】



【圖 8】



【図 9】

(1) 積集合タイプ

ルールに対応する文節	検索キー文に含まれる文節	結果
(a) (サッカー 観戦)	(サッカー 少年)	○
	(観戦 ツアー)	○
	(戦)	×

(2) 全要素タイプ

ルールに対応する文節	検索キー文に含まれる文節	結果
(b) (サッカー 観戦)	(サッカー 観戦)	○
	(観戦)	×

(3) 部分集合タイプ

ルールに対応する文節	検索キー文に含まれる文節	結果
(c) (サッカー 観戦)	(サッカー 観戦 ツアー)	○
	(観戦 ツアー)	×

<単語要素マッチングの例>

【図 10】

(1) ノード親子関係タイプ

ルールに対応する文節	検索キー文に含まれる文節	結果
(a) (サッカー 観戦) - (行く)	(サッカー) - (行く)	○
	(サッカー 観戦)	×
	(行く)	×

(2) ノード単独タイプ

ルールに対応する文節	検索キー文に含まれる文節	結果
(b) (サッカー 観戦) - (行く)	(サッカー)	○
	(行く)	○
	(する)	×

<係り受けマッチングの例>

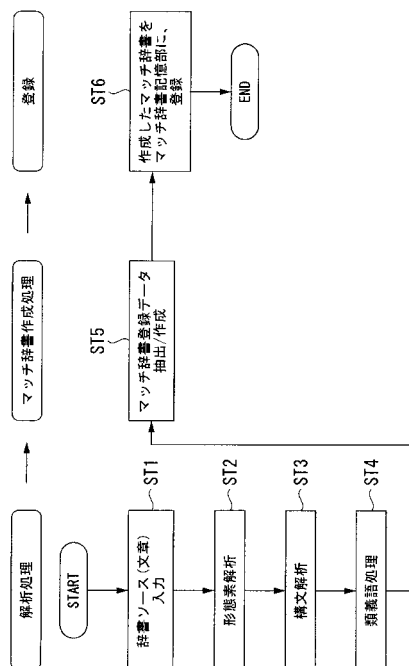
【図 11】

(1) 文属性一致

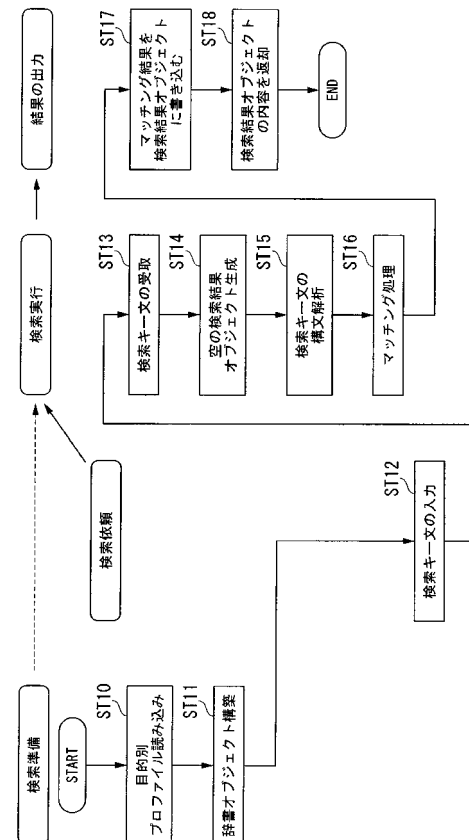
ルールに対応する文節	検索キー文に含まれる文節	結果
(a) (サッカー 観戦) + 否定	(観戦) + 否定	○
	(サッカー 観戦) + 可能	×

<属性マッチングの例>

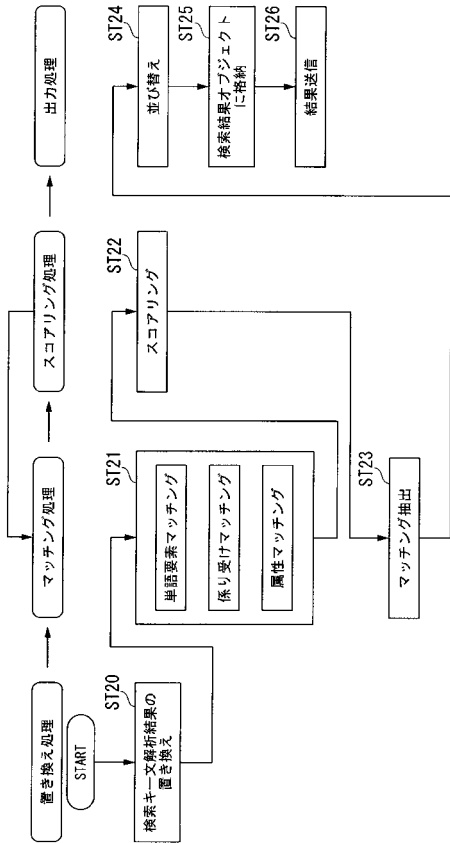
【図 12】



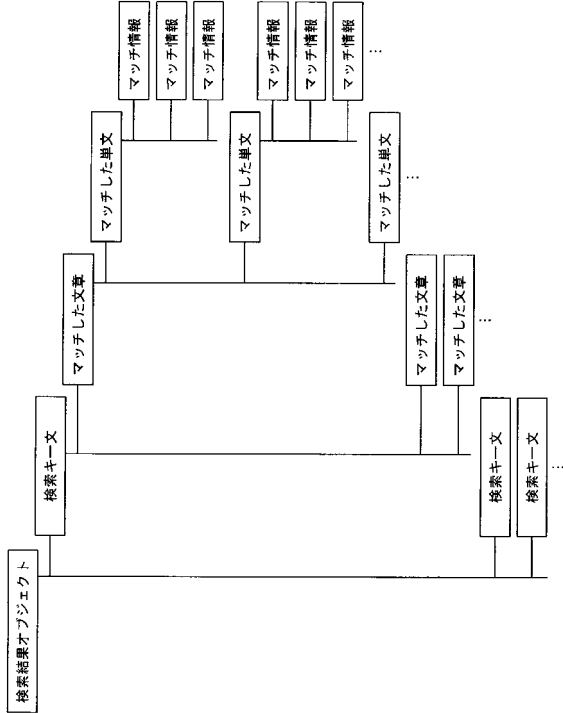
【図 13】



【図 1 4】



【図 1 5】



【図 1 6】

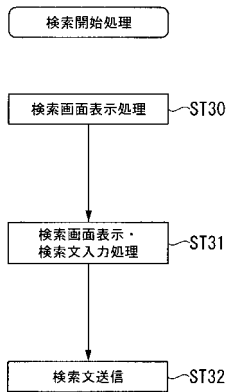
(a) 検索文：インターネットがつながらない
結果：
検索結果1：001-1 インターネットがつながらない場合でも操作は可能ですか？
検索結果2：002-3 突然、インターネットができなくなりました

(b) マッチ位置の考え方
インターネットがつながらない
1 2 3 4 5 6 7 8 9 10 11 12 13 14

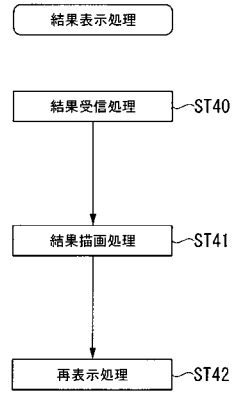
(c)

文章ID	単文ID	マッチモード	スコア	マッチ位置
001	001-1	係り受け	8.9	key1:7, res1:7
		属性		key9:14, res9:14
		単語要素		key9:14, res9:14
002	002-3	単語要素	1	Key1:7, res1:7
		単語要素		Key1:7, res4:10

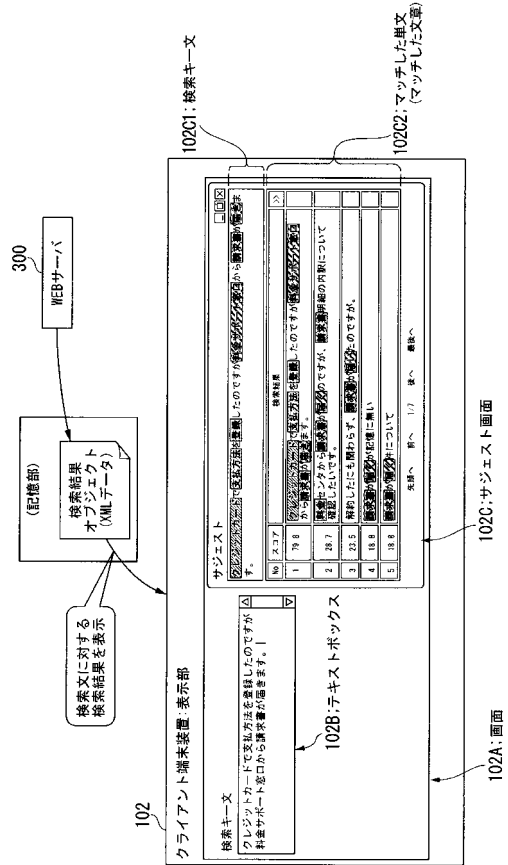
【図 1 7】



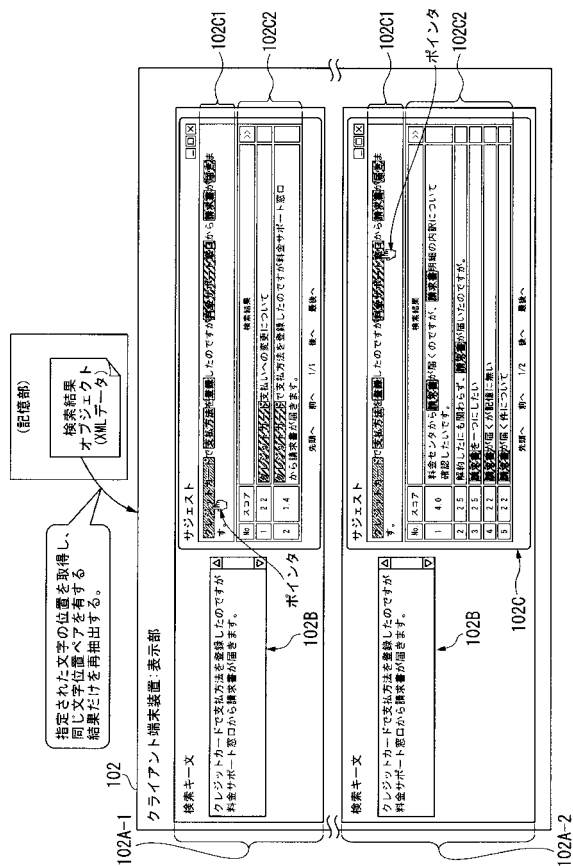
【 図 1 8 】



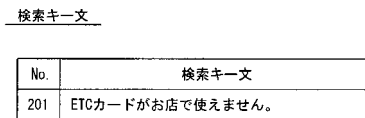
【 図 1 9 】



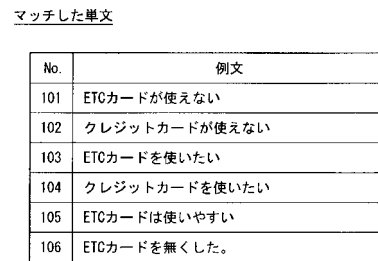
【 図 2 0 】



【 図 2 1 】



【 図 2 2 】



【図 2 3】

スコアモード										
No	マッチ ファイル	マッチ ファイル	文庫集合における 単語の 重み付け	文庫 出現位置	検索 出現位置	逆順単語マッチ 結果	係り受けマッチに おける 重み付け	品詞カテゴリに おける 重み付け	指定期間における 重み付け	同語反意語 重み付け
301	A	(2)	なし(0)	なし(0)	なし(0)	なし(1.0)	なし(1.0)	なし (すべての品詞 カテゴリで1.0を指定)	なし (すべての品詞で 1.0を指定)	なし(1.0)
302	B	(2)	なし(0)	なし(0)	なし(0)	なし(1.0)	あり(2.0)	なし (すべての品詞 カテゴリで1.0を指定)	なし (すべての品詞で 1.0を指定)	なし(1.0)
303	C	(2)	なし(0)	なし(0)	なし(0)	あり(2.0)	なし(1.0)	なし (すべての品詞 カテゴリで1.0を指定)	なし (すべての品詞で 1.0を指定)	なし(1.0)

PA1

PA2
(PA3)

PA4

PA5

PA6

PA7

PA8

PA9

PA10

PA1 PA2 (PA3) PA4 PA5 PA6 PA7 PA8 PA9 PA10

(38)

JP 5439028 B2 2014.3.12

【図 2 4】

サジェスト

検索結果

No	スコア	検索結果
1	6.0	検索結果が低い
2	2.4	検索結果が低い
3	2.4	検索結果が低い
4	2.4	検索結果が低い
5	1.2	検索結果が低い

先頭へ 前へ 1/1 後へ 最後へ

<マッチプロファイルAによるマッチング結果>

【図 2 5】

サジェスト

検索結果

No	スコア	検索結果
1	7.2	検索結果が低い
2	2.4	検索結果が低い
3	2.4	検索結果が低い
4	2.4	検索結果が低い
5	1.2	検索結果が低い

先頭へ 前へ 1/1 後へ 最後へ

<マッチプロファイルBによるマッチング結果>

【図 2 6】

サジェスト

検索結果

No	スコア	検索結果
1	7.2	検索結果が低い
2	2.4	検索結果が低い
3	2.4	検索結果が低い
4	2.4	検索結果が低い
5	2.4	検索結果が低い

先頭へ 前へ 1/1 後へ 最後へ

<マッチプロファイルCによるマッチング結果>

【図 27】

No.	マッチモード定義		マッチングモード			
	単語要素マッチ	属性要素マッチ	単語要素+属性要素	係りうけ+属性要素	係りうけ+属性要素+単語要素	係りうけ+属性要素+単語要素+属性要素
1	(1) 単語要素マッチ	属性要素タイプ	属性要素タイプ	属性要素タイプ	属性要素タイプ	属性要素タイプ
2	(2) 単語要素+属性要素	属性要素タイプ	属性要素タイプ	属性要素タイプ	属性要素タイプ	属性要素タイプ
3	(3) 係りうけ+属性要素	属性要素タイプ	属性要素タイプ	属性要素タイプ	属性要素タイプ	属性要素タイプ
4	(4) 係りうけ+属性要素+単語要素	属性要素タイプ	属性要素タイプ	属性要素タイプ	属性要素タイプ	属性要素タイプ

(a)

【図 28】

マッチプロファイル: 単語要素+属性要素マッチモード
「ETCカード」、「お店」、「使う」のいずれかが含まれているとマッチする。

サジェスト		
ETCカードがお店で使えません。		
No.	スコア	検索結果
1	20.4	ETCカードが使えない
2	15.9	ETCカードを使いたい
3	15.9	ETCカードは使いやすい
4	2.3	ETCカードを失くした。
5	1.8	クレジットカードが使えない
6	0.9	クレジットカードを使いたい

【図 29】

マッチプロファイル: 単語要素+属性要素マッチモード
「ETCカード」、「お店」、「使う(否定)」のいずれかが含まれているとマッチする。

サジェスト		
ETCカードがお店で使えません。		
No.	スコア	検索結果
1	20.4	ETCカードが使えない
2	2.3	ETCカードを使いたい
3	2.3	ETCカードは使いやすい
4	2.3	ETCカードを失くした。
5	1.8	クレジットカードが使えない

【図 30】

マッチプロファイル: 係りうけ+属性要素マッチモード
「ETCカード+使う」、「お店+使う」のいずれかが含まれているとマッチする。

サジェスト		
ETCカードがお店で使えません。		
No.	スコア	検索結果
1	18.9	ETCカードが使えない
2	15.1	ETCカードを使いたい
3	15.1	ETCカードは使いやすい

【図 31】

マッチプロファイル: 係りうけ+属性要素マッチモード
「ETCカード+使う(否定)」、「お店+使う(否定)」のいずれかが含まれているとマッチする。

サジェスト		
ETCカードがお店で使えません。		
No.	スコア	検索結果
1	18.9	ETCカードが使えない

フロントページの続き

- (72)発明者 岡田 達彦
東京都江東区豊洲三丁目3番3号 株式会社エヌ・ティ・ティ・データ内
- (72)発明者 亘 健典
東京都江東区豊洲三丁目3番3号 株式会社エヌ・ティ・ティ・データ内
- (72)発明者 溝渕 敬司
東京都江東区豊洲三丁目3番3号 株式会社エヌ・ティ・ティ・データ内
- (72)発明者 高井 貞治
東京都渋谷区広尾五丁目4番12号 合同会社シンタックス内
- (72)発明者 石岡 隆光
東京都渋谷区広尾五丁目4番12号 合同会社シンタックス内
- (72)発明者 井上 世紀
東京都豊島区高田三丁目28番2号 株式会社ウェブプロデュース内

審査官 鹿野 博嗣

- (56)参考文献 特開2006-338457(JP,A)
特開2003-167898(JP,A)
特開平10-105555(JP,A)
特開2001-084248(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06F 17/30

G06F 3/0482