

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6065844号
(P6065844)

(45) 発行日 平成29年1月25日 (2017. 1. 25)

(24) 登録日 平成29年1月6日 (2017. 1. 6)

(51) Int. Cl. F 1
GO 6 F 17/30 (2006. 01)
GO 6 F 17/30 2 2 0 Z
GO 6 F 17/30 4 1 4 Z

請求項の数 9 (全 20 頁)

(21) 出願番号	特願2013-553268 (P2013-553268)	(73) 特許権者	000004237
(86) (22) 出願日	平成25年1月10日 (2013. 1. 10)		日本電気株式会社
(86) 国際出願番号	PCT/JP2013/000067		東京都港区芝五丁目7番1号
(87) 国際公開番号	W02013/105505	(74) 代理人	100110928
(87) 国際公開日	平成25年7月18日 (2013. 7. 18)		弁理士 速水 進治
審査請求日	平成27年12月3日 (2015. 12. 3)	(72) 発明者	西村 祥治
(31) 優先権主張番号	特願2012-5562 (P2012-5562)		東京都港区芝五丁目7番1号 日本電気株
(32) 優先日	平成24年1月13日 (2012. 1. 13)		式会社内
(33) 優先権主張国	日本国 (JP)		
		審査官	川▲崎▼ 博章

最終頁に続く

(54) 【発明の名称】 インデックス走査装置及びインデックス走査方法

(57) 【特許請求の範囲】

【請求項 1】

複数の部分空間に分割された多次元空間にマッピングされたデータベースにおける、該複数の部分空間の各々を特定するための各インデックスデータを、各部分空間に属するデータが格納されるデータページを特定するための各ページ特定データと関連付けて格納するインデックス格納部と、

クエリの検索範囲に対応する前記多次元空間内の検索空間の情報、又は、該検索空間内の部分検索空間の情報を取得する検索空間取得部と、

前記検索空間又は前記部分検索空間に含まれるデータ点のうち、前記多次元空間の各データ点を一次元化することにより得られる値が最小及び最大の少なくとも一方となるデータ点を検索点として設定する検索点設定部と、

前記インデックス格納部において、前記検索点設定部により設定された前記検索点を含む部分空間のインデックスデータを特定し、該特定されたインデックスデータと関連付けられたページ特定データを前記検索空間の少なくとも一部と重なる部分空間のページ特定データとして抽出するインデックス検索部と、

前記インデックス検索部により特定されたインデックスデータから得られる部分空間の特定情報に基づいて、前記検索空間又は前記部分検索空間を通る、前記多次元空間を前記複数の部分空間に分割する境界線を推定し、該推定された境界線により前記検索空間又は前記部分検索空間を分割することにより得られる分割検索空間のうち、前記検索点を含む部分空間以外の分割検索空間を前記部分検索空間として決定する検索領域決定部と、

10

20

を備えるインデックス走査装置。

【請求項 2】

前記各インデックスデータは、各部分空間の位置及び範囲を特定し得るデータ形式を有する請求項 1 に記載のインデックス走査装置。

【請求項 3】

前記多次元空間の空間分割には、KD (K-dimensional) - Tree が利用され、

前記多次元空間の一次元化には、空間充填曲線が用いられ、

前記多次元空間の各データ点を一次元化することにより得られる値は、2 進数表記され

、
前記インデックス格納部は、各部分空間に含まれるデータ点の最長共通プレフィックスを各インデックスデータとし、

前記検索領域決定部は、各軸の各データ点のうち、最下位桁から上位桁方向に向けて初めて 0 が現れる桁位置までの全桁に 1 が設定されている値で示されるデータ点と、該値に 1 を加算して得られる値で示されるデータ点との間を前記境界線として推定する、

請求項 1 又は 2 に記載のインデックス走査装置。

【請求項 4】

複数の部分空間に分割された多次元空間にマッピングされたデータベースにおける、該複数の部分空間の各々を特定するための各インデックスデータを、各部分空間に属するデータが格納されるデータページを特定するための各ページ特定データと関連付けて格納するインデックス格納部を走査するインデックス走査方法において、

クエリの検索範囲に対応する前記多次元空間内の検索空間の情報、又は、該検索空間内の部分検索空間の情報を取得し、

前記検索空間又は前記部分検索空間に含まれるデータ点のうち、前記多次元空間の各データ点を一次元化することにより得られる値が最小及び最大の少なくとも一方となるデータ点を検索点として設定し、

前記インデックス格納部において、前記設定された検索点を含む部分空間のインデックスデータを特定し、

前記特定されたインデックスデータと関連付けられたページ特定データを前記検索空間の少なくとも一部と重なる部分空間のページ特定データとして抽出し、

前記特定されたインデックスデータから得られる部分空間の特定情報に基づいて、前記検索空間又は前記部分検索空間を通る、前記多次元空間を前記複数の部分空間に分割する境界線を推定し、

前記推定された境界線により前記検索空間又は前記部分検索空間を分割することにより得られる分割検索空間のうち、前記検索点を含む部分空間以外の分割検索空間を前記部分検索空間として決定する、

ことを含むインデックス走査方法。

【請求項 5】

前記各インデックスデータは、各部分空間の位置及び範囲を特定し得るデータ形式を有する請求項 4 に記載のインデックス走査方法。

【請求項 6】

前記多次元空間の空間分割には、KD (K-dimensional) - Tree が利用され、

前記多次元空間の一次元化には、空間充填曲線が用いられ、

前記多次元空間の各データ点を一次元化することにより得られる値は、2 進数表記され

、
前記インデックス格納部は、各部分空間に含まれるデータ点の最長共通プレフィックスを各インデックスデータとし、

前記境界線の推定は、各軸の各データ点のうち、最下位桁から上位桁方向に向けて初めて 0 が現れる桁位置までの全桁に 1 が設定されている値で示されるデータ点と、該値に 1

10

20

30

40

50

を加算して得られる値で示されるデータ点との間を前記境界線として推定する、
請求項 4 又は 5 に記載のインデックス走査方法。

【請求項 7】

コンピュータに、

複数の部分空間に分割された多次元空間にマッピングされたデータベースにおける、該複数の部分空間の各々を特定するための各インデックスデータを、各部分空間に属するデータが格納されるデータページを特定するための各ページ特定データと関連付けて格納するインデックス格納部と、

クエリの検索範囲に対応する前記多次元空間内の検索空間の情報、又は、該検索空間内の部分検索空間の情報を取得する検索空間取得部と、

前記検索空間又は前記部分検索空間に含まれるデータ点のうち、前記多次元空間の各データ点を一次元化することにより得られる値が最小及び最大の少なくとも一方となるデータ点を検索点として設定する検索点設定部と、

前記インデックス格納部において、前記検索点設定部により設定された前記検索点を含む部分空間のインデックスデータを特定し、該特定されたインデックスデータと関連付けられたページ特定データを前記検索空間の少なくとも一部と重なる部分空間のページ特定データとして抽出するインデックス検索部と、

前記インデックス検索部により特定されたインデックスデータから得られる部分空間の特定情報に基づいて、前記検索空間又は前記部分検索空間を通る、前記多次元空間を前記複数の部分空間に分割する境界線を推定し、該推定された境界線により前記検索空間又は前記部分検索空間を分割することにより得られる分割検索空間のうち、前記検索点を含む部分空間以外の分割検索空間を前記部分検索空間として決定する検索領域決定部と、

を実現させるプログラム。

【請求項 8】

前記各インデックスデータは、各部分空間の位置及び範囲を特定し得るデータ形式を有する請求項 7 に記載のプログラム。

【請求項 9】

前記多次元空間の空間分割には、KD (K-dimensional) - Tree が利用され、

前記多次元空間の一次元化には、空間充填曲線が用いられ、

前記多次元空間の各データ点を一次元化することにより得られる値は、2 進数表記され、

前記インデックス格納部は、各部分空間に含まれるデータ点の最長共通プレフィックスを各インデックスデータとし、

前記検索領域決定部は、各軸の各データ点のうち、最下位桁から上位桁方向に向けて初めて 0 が現れる桁位置までの全桁に 1 が設定されている値で示されるデータ点と、該値に 1 を加算して得られる値で示されるデータ点との間を前記境界線として推定する、

請求項 7 又は 8 に記載のプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、複数の部分空間に分割された多次元空間にマッピングされたデータベースに対する多次元範囲検索におけるインデックス走査技術に関する。

【背景技術】

【0002】

データ数の増加に伴い、拡張性のある分散型データベースが提案されている。これら分散型データベースの多くは、一次元のみ範囲検索のような単純な問い合わせ処理のみをサポートすることで拡張性を実現している。一方で、多次元値を一次元値にマップすることで、複数次元の範囲検索を可能としている分散型データベースも提案されている（下記非特許文献 1）。

10

20

30

40

50

【0003】

下記非特許文献1で提案される多次元範囲検索手法では、まず、データベースがマッピングされている多次元空間内の検索対象範囲（検索空間）を一次元化することで得られるデータ点の中から、最大値を示すデータ点（最大データ点と表記する）及び最小値を示すデータ点（最小データ点と表記する）が取得される。続いて、当該手法は、その多次元空間内の各部分空間を特定するための各インデックスデータが当該データ点の示す値の順にソートされてなるインデックス列を、最小データ点から最大データ点までの範囲で走査する。このとき、その走査範囲内にある各インデックスで特定される各部分空間について、検索空間と少なくとも一部が重なっているか否かがそれぞれチェックされ、重なっている部分空間に対応付けられたデータページが検索対象に決定される。

10

【先行技術文献】

【非特許文献】

【0004】

【非特許文献1】Shoji Nishimura, SudiptoDas, Divyakant Agrawal, Amr El Abbadi, "MD-HBase: A Scalable Multi-Dimensional Data Infrastructure for Location Aware Services", IEEE computer society, 2011.6.6.

【発明の概要】

【0005】

しかしながら、上述の手法では、インデックス走査が非効率になる場合がある。つまり、検索空間が多次元空間全体に比べて小さいにもかかわらず、インデックスの走査範囲（最小データ点から最大データ点までの間）に多くの部分空間が含まれる場合があり得る。最悪ケースとして、インデックス走査範囲に、各インデックスで特定される全部分空間が含まれる場合もあり得る。このような場合、全部分空間について検索空間と重なっているか否かをチェックすることになり、インデックス走査で検索空間と重なる部分空間を抽出するのに多くの時間が費やされてしまう。

20

【0006】

このような問題は、多次元空間を一次元化するときに生じるひずみに起因する。例えば、検索範囲として、複数の属性の中の一部の属性値が指定されず、或る属性の全値域が指定される場合、指定した属性の範囲が十分狭いにもかかわらず、全インデックスを走査することになる可能性がある。このような問題は、検索対象の次元数が多い場合や、検索時に指定する各次元の範囲の幅に差がある場合等に顕著となる。

30

【0007】

本発明は、上述のような課題に鑑みてなされたものであり、多次元範囲検索におけるインデックス走査を高速化する技術を提供する。

【0008】

本発明の各態様では、上述した課題を解決するために、それぞれ以下の構成を採用する。

【0009】

第1の態様に係るインデックス走査装置は、複数の部分空間に分割された多次元空間にマッピングされたデータベースにおける、複数の部分空間の各々を特定するための各インデックスデータを、各部分空間に属するデータが格納されるデータページを特定するための各ページ特定データと関連付けて格納するインデックス格納部と、クエリの検索範囲に対応する上記多次元空間内の検索空間の情報、又は、その検索空間内の部分検索空間の情報を取得する検索空間取得部と、検索空間又は部分検索空間に含まれるデータ点のうち、多次元空間の各データ点を一次元化することにより得られる値が最小及び最大の少なくとも一方となるデータ点を検索点として設定する検索点設定部と、インデックス格納部において、検索点設定部により設定された検索点を含む部分空間のインデックスデータを特定し、この特定されたインデックスデータと関連付けられたページ特定データを検索空間の少なくとも一部と重なる部分空間のページ特定データとして抽出するインデックス検索部と、インデックス検索部により特定されたインデックスデータから得られる部分空間の特

40

50

定情報に基づいて、検索空間又は部分検索空間を通る、上記多次元空間を上記複数の部分空間に分割する境界線を推定し、この推定された境界線により検索空間又は部分検索空間を分割することにより得られる分割検索空間のうち、当該検索点を含む部分空間以外の分割検索空間を部分検索空間として決定する検索領域決定部と、を備える。

【0010】

第2の態様に係るインデックス走査方法は、複数の部分空間に分割された多次元空間にマッピングされたデータベースにおける、複数の部分空間の各々を特定するための各インデックスデータを、各部分空間に属するデータが格納されるデータページを特定するための各ページ特定データと関連付けて格納するインデックス格納部を走査する。第2の態様に係るインデックス走査方法は、クエリの検索範囲に対応する上記多次元空間内の検索空間の情報、又は、その検索空間内の部分検索空間の情報を取得し、検索空間又は部分検索空間に含まれるデータ点のうち、多次元空間の各データ点を一次元化することにより得られる値が最小及び最大の少なくとも一方となるデータ点を検索点として設定し、インデックス格納部において、検索点設定部により設定された検索点を含む部分空間のインデックスデータを特定し、この特定されたインデックスデータと関連付けられたページ特定データを検索空間の少なくとも一部と重なる部分空間のページ特定データとして抽出し、その特定されたインデックスデータから得られる部分空間の特定情報に基づいて、検索空間又は部分検索空間を通る、上記多次元空間を上記複数の部分空間に分割する境界線を推定し、この推定された境界線により検索空間又は部分検索空間を分割することにより得られる分割検索空間のうち、当該検索点を含む部分空間以外の分割検索空間を部分検索空間として決定することを含む。

10

20

【0011】

なお、本発明の他の態様としては、上記第1態様の構成をコンピュータに実現させるプログラムであってもよいし、このようなプログラムを記録したコンピュータが読み取り可能な記録媒体であってもよい。この記録媒体は、非一時的な有形の媒体を含む。

【0012】

本発明の各態様によれば、多次元範囲検索におけるインデックス走査を高速化する技術を提供することができる。

【図面の簡単な説明】

【0013】

上述した目的、およびその他の目的、特徴および利点は、以下に述べる好適な実施の形態、およびそれに付随する以下の図面によってさらに明らかになる。

30

【0014】

【図1】図1は、第1実施形態におけるインデックス走査装置のハードウェア構成例を概念的に示す図である。

【図2】図2は、第1実施形態におけるインデックス走査装置の処理構成例を概念的に示す図である。

【図3】図3は、第1実施形態におけるインデックス走査装置の動作例を示すフローチャートである。

【図4】図4は、実施例におけるデータベースにマッピングされた2次元空間及び2次元検索範囲を概念的に示す図である。

40

【図5】図5は、実施例におけるインデックス格納部に格納されるエントリの例を示す図である。

【図6】図6は、最初に特定された部分空間（最小データ点を含む）とこの部分空間から推定される境界線を概念的に示す図である。

【図7】図7は、最初に特定された部分空間（最大データ点を含む）とこの部分空間から推定される境界線を概念的に示す図である。

【発明を実施するための形態】

【0015】

以下、本発明の実施の形態について説明する。なお、以下に挙げる実施形態は例示であ

50

り、本発明は以下の実施形態の構成に限定されない。

【0016】

本実施形態に係るインデックス走査装置は、複数の部分空間に分割された多次元空間にマッピングされたデータベースにおける、複数の部分空間の各々を特定するための各インデックスデータを、各部分空間に属するデータが格納されるデータページを特定するための各ページ特定データと関連付けて格納するインデックス格納部と、クエリの検索範囲に対応する上記多次元空間内の検索空間の情報、又は、その検索空間内の部分検索空間の情報を取得する検索空間取得部と、検索空間又は部分検索空間に含まれるデータ点のうち、多次元空間の各データ点を一次元化することにより得られる値が最小及び最大の少なくとも一方となるデータ点を検索点として設定する検索点設定部と、インデックス格納部において、検索点設定部により設定された検索点を含む部分空間のインデックスデータを特定し、この特定されたインデックスデータと関連付けられたページ特定データを検索空間の少なくとも一部と重なる部分空間のページ特定データとして抽出するインデックス検索部と、インデックス検索部により特定されたインデックスデータから得られる部分空間の特定情報に基づいて、検索空間又は部分検索空間を通る、上記多次元空間を上記複数の部分空間に分割する境界線を推定し、この推定された境界線により検索空間又は部分検索空間を分割することにより得られる分割検索空間のうち、当該検索点を含む部分空間以外の分割検索空間を部分検索空間として決定する検索領域決定部と、を備える。

10

【0017】

本実施形態におけるインデックス走査方法は、複数の部分空間に分割された多次元空間にマッピングされたデータベースにおける、複数の部分空間の各々を特定するための各インデックスデータを、各部分空間に属するデータが格納されるデータページを特定するための各ページ特定データと関連付けて格納するインデックス格納部を走査する。当該インデックス走査方法は、クエリの検索範囲に対応する上記多次元空間内の検索空間の情報、又は、その検索空間内の部分検索空間の情報を取得し、検索空間又は部分検索空間に含まれるデータ点のうち、多次元空間の各データ点を一次元化することにより得られる値が最小及び最大の少なくとも一方となるデータ点を検索点として設定し、インデックス格納部において、検索点設定部により設定された検索点を含む部分空間のインデックスデータを特定し、この特定されたインデックスデータと関連付けられたページ特定データを検索空間の少なくとも一部と重なる部分空間のページ特定データとして抽出し、その特定されたインデックスデータから得られる部分空間の特定情報に基づいて、検索空間又は部分検索空間を通る、上記多次元空間を上記複数の部分空間に分割する境界線を推定し、この推定された境界線により検索空間又は部分検索空間を分割することにより得られる分割検索空間のうち、当該検索点を含む部分空間以外の分割検索空間を部分検索空間として決定することを含む。

20

30

【0018】

本実施形態では、検索空間又は部分検索空間に含まれるデータ点の中から、一次元化された値が最大となるデータ点（以降、最大データ点と表記する）及び一次元化された値が最小となるデータ点（以降、最小データ点と表記する）の少なくとも一方が検索点として設定され、この検索点を含む部分空間のインデックスデータが特定され、この特定されたインデックスデータと関連付けられたページ特定データが抽出される。抽出されたページ特定データは、当該検索空間の少なくとも一部と重なる部分空間に属するデータが格納されるデータページを特定するものである。

40

【0019】

更に、本実施形態では、上述のように特定されたインデックスデータから得られる部分空間の特定情報に基づいて、検索空間又は部分検索空間を通る境界線が推定され、この推定された境界線により検索空間又は部分検索空間を分割することにより得られる分割検索空間のうち、当該検索点を含む部分空間以外の分割検索空間が部分検索空間として決定される。これにより、ここで決定された部分検索空間を対象にして、上述のように、検索点の設定、インデックスデータの特定及びページ特定データの抽出が実行される。このよう

50

な処理は、例えば、検索空間の少なくとも一部と重なる全ての部分空間のページ特定データが抽出されるまで、繰り返される。

【0020】

このように、本実施形態では、検索空間と重なる1つの部分空間が見つかり、その部分空間の存在から、多次元空間全体がどのように空間分割されているか（境界線）が出来る範囲で推定され、推定された境界線により、その見つけられた1つの部分空間を除いた検索空間がより小さな検索空間の集合に分割され、インデックスデータの走査範囲がその小さな検索空間に置き換えられる。

【0021】

結果、本実施形態によれば、クエリの検索範囲に対応する検索空間と少なくとも1部が重なる部分空間のみを対象にしてインデックス走査が行われるため、インデックス走査を高速化することができる。

【0022】

〔第1実施形態〕

以下、上述の実施形態について更に詳細を説明する。

【0023】

〔装置構成〕

図1は、第1実施形態におけるインデックス走査装置10のハードウェア構成例を概念的に示す図である。第1実施形態におけるインデックス走査装置10は、いわゆるコンピュータであり、例えば、バス5で相互に接続される、CPU（Central Processing Unit）2、メモリ3、入出力インタフェース（I/F）4等を有する。メモリ3は、RAM（Random Access Memory）、ROM（Read Only Memory）、ハードディスク、可搬型記憶媒体等である。

【0024】

入出力I/F4は、ネットワーク（図示せず）を介して他のコンピュータと通信を行う通信装置等と接続される。入出力I/F4は、キーボード、マウス等のようなユーザ操作の入力を受け付ける入力装置、ディスプレイ装置やプリンタ等のようなユーザに情報を提供する出力装置と接続されてもよい。インデックス走査装置10のハードウェア構成は制限されない。

【0025】

図2は、第1実施形態におけるインデックス走査装置10の処理構成例を概念的に示す図である。第1実施形態におけるインデックス走査装置10は、検索空間取得部11、検索点設定部13、検索領域決定部15、インデックス検索部17、インデックス格納部21、ページ特定データ保持部23、検索空間保持部25等を有する。これら各処理部は、例えば、CPU2によりメモリ3に格納されるプログラムが実行されることにより実現される。また、当該プログラムは、例えば、CD（Compact Disc）、メモリカード等のような可搬型記録媒体やネットワーク上の他のコンピュータから入出力I/F4を介してインストールされ、メモリ3に格納されてもよい。

【0026】

インデックス走査装置10は、インデックス格納部21とデータ格納部（図示せず）とからなるデータベースシステムに適用される。データ格納部は、複数のデータページにより複数のデータを格納する。各データページには、所定閾値以下の件数のデータがそれぞれ格納される。

【0027】

インデックス格納部21は、データ格納部に格納される複数データがマッピングされた多次元空間を構成する各部分空間を特定するための各インデックスデータ、及び、各部分空間に対応する各データページを特定するための各ページ特定データを含む複数エントリを格納する。インデックス格納部21に格納される各エントリにより、1つの部分空間を示す1つのインデックスデータと、その部分空間に対応するデータページを特定するための1つのページ特定データとがそれぞれ特定される。なお、1つのインデックスデータは

、複数の部分空間を示し、複数のページ特定データと共に1つのエントリに含まれてもよい。

【0028】

各インデックスデータは、各部分空間の位置及び範囲を特定し得るデータ形式を持つ。例えば、一次元化された多次元空間内の各データ点が2進数表記される場合、各インデックスデータは、対応する各部分空間に含まれるデータ点の最長共通プレフィックスに設定される。多次元空間を複数の部分空間に分割する手法には、例えば、KD (K-Dimensional) - Tree 等のような空間分割法が利用される。多次元空間の一次元化には、例えば、Zカーブ、ヒルベルトカーブ等のような空間充填曲線が利用される。

【0029】

検索空間取得部11は、多次元範囲検索のクエリ50を取得し、このクエリ50を解析することによりクエリ50の検索範囲に対応する多次元空間内の検索空間の情報を取得する。検索空間の情報には、少なくとも、当該検索空間に含まれるデータ点の中の最小データ点及び最大データ点が含まれる。上記検索空間情報の取得の後、検索空間取得部11は、検索空間保持部25に保持される部分検索空間の情報を取得する。部分検索空間とは、検索空間内の空間であって、後述のインデックス検索部17により、その空間に対応するページ特定データが未だ抽出されていない未処理空間を意味する。部分検索空間は、少なくとも1つの部分空間と重複する。

【0030】

検索点設定部13は、検索空間に含まれるデータ点の中の最小データ点及び最大データ点の少なくとも一方を検索点として設定する。この場合、検索点には、最小データのみ、最大データ点のみ、又は、最小データ点及び最大データ点の両方が設定される。また、検索点設定部13は、検索空間取得部11により部分検索空間が取得された場合には、部分検索空間に含まれるデータ点の中の最小データ点及び最大データ点の少なくとも一方を検索点として設定する。

【0031】

インデックス検索部17は、インデックス格納部21を検索することにより、検索点設定部13により設定された検索点を含む部分空間のインデックスデータを含むエントリを特定し、この特定されたエントリに含まれるページ特定データを抽出する。この抽出されたページ特定データは、検索空間の少なくとも一部と重なる部分空間に対応する。抽出されたページ特定データは、ページ特定データ保持部23に保持される。

【0032】

検索領域決定部15は、インデックス検索部17により特定されたエントリに含まれるインデックスデータから得られる部分空間の特定情報（例えば、位置及び範囲）に基づいて、当該検索空間又は当該部分検索空間から、対応するページ特定データが既に抽出された部分空間を除外して残る空間を部分検索空間に決定する。以降、この残空間を未踏検索空間と表記する場合もある。このとき、検索領域決定部15は、当該検索空間又は当該部分検索空間を通る、多次元空間を複数の部分空間に分割する境界線を推定し、上記未踏検索空間をその境界線で分割することにより得られる各空間を当該部分検索空間に決定する。このように決定された部分検索空間の情報は検索空間保持部25に保持される。

【0033】

ページ特定データ保持部23は、インデックス検索部17により抽出されたページ特定データを保持する。つまり、ページ特定データ保持部23は、検索空間の少なくとも一部と重なる部分空間に対応するページ特定データを保持する。ページ特定データは、例えば、データページのアドレスである。

【0034】

検索空間保持部25は、検索領域決定部15により決定された部分検索空間の情報を保持する。部分検索空間の情報には、少なくとも、当該部分検索空間に含まれるデータ点の中の最小データ点又は最大データ点が含まれる。

【0035】

〔動作例〕

以下、第1実施形態におけるインデックス走査方法について図3を用いて説明する。図3は、第1実施形態におけるインデックス走査装置10の動作例を示すフローチャートである。

【0036】

インデックス走査装置10は、多次元範囲検索のクエリ50を取得する(S31)。インデックス走査装置10は、このクエリ50で指定された検索範囲に対応する多次元空間内の検索空間の情報を取得する(S32)。

【0037】

続いて、インデックス走査装置10は、その検索空間内において検索点を設定する(S33)。例えば、インデックス走査装置10は、検索空間に含まれるデータ点の中の最小データ点及び最大データ点の少なくとも一方を検索点として設定する。

【0038】

インデックス走査装置10は、インデックス格納部21に格納されるインデックスデータの中から、当該検索点を含む部分空間に対応するインデックスデータを特定する(S34)。言い換えれば、インデックス走査装置10は、インデックス格納部21に格納されるインデックスデータを走査することにより、多次元空間を構成する複数の部分空間の中から、この検索点を含む部分空間を特定する。

【0039】

インデックス走査装置10は、特定されたインデックスデータと関連付けられたページ特定データを抽出する(S35)。インデックス走査装置10は、抽出されたページ特定データをページ特定データ保持部23に保持させる。

【0040】

続いて、インデックス走査装置10は、検索空間内における、これまで(S34)で特定された全ての部分空間以外の残空間(未踏検索空間)を部分検索空間に決定する(S36)。このとき、インデックス走査装置10は、(S34)で特定された部分空間の特定情報に基づいて、当該検索空間を通る境界線を推定し、上記未踏検索空間をその境界線で分割することにより得られる各空間を当該部分検索空間に決定する。これにより、検索空間を通る境界線が1つ以上見つければ、1つ以上の部分検索空間が決定されることになる。インデックス走査装置10は、このように決定された部分検索空間の情報を検索空間保持部25に保持させる。

【0041】

インデックス走査装置10は、部分検索空間が有るか否かを判定する(S37)。具体的には、インデックス走査装置10は、検索空間保持部25に部分検索空間の情報が保持されているか否かでこの判定を行う。インデックス走査装置10は、部分検索空間が無いと判定した場合(S37; NO)、当該検索空間の少なくとも一部と重複する部分空間に対応付けられた全てのページ特定データが抽出されたと判断し、処理を終了する。

【0042】

一方、インデックス走査装置10は、部分検索空間が有ると判定した場合(S37; YES)、検索空間保持部25から1つの部分検索空間の情報を取得する(S38)。インデックス走査装置10により取得された部分検索空間の情報は、検索空間保持部25から削除される。

【0043】

インデックス走査装置10は、その部分検索空間内において検索点を設定する(S39)。この検索点は、例えば、その部分検索空間に含まれるデータ点の中の最小データ点及び最大データ点の少なくとも一方である。

【0044】

インデックス走査装置10は、この新たに設定された検索点に関し、(S34)、(S35)及び(S36)を上述のように実行する。このとき、(S36)で部分検索空間が新たに決定された場合には、その新たな部分検索空間の情報が検索空間保持部25に追加

される。(S 3 6)において新たな部分検索空間がないと判定された場合には、検索空間保持部 2 5 に保持される部分検索空間の情報が取得される(S 3 7)。インデックス走査装置 1 0 は、検索空間保持部 2 5 に保持される部分検索空間の情報がなくなるまで(S 3 7; NO)、(S 3 8)、(S 3 9)、(S 3 4)、(S 3 5)、(S 3 6)を実行する。

【0 0 4 5】

なお、上述の動作例では、部分検索空間が 1 つずつ処理されていたが、複数の部分検索空間が検索空間保持部 2 5 に保持されている場合には、部分検索空間毎に、(S 3 4) から(S 3 9)が並列かつ独立して実行されるようにしてもよい。これは、例えば、検索空間取得部 1 1、検索点設定部 1 3、検索領域決定部 1 5 及びインデックス検索部 1 7 のセットが部分検索空間毎に並列して動作するように構成されることで実現される。

10

【0 0 4 6】

〔第 1 実施形態の作用及び効果〕

上述のように、第 1 実施形態では、インデックス格納部 2 1 において、検索空間と重なる部分空間に対応するインデックスデータを探すにあたり、多次元空間の各データ点を一次元化することにより得られるバイナリ値が最小及び最大の少なくとも一方となるデータ点を含む部分空間に対応するインデックスデータが特定される。そして、この特定されたインデックスデータから得られる対応する部分空間の特定情報(位置及び範囲)から、全体の多次元空間がどのように空間分割されているのかが推定される。この推定により、検索空間を通る境界線情報が取得され、検索空間におけるその特定された部分空間以外の残

20

【0 0 4 7】

従って、第 1 実施形態によれば、クエリ 5 0 の検索範囲に対応する検索空間の少なくとも一部と重なる部分空間のみを対象としてインデックス走査が行われるため、インデックス走査を高速化することができる。

【実施例】

【0 0 4 8】

以下、上述の各実施形態の更なる具体例を説明する。以下の実施例では、データベースがマッピングされた検索対象空間の次元が 2 次元であり、この 2 次元空間が Z カーブ(空間充填曲線)により一次元化され、各データ点が 2 進数表記され、K D - T r e e により当該 2 次元空間が空間分割された場合の例が示される。

30

【0 0 4 9】

図 4 は、実施例におけるデータベースにマッピングされた 2 次元空間及び 2 次元検索範囲を概念的に示す図である。なお、以降、図 4 における横軸を X 軸と表記し、縦軸を Y 軸と表記する。図 4 の太線は、空間分割の境界線を示し、図 4 の太点線は、検索空間を示す。2 次元空間の各データ点は、図 4 に示されるバイナリ値(上段が上位 3 桁、下段が下位 3 桁)で特定される。各バイナリ値は、各軸のデータ点の値(各々 3 桁のバイナリ値)を上位桁から下位桁へ方向に 1 桁ずつ交互に並べることにより形成される。具体的には、X 軸の値「0 0 1」と Y 軸の値「0 1 0」とで特定されるデータ点は、X 軸の最上位桁「0」、Y 軸の最上位桁「0」、X 軸の 2 番目の桁「0」、Y 軸の 2 番目の桁「1」、X 軸の最下位桁「1」、Y 軸の最下位桁「0」が並べられることにより形成される(「0 0 0 1 1 0」)。

40

【0 0 5 0】

図 5 は、実施例におけるインデックス格納部 2 1 に格納されるエントリの例を示す図である。図 5 の例によれば、各エントリには、図 4 に示される部分空間のいずれか 1 つを特定するインデックスデータと、そのインデックスデータに対応するページアドレスとがそれぞれ含まれる。各インデックスデータは、対応する部分空間に含まれるバイナリ値の最長共通プレフィックスと、他の各桁を示す「*」とで表記されている。「*」は、その桁の値を制限しないことを意味する。

50

【0051】

例えば、データ点「000000」、「000001」、「000010」及び「000011」を含む部分空間（図4の左下）のインデックスデータは、最長共通プレフィックス「0000」と、他の2桁を示す「**」とで形成される。また、データ点「001000」、「001001」、「001010」、「001011」、「001100」、「001101」、「001110」及び「001111」を含む部分空間（図4における上記部分空間の右隣）のインデックスデータは、最長共通プレフィックス「001」と他の3桁を示す「***」とで形成される。実施例におけるインデックス格納部21の各エントリは、インデックスデータを昇順にソートすることにより、並べられている。

【0052】

以下、この実施例におけるインデックス走査装置10の動作例について図3を用いて説明する。なお、第1実施形態と同じ内容については適宜省略する。

【0053】

インデックス走査装置10は、(S31)で取得されたクエリ50で指定された検索範囲に対応する2次元空間内の検索空間の情報を取得する(S32)。実施例では、当該検索空間の情報には、最大データ点「100111」、最小データ点「000011」が含まれる。

【0054】

インデックス走査装置10は、当該検索空間内において検索点を設定する(S33)。ここでは、インデックス走査装置10は、検索空間に含まれるデータ点の中の最小データ点「000011」を検索点に設定したものと仮定する。

【0055】

インデックス走査装置10は、図5に示されるインデックスデータの中から、当該検索点「000011」を含む部分空間に対応するインデックスデータを特定する(S34)。本実施例では、インデックス格納部21において、インデックスデータと検索点とがプレフィックスマッチするエントリが二分探索されることにより、特定すべき当該インデックスデータを効率的に見つけることができる。ここでは、検索点「000011」とプレフィックスマッチするインデックスデータとして、インデックスデータ「0000**」が特定される。

【0056】

図6は、最初に特定された部分空間とこの部分空間から推定される境界線を概念的に示す図である。図6の網掛けで示される領域がインデックスデータ「0000**」で特定される部分空間である。インデックス走査装置10は、特定されたインデックスデータ「0000**」と関連付けられたページ特定データ「0x0001」を抽出する(S35)。インデックス走査装置10は、抽出されたページ特定データ「0x0001」をページ特定データ保持部23に保持させる。

【0057】

続いて、インデックス走査装置10は、検索空間内における、上述のように特定された部分空間（インデックスデータ「0000**」）以外の残空間（未踏検索空間）を部分検索空間に決定する(S36)。

【0058】

ここで、図6の網掛けで示される部分空間の存在が判明すると、全体の2次元空間は、図6の太線で示された境界線で空間分割されていると推定することができる。これは、本実施例で用いられた空間ベースのKD-Treeでの空間分割手法の特性に基づく。一般的に、空間ベースの空間分割手法では、多次元空間が規則的かつ再帰的に分割されるため、或る部分空間の位置及び範囲に基づいて、その部分空間が切り出されるためにどのように空間が分割されてきたかを逆算することができる。一方で、データ点ベースの空間分割手法は、空間分割境界がデータ点に基づいて決定されるため、一般的に、位置と範囲から空間分割形態を逆算によって求めることができない。

【0059】

上述の各実施形態では、このような事実に基づいて、既に特定された部分空間以外の残空間（未踏検索空間）を分割して得られる各空間が部分検索空間として決定される。具体的には、既に特定された部分空間の特定情報（位置及び範囲）から導出される空間分割の情報（境界線情報）により、検索空間内の未踏検索空間が分割される。検索空間内を通る境界線は、各軸についてそれぞれ推定される。具体的には、当該境界線は、各軸の各データ点のうち、最下位桁から上位桁方向に向けて初めて0が現れる桁位置までの全桁に1が設定されている値で示されるデータ点（以降、負側境界値と表記する）と、この値に1を加算して得られる値で示されるデータ点（以降、正側境界値と表記する）との間に位置すると推定される。

【0060】

10

以下、図6の例に基づいて、具体的な境界線の推定方法について説明する。

まず、未踏検索空間における各次元（各軸）の下限值及び上限値が取得される。X軸については上限値が「101」であり、下限値が「001」であり、Y軸については上限値が「011」であり、下限値が「001」である。

【0061】

続いて、各次元の下限值をNOT演算した値のサフィックス列が取得される。X軸については、下限値「001」をNOT演算した値が「110」であり、この値から得られるサフィックス列は、「0」、「10」、「110」である。Y軸についても同様に、サフィックス列「0」、「10」、「110」が取得される。

【0062】

20

取得された各サフィックス列に下限値と「1」とが加算された値が境界値候補として取得される。X軸及びY軸の各境界値候補は以下のように取得される。

X軸のサフィックス列「0」⇒境界値候補「010」

X軸のサフィックス列「10」⇒境界値候補「100」

X軸のサフィックス列「110」⇒境界値候補「1000」

Y軸のサフィックス列「0」⇒境界値候補「010」

Y軸のサフィックス列「10」⇒境界値候補「100」

Y軸のサフィックス列「110」⇒境界値候補「1000」

【0063】

このように取得された境界値候補のうち、未踏検索空間における各軸の下限值及び上限値の範囲に含まれる境界値候補が、その次元の境界線の正側（X軸の右側、Y軸の上側）に位置する値（正側境界値）と推定される。なお、上述の境界値から1を減算した値は、その次元の境界線の負側（X軸の左側、Y軸の下側）に位置する値（負側境界値）となる。

30

X軸の正側境界値＝「010」、「100」

X軸の負側境界値＝「001」、「011」

Y軸の正側境界値＝「010」

Y軸の負側境界値＝「001」

【0064】

ここで、X軸の正側境界値と検索空間におけるY軸の下限值とで示されるデータ点、及び、Y軸の正側境界値と検索空間におけるX軸の下限值とで示されるデータ点が、その部分検索空間の最小データ点となる。

40

第1最小データ点（「010」、「001」）＝「001001」

第2最小データ点（「100」、「001」）＝「100001」

第3最小データ点（「001」、「010」）＝「000110」

【0065】

これにより、インデックス走査装置10は、網掛けで示される既に特定された部分空間以外の残空間が少なくとも3つに分割されていると決定することができる。この3つの残空間が部分検索空間として検索空間保持部25に追加される。この部分検索空間の情報としては、上述のように特定された最小データ点が含まれてもよい。これにより、その最小

50

データ点がその部分検索空間の検索点に設定されてもよい。

【0066】

また、X軸の負側境界値と検索空間におけるY軸の上限値とで示されるデータ点、及び、検索空間の最大データ点により、その部分検索空間の最大データ点が取得されてもよい。

第1最大データ点（「001」、「011」）＝「000111」

第2最大データ点（「011」、「011」）＝「001111」

第3最大データ点（「101」、「011」）＝「100111」

【0067】

このように決定される部分検索空間を以降の検索範囲とすることにより、クエリ50に対応する検索範囲に対応する元の検索空間をより狭い範囲検索へと絞り込むことになる。即ち、インデックス格納部21において、検索点（第1最小データ点）「001001」とプレフィックスマッチするインデックスデータとして、インデックスデータ「001***」が特定される。更に、検索点（第2最小データ点）「100001」とプレフィックスマッチするインデックスデータとして、インデックスデータ「10****」が特定される。更に、検索点（第3最小データ点）「000110」とプレフィックスマッチするインデックスデータとして、インデックスデータ「0001**」が特定される。なお、本実施例では、このように特定された各部分空間により新たに推定される部分検索空間はない。

【0068】

このように、本実施例によれば、図6に示されるインデックスデータのエントリのうち、インデックスデータ「0000**」、「001***」、「10****」、「0001**」を含むエントリのみにアクセスすれば、クエリ50の検索範囲に対応する目的のデータページの集合を得ることができる。

【0069】

上述の例では、最初の検索点が、検索空間に含まれるデータ点の中の最小データ点「000011」に設定されたが、最大データ点「100111」に設定されてもよい。この場合には、次のようにインデックス走査が実行される。

【0070】

（S34）において、検索点「100111」とプレフィックスマッチするインデックスデータとして、インデックスデータ「10****」が特定される。

【0071】

図7は、最初に特定された部分空間とこの部分空間から推定される境界線を概念的に示す図である。図7の網掛けで示される領域がインデックスデータ「10****」で特定される部分空間である。インデックス走査装置10は、特定されたインデックスデータ「10****」と関連付けられたページ特定データ「0x06a8」を抽出する（S35）。インデックス走査装置10は、抽出されたページ特定データ「0x06a8」をページ特定データ保持部23に保持させる。

【0072】

インデックス走査装置10は、検索空間内における、上述のように特定された部分空間（インデックスデータ「10****」）以外の残空間（未踏検索空間）を部分検索空間に決定し（S36）、特定された部分空間に基づいて、次のように境界線を推定する。このとき、未踏検索空間における各軸の上限値及び下限値が取得される。ここでは、X軸については、下限値「001」、上限値「011」が取得され、Y軸については、下限値「001」、上限値「011」が取得される。

【0073】

続いて、各次元の上限値をNOT演算した値のサフィックス列が取得される。X軸については、上限値「011」をNOT演算した値が「100」であり、この値から得られるサフィックス列は、「0」、「100」である。Y軸についても同様に、サフィックス列「0」、「100」が取得される。

【0074】

ここで、取得されたサフィックス列のうち、未踏検索空間における各軸の上限値以上のサフィックス列が除外される。ここでは、X軸のサフィックス列「100」及びY軸のサフィックス列「100」が除外される。そして、各軸の上限値から、各サフィックス列及び「1」が減算された値が境界値候補として取得される。X軸及びY軸の各境界値候補は以下のように取得される。

X軸のサフィックス列「0」⇒境界値候補「010」

Y軸のサフィックス列「0」⇒境界値候補「010」

【0075】

このように取得された境界値候補のうち、未踏検索空間における各軸の下限值及び上限値の範囲に含まれる境界値候補が、その次元の境界線の正側（X軸の右側、Y軸の上側）に位置する値（正側境界値）と推定される。なお、上述の境界値から1を減算した値は、その次元の境界線の負側（X軸の左側、Y軸の下側）に位置する値（負側境界値）となる。

X軸の正側境界値＝「010」

X軸の負側境界値＝「001」

Y軸の正側境界値＝「010」

Y軸の負側境界値＝「001」

【0076】

ここで、X軸の正側境界値と検索空間におけるY軸の下限值とで示されるデータ点、Y軸の正側境界値と検索空間におけるX軸の下限值とで示されるデータ点、及び、検索空間の最小データ点が、その部分検索空間の最小データ点となる。

第1最小データ点（「010」、「001」）＝「001001」

第2最小データ点（「001」、「010」）＝「000110」

第3最小データ点（検索空間の最小データ点）＝「000011」

【0077】

〔比較例〕

以下、上述の各実施形態及び実施例の効果を示すために、本実施形態とは異なるインデックス走査方法を比較例として説明する。以下の説明においても、図4及び図5で示される、2次元空間と2次元範囲検索、並びに、インデックス格納部21が用いられる。

【0078】

比較例の手法は、インデックス格納部21における、検索空間の最大データ点に対応するインデックスデータを含むエン트리からその最小データ点に対応するインデックスデータを含むエン트리までの全エントリを走査対象とする。つまり、図5に示されるエントリのうち、インデックスデータ「0000**」を含むエン트리からインデックスデータ「10****」を含むエン트리までの全エントリに関して、対応する部分空間と検索空間とが重なっていないかがチェックされる。

【0079】

ところが、上述したように、特定されるべき部分空間は、インデックスデータ「0000**」、「001***」、「10****」、「0001**」に対応する4つの部分空間しかない。よって、比較例の手法によれば、インデックスデータ「010000」からインデックスデータ「011111」に対応する各部分空間は、重複チェックの対象となるが、検索空間と重複しておらず、特定されないことになる。

【0080】

つまり、比較例の手法によれば、検索対象範囲は空間全体に比べて狭いにもかかわらず、インデックス走査される部分空間数が、インデックス格納部21に格納される全インデックスに対応する部分空間数と同程度となる場合があり得る。その結果、比較例によれば、インデックスの走査に時間がかかってしまうことになる。

【0081】

一方で、上述の各実施形態及び実施例によれば、クエリの検索範囲に対応する検索空間

10

20

30

40

50

と少なくとも1部が重なる部分空間のみを対象にしてインデックス走査が行われるため、インデックス走査が高速化される。このような各実施形態及び実施例と比較例との効果の差は、検索対象の次元数が大きい場合や、検索時に指定する各次元の範囲の幅に差がある場合に、顕著となる。

【0082】

なお、上述の説明で用いた複数のフローチャートでは、複数のステップ（処理）が順番に記載されているが、本実施形態で実行される処理ステップの実行順序は、その記載の順番に制限されない。本実施形態では、図示される処理ステップの順番を内容的に支障のない範囲で変更することができる。また、上述の各実施形態及び各変形例は、内容が相反しない範囲で組み合わせることができる。

10

【0083】

上記の各実施形態及び各変形例の一部又は全部は、以下の付記のようにも特定され得る。但し、各実施形態及び各変形例が以下の記載に限定されるものではない。

【0084】

（付記1）

複数の部分空間に分割された多次元空間にマッピングされたデータベースにおける、該複数の部分空間の各々を特定するための各インデックスデータを、各部分空間に属するデータが格納されるデータページを特定するための各ページ特定データと関連付けて格納するインデックス格納部と、

クエリの検索範囲に対応する前記多次元空間内の検索空間の情報、又は、該検索空間内の部分検索空間の情報を取得する検索空間取得部と、

20

前記検索空間又は前記部分検索空間に含まれるデータ点のうち、前記多次元空間の各データ点を一次元化することにより得られる値が最小及び最大の少なくとも一方となるデータ点を検索点として設定する検索点設定部と、

前記インデックス格納部において、前記検索点設定部により設定された前記検索点を含む部分空間のインデックスデータを特定し、該特定されたインデックスデータと関連付けられたページ特定データを前記検索空間の少なくとも一部と重なる部分空間のページ特定データとして抽出するインデックス検索部と、

前記インデックス検索部により特定されたインデックスデータから得られる部分空間の特定情報に基づいて、前記検索空間又は前記部分検索空間を通る、前記多次元空間を前記複数の部分空間に分割する境界線を推定し、該推定された境界線により前記検索空間又は前記部分検索空間を分割することにより得られる分割検索空間のうち、前記検索点を含む部分空間以外の分割検索空間を前記部分検索空間として決定する検索領域決定部と、

30

を備えるインデックス走査装置。

【0085】

（付記2）

前記各インデックスデータは、各部分空間の位置及び範囲を特定し得るデータ形式を有する付記1に記載のインデックス走査装置。

【0086】

（付記3）

40

前記多次元空間の空間分割には、KD（K-dimensional）-Treeが利用され、

前記多次元空間の一次元化には、空間充填曲線が用いられ、

前記多次元空間の各データ点を一次元化することにより得られる値は、2進数表記され、

前記インデックス格納部は、各部分空間に含まれるデータ点の最長共通プレフィックスを各インデックスデータとし、

前記検索領域決定部は、各軸の各データ点のうち、最下位桁から上位桁方向に向けて初めて0が現れる桁位置までの全桁に1が設定されている値で示されるデータ点と、該値に1を加算して得られる値で示されるデータ点との間を前記境界線として推定する、

50

付記 1 又は 2 に記載のインデックス走査装置。

【0087】

(付記 4)

複数の部分空間に分割された多次元空間にマッピングされたデータベースにおける、該複数の部分空間の各々を特定するための各インデックスデータを、各部分空間に属するデータが格納されるデータページを特定するための各ページ特定データと関連付けて格納するインデックス格納部を走査するインデックス走査方法において、

クエリの検索範囲に対応する前記多次元空間内の検索空間の情報、又は、該検索空間内の部分検索空間の情報を取得し、

前記検索空間又は前記部分検索空間に含まれるデータ点のうち、前記多次元空間の各データ点を一次元化することにより得られる値が最小及び最大の少なくとも一方となるデータ点を検索点として設定し、

前記インデックス格納部において、前記検索点設定部により設定された前記検索点を含む部分空間のインデックスデータを特定し、

前記特定されたインデックスデータと関連付けられたページ特定データを前記検索空間の少なくとも一部と重なる部分空間のページ特定データとして抽出し、

前記特定されたインデックスデータから得られる部分空間の特定情報に基づいて、前記検索空間又は前記部分検索空間を通る、前記多次元空間を前記複数の部分空間に分割する境界線を推定し、

前記推定された境界線により前記検索空間又は前記部分検索空間を分割することにより得られる分割検索空間のうち、前記検索点を含む部分空間以外の分割検索空間を前記部分検索空間として決定する、

ことを含むインデックス走査方法。

【0088】

(付記 5)

前記各インデックスデータは、各部分空間の位置及び範囲を特定し得るデータ形式を有する付記 4 に記載のインデックス走査方法。

【0089】

(付記 6)

前記多次元空間の空間分割には、KD (K-dimensional) - Tree が利用され、

前記多次元空間の一次元化には、空間充填曲線が用いられ、

前記多次元空間の各データ点を一次元化することにより得られる値は、2 進数表記され、

前記インデックス格納部は、各部分空間に含まれるデータ点の最長共通プレフィックスを各インデックスデータとし、

前記境界線の推定は、各軸の各データ点のうち、最下位桁から上位桁方向に向けて初めて 0 が現れる桁位置までの全桁に 1 が設定されている値で示されるデータ点と、該値に 1 を加算して得られる値で示されるデータ点との間を前記境界線として推定する、

付記 4 又は 5 に記載のインデックス走査方法。

【0090】

(付記 7)

コンピュータに、

複数の部分空間に分割された多次元空間にマッピングされたデータベースにおける、該複数の部分空間の各々を特定するための各インデックスデータを、各部分空間に属するデータが格納されるデータページを特定するための各ページ特定データと関連付けて格納するインデックス格納部と、

クエリの検索範囲に対応する前記多次元空間内の検索空間の情報、又は、該検索空間内の部分検索空間の情報を取得する検索空間取得部と、

前記検索空間又は前記部分検索空間に含まれるデータ点のうち、前記多次元空間の各デ

10

20

30

40

50

ータ点を一次元化することにより得られる値が最小及び最大の少なくとも一方となるデータ点を検索点として設定する検索点設定部と、

前記インデックス格納部において、前記検索点設定部により設定された前記検索点を含む部分空間のインデックスデータを特定し、該特定されたインデックスデータと関連付けられたページ特定データを前記検索空間の少なくとも一部と重なる部分空間のページ特定データとして抽出するインデックス検索部と、

前記インデックス検索部により特定されたインデックスデータから得られる部分空間の特定情報に基づいて、前記検索空間又は前記部分検索空間を通る、前記多次元空間を前記複数の部分空間に分割する境界線を推定し、該推定された境界線により前記検索空間又は前記部分検索空間を分割することにより得られる分割検索空間のうち、前記検索点を含む部分空間以外の分割検索空間を前記部分検索空間として決定する検索領域決定部と、

を実現させるプログラム。

【0091】

(付記8)

前記各インデックスデータは、各部分空間の位置及び範囲を特定し得るデータ形式を有する付記7に記載のプログラム。

【0092】

(付記9)

前記多次元空間の空間分割には、KD (K-dimensional) - Tree が利用され、

前記多次元空間の一次元化には、空間充填曲線が用いられ、

前記多次元空間の各データ点を一次元化することにより得られる値は、2進数表記され、

前記インデックス格納部は、各部分空間に含まれるデータ点の最長共通プレフィックスを各インデックスデータとし、

前記検索領域決定部は、各軸の各データ点のうち、最下位桁から上位桁方向に向けて初めて0が現れる桁位置までの全桁に1が設定されている値で示されるデータ点と、該値に1を加算して得られる値で示されるデータ点との間を前記境界線として推定する、

付記7又は8に記載のプログラム。

【0093】

(付記10) 付記7から9のいずれか1つに記載のプログラムを記録したコンピュータが読み取り可能な記録媒体。

【0094】

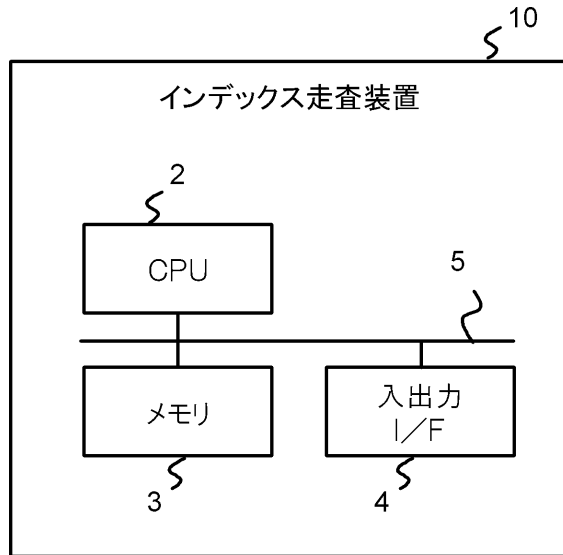
この出願は、2012年1月13日に出願された日本出願特願2012-005562号を基礎とする優先権を主張し、その開示の全てをここに取り込む。

10

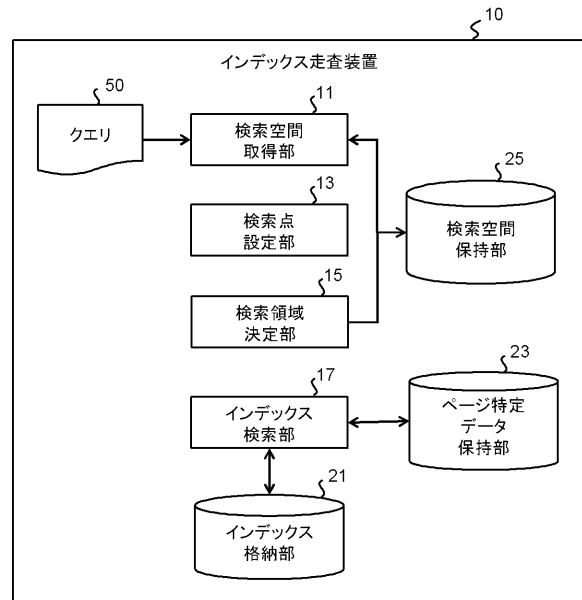
20

30

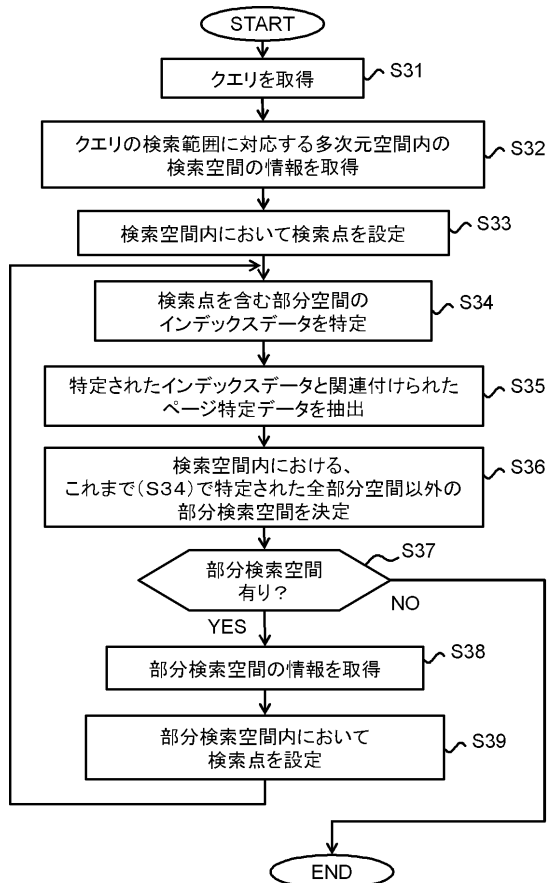
【図 1】



【図 2】



【図 3】



【図 4】

	010	010	011	011	110	110	111	111
111	101	111	101	111	101	111	101	111
	010	010	011	011	110	110	111	111
110	100	110	100	110	100	110	100	110
	010	010	011	011	110	110	111	111
101	001	011	001	011	001	011	001	011
	010	010	011	011	110	110	111	111
100	000	010	000	010	000	010	000	010
	000	000	001	001	100	100	101	101
011	101	111	101	111	101	111	101	111
	000	000	001	001	100	100	101	101
010	100	110	100	110	100	110	100	110
	000	000	001	001	100	100	101	101
001	001	011	001	011	001	011	001	011
	000	000	001	001	100	100	101	101
000	000	010	000	010	000	010	000	010
	000	001	010	011	100	101	110	111

【図 5】

インデックスデータ	ページアドレス
0000**	0x0001
0001**	0x005b
001***	0x008e
010000	0x0103
010001	0x0114
010010	0x012a
⋮	⋮
011100	0x0???
011101	0x0???
011110	0x0???
011111	0x0???
10****	0x06a8
11****	0x0???

【図 6】

111	010 101	010 111	011 101	011 111	110 101	110 111	111 101	111 111
110	010 100	010 110	011 100	011 110	110 100	110 110	111 100	111 110
101	010 001	010 011	011 001	011 011	110 001	110 011	111 001	111 011
100	010 000	010 010	011 000	011 010	110 000	110 010	111 000	111 010
011	000 101	000 111	001 101	001 111	100 101	100 111	101 101	101 111
010	000 100	000 110	001 100	001 110	100 100	100 110	101 100	101 110
001	000 001	000 011	001 001	001 011	100 001	100 011	101 001	101 011
000	000 000	000 010	001 000	001 010	100 000	100 010	101 000	101 010
	000	001	010	011	100	101	110	111

【図 7】

111	010 101	010 111	011 101	011 111	110 101	110 111	111 101	111 111
110	010 100	010 110	011 100	011 110	110 100	110 110	111 100	111 110
101	010 001	010 011	011 001	011 011	110 001	110 011	111 001	111 011
100	010 000	010 010	011 000	011 010	110 000	110 010	111 000	111 010
011	000 101	000 111	001 101	001 111	100 101	100 111	101 101	101 111
010	000 100	000 110	001 100	001 110	100 100	100 110	101 100	101 110
001	000 001	000 011	001 001	001 011	100 001	100 011	101 001	101 011
000	000 000	000 010	001 000	001 010	100 000	100 010	101 000	101 010
	000	001	010	011	100	101	110	111

フロントページの続き

(56)参考文献 特開2001-209651 (JP, A)

Shoji Nishimura, Sudipto Das, Divyakant Agrawal, Amr El Abbadi, MD-HBase: A Scalable Multi-dimensional Data Infrastructure for Location Aware Services, Mobile Data Management (MDM), 2011 12th IEEE International Conference on, 米国, IEEE, 2011年 6月 9日, p. 7-16, [online], [平成28年9月1日検索], <URL:<http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6068416>>

上川 伸彦、岩崎 一正, 画像の類似検索に向けた多次元インデックス手法, 情報処理学会論文誌, 日本, 社団法人情報処理学会, 2001年 1月15日, 第42巻、No. SIG1 (TOD8), p. 140-147

中台 慎二、谷口 邦弘, PLATON: 超分散環境におけるデータ共有のためのP2P多次元範囲検索システム, マルチメディア, 分散, 協調とモバイル (DICOMO2007) シンポジウム論文集, 日本, 社団法人情報処理学会, 2007年 6月29日, Vol. 2007, No. 1, p. 173-184

(58)調査した分野(Int.Cl., DB名)

G06F 17/30