フレームデータモデルに基づく機械材料データベースにおけるデータの検索*

小野 陽*¹, 稲田 康徳*¹

Data Retrieval in a Material Database with Frame Data Model

Noboru ONO and Yasunori INADA

From database consisting of a set of frame-represented material data, data retrieval is made by determining data objects which are subsumed by the object that is also a frame-representation of a query. The discrimination of data can be successfully made down to every detail of data objects such as, for example, the order of successively given thermo-mechanical treatments, their temperatures, and the dimensions of test specimens. Nevertheless, the fact that expected slot values are often missing in original report on the results of material testings impairs the completeness of data retrieval. This problem is coped with through the use of default reasoning, so that to prepare a set of appropriate default rules is an important task to render the database system more useful.

Key Words: Material Testing, Artificial Intelligence, Database System, Material Database, Frame, Data Retrieval, Default Reasoning

1. はじめに

一般にデータベースの利用者が知りたい事実そのものが収録、検索されるデータベースは文献検索のためのデータベースなど知りたい情報の所在を知らせる案内データベースと区別してファクトデータベースと呼ばれ、その充実が望まれている⁽¹⁾。機械材料の諸特性に関してもそれは同様だが、それら、特に機械的な性質の記述には材料自身の組成や履歴、試験環境、それに試験方法や特性値の定義など様々な事項が含まれ、さらにそのような事項はデータごとに必ずしも同じでない。そのため機械材料のファクトデータベースの実現には通常のデータベースシステムの表現と検索の能力は不十分であると感じられる⁽²⁾⁽³⁾。そこで我々はフレームデータモデルによる材料データの記述を試み、適切なフレームデザインによって自然言語、表、図などを用いて行われる材料試験結果の報告の内容をその細部にわたってほぼ余すところなく計算機中のデータオブジェクトとして表現、記録できることを示した⁽³⁾。これに対応して本報告ではそのようなデータベース

における検索の基本的な論理とアルゴリズム、期待する検索結果を得るために必要な付加的な知識と推論についてまとめる。そこではデータの記述におけるのと同じ表現力が問い合わせの記述にも許され、極めて詳細なデータの弁別が可能になる。しかしそのように強力な問い合わせの記述能力と検索の能力のために、検索結果の意味についてそれだけ慎重な解釈が必要となる。特に材料データの記述においてはスロット値の欠落がしばしば存在して⁽³⁾、そのために検索の完全性が損なわれる場合がある。これを避けるためには欠落したスロットを常識によって自動的に補うこと、すなわちデフォルト推論が不可欠である。それを含めた検索実験の結果を挙げて本システムの検索能力を実証し、問題点を考察する。

2. データベースシステム

本研究で用いたデータベースシステムの概略を図1に示す。前報⁽³⁾でフレームデータモデルを用いたデータベースシステムにおいては、フレームの定義がそこに含まれるフレーム名とスロット名も含めて材料データを記述するための言語の役割を果たすと述べた。それらはシステムデータベースの一つ、フレーム定義データベースに蓄えられる。例えば材料はその素材、加えられた処理、化学組成、それに呼び名の4つをス

* 平成3年3月16日 東北支部第26期総会講演会において講演、原稿受付 平成3年2月1日。

*¹ 正員、東北大学工学部 (〒980 仙台市青葉区荒巻字青葉)。

(a) データの入力

(b) データの検索と編集

図1 データベースシステムの構成

ロットに持つフレームにより記述できると考えた。そのフレーム名をmaterialとしてこれを

```
material(raw_material,process,
          chemical_composition,name)
```

と書く。同様に熱処理

```
heat_treatment(name,temperature,
                duration,atmosphere)
```

などのフレーム定義を与えた。さらに、材料データ自身も試験に用いられた試験片、試験の環境、そこで観察された特性の記述をスロットとして持つフレームのように定義した。

個々のデータはそこに現れる個々の事項のフレーム表現を他のフレームのスロット値とする事で互いに連結され、最終的にフレームdataを根とする木構造のオブジェクトとしてオブジェクトデータベースに記憶、蓄積される。前報で用いたのと同じ文献⁽⁴⁾の焼入れと時効を受けたTi-6Al-4V合金の降伏応力のデータのフレーム表現をPROLOGリストで書き、PROLOGの事実節として記録した例を図2に挙げた。

一方、データベースシステムに必要なもう一つ機能であるデータの検索も同じくフレーム定義に沿って検

```
object(data,d6,
      [ [specimen,object(s37)],[property,object(p26)]]).

object(specimen,s37,
      [ [made_of,object(m36)],
        [process,
          [machining,
            [direction,longitudinal],
            [geometry,
              [round_bar,
                [ [gauge_diam,3],
                  [gauge_length,12]]]]]]]]].

object(yield_strength,p26,
      [ [yield_strength,1000],
        [definition,
          [proof_stress,
            [ [offset,0.2]]]]]]].

object(material,m36,
      [ [process,
          [heat_treatment,
            [ [cooling_rate,air_cooling],
              [duration,7200],
              [temp,973]]]],
          [raw material,object(m35)]]].

object(material,m35,
      [ [process,
          [heat_treatment,
            [ [cooling_rate,water_quench],
              [duration,3600],
              [temp,1233]]]],
          [raw material,object(m34)]]].

object(material,m34,
      [ [name,ti-6al-4v],
        [process,object(c32)],
        [chemical_composition,
          [chemical_composition,
            [ [ti,61],[al,6.1],[v,4.1],[fe,0.21],
              [c,0.01],[n,0.026],[o,0.198],[h,0.0177],
              [w,0.004]]]],
          [raw material,object(s33)]]].

object(powder,s33,
      [ [name,prealloyed REP powder],
        [made_from,[material,[name,ti-6al-4v]]],
        [process,rotary electrode process]]].

object(hip,c32,
      [ [temp,1188],
        [duration,21600],
        [pressure,1],
        [geometry,
          [cylinder,
            [ [diam,95],
              [length,200],
              [weight,4]]]]]]].
```

図2 HIPと熱処理を受けた粉末冶金Ti-6Al-4V合金の降伏応力のフレーム表現を記録したPROLOGの事実節

索したいデータのフレーム表現を作ることで行われる。検索プログラムはその表現を受け取り、それとマッチするデータをデータベース中から探し出して出力する。そのような検索のための論理とアルゴリズム、問題点について次章以下で考察する。

3. データ検索の論理とアルゴリズム

自然言語のフレーズにせよ計算機中のデータオブジェクトにせよ、材料の特性の記述において現れるあらゆる「表現」はそれぞれに物質、操作、事象などの物理的実体の集合を指し示すものと考えられる⁽⁵⁾。我々の実験システムにおける事象のフレーム表現あるいはそのPROLOGリストによる記述を用いて例を挙げれば

O1: [material,[[name,ti-6al-4v]]]

はTi-6Al-4Vと呼ばれる全ての合金の集合を指し示す抽象的な概念である。もし、

O2: [material,[[name,ti-6al-4v],
[process,hip]]]

と書けば、溶解で作られたTi-6Al-4Vは排除され、O1の集合の部分集合を指し示すことになる。このようにある事物に関して記述が詳細あるいは限定的になるほどそれだけ小さな集合を規定することになる。例えばHIP処理の条件が一部書き込まれた次のオブジェクト

O3: [material,[[name,ti-6al-4v],
[process,hip,[[pressure,1]]]]]

はさらにO2の部分集合になる、

これから推察されるように、ある事物の記述がより小さな集合を指すことはそのフレーム表現においては、イ) スロットが値を持つことおよび、ロ) スロット値として与えられたオブジェクトがより限定的なものであること、に他ならない。したがってあるオブジェクトQが他のオブジェクトDを包含することは基本的に次の規則により判定して良いであろう。

M1: 「QとDのフレーム名が同じであり、かつQが持つ全てのスロットについてそのスロット値がDの対応するスロットのスロット値を包含する時QはDを包含する」

あるオブジェクトのスロット値は再びオブジェクトであるから、その間の包含関係の判定にはこの規則を再帰的に用いれば良い。すると判定は最終的にPROLOGアトムあるいは文字列同士の比較、あるいは数値や数値範囲同士の比較に行き着く。前者の間のマッチングはそれらを再びQ、Dと書けば規則

M2: 「QとDが同一であるかあるいは包含関係の予備知識、ako(Q,D)が別に定義されている時QはDを包含する」

によって行う必要がある。これは例えば $\alpha+\beta$ 型Ti-合金と呼ばれる材料の集合がより具体的にTi-6Al-4Vと呼ばれる材料の集合の上集合であるといった場合である。数値については例えば

M3: [range,[[min,900]]] $\supset$
[range,[[min,950],[max,1000]]] $\supset$
973

のような包含関係があるのは明かである。

上で述べたように我々のデータベースシステムでは個々のデータもフレーム表現されたオブジェクトとしてデータベースに記憶されている。それらのデータは材料と試験片、試験環境、試験方法や結果など試験結果を記述するのに必要な全ての項目について可能な限

り具体的、限定的な記述が与えられたオブジェクトであり、したがってあらゆる材料データという集合の中の極めて小さな部分集合になっている。これに対して問い合わせは一般にデータの記述よりも大ざっぱだからデータベースにおけるデータの検索を「与えられた問い合わせのオブジェクトについてそれに包含されるデータオブジェクトを弁別し、出力すること」と定義して良いであろう。この時の包含関係を判定するデータオブジェクトの間のマッチング規則には上で述べた規則、M1からM3を用いることが出来る。これらの規則にしたがって包含関係の判定を行うマッチングプログラムを書くことは容易である。データの記述に用いられる語彙の間の包含関係、言い替えればそれらの分類はPROLOGの事実節 $ako(A,B)$ の形で図1の語彙分類データベースに蓄えられる。

4. 欠落したスロット

4・1 欠落したスロットの一般的取扱い 2で述べたマッチング規則M1では、Qでは欠落しているがDには存在するスロットがマッチングにおいて問題にされないことに注意しよう。これは例えば processのスロットが欠落したO1が問い合わせの中に現れたとき製造方法や加えられた処理は何でもかまわず、あらゆるTi-6Al-4Vに興味があると解釈する事に相当する。

逆にDに欠落したスロットがある場合にはこの規則ではマッチングは失敗する。その意味は次のごとくである。例としてデータと問い合わせの中の材料の記述がそれぞれO2、O3であった場合を考えよう。このデータはある特定の圧力でHIPされた材料から得られたものに違いないが、その圧力がこのオブジェクトが指し示すあらゆる圧力の中のいずれであるかを知ることはできない。したがってとりたてて圧力1kbarでHIPされた材料への興味を表現するO3の問い合わせに対して積極的な情報を提供しないし、もし検索され出力されればそれは誤った情報であるかも知れない。したがってこの場合にはマッチングが失敗し、そのデータが検索されないのが検索アルゴリズムの望まれる動作であり、上で述べた検索規則M1からM3はそれを保証する。このような理由によるマッチングの失敗を以下では「未知によるマッチングの失敗」と呼ぶことにする。

4・2 デフォルト推論 前報¹³⁾で述べたように、実際の材料試験結果の記述にはしばしば記述の欠落があり、したがってそのフレーム表現にもスロットの欠落がしばしば現れる。そのためその全てを前節の検索規則で処理すると多くのデータが検索されなくなる。例えば図2に挙げたHIPしたTi-6Al-4V合金の降

伏応力の報告⁽⁴⁾には試験環境の記述がない。したがってそのフレーム表現でもそのスロットを欠落したままにするのが原情報に最も忠実である。一方、室温の空気中での結果にだけ興味があるとき、液中や高温での実験の結果が検索結果に含まれることを防ぐために問い合わせの記述の中に室温の空気中という項を含めなければならない。その結果上の検索規則により常識的に室温の空気中で得られたと考えられるこのデータは検索されなくなる。このような不都合を避けるためにはデータにおいて欠落したスロットを常識によって補うことが必要である。これを一般にデフォルト推論⁽⁶⁾と言うが、それが全ての欠落したスロットに適用できるとは限らない。それを前報および図2に見られる例について検討する。

4・2・1 自明な記述の省略 上の例の試験環境の記述のように何も書かないことの意味が明かな場合その記述を省略することが多い。これ以外の例を以下に挙げる。

a. 材料の名前は始めの材料、あるいは受け入れ材についてだけ与え、その後の処理を受けた後の材料の名前のスロットは欠落させるものとしたが⁽⁵⁾、その処理によって材料の呼び名は変わらないことを前提として、処理前の材料の名前を補って良いであろう。溶解や圧粉焼結など呼び名を変える可能性のある製造プロセスがある時はその直後の材料に呼び名が与えられるから問題は起こらないであろう。

b. 材料の化学組成も処理履歴の中の始めの材料についてだけ与えられることが多い。これも名前と同様に処理前の材料の組成を参照することで良いと考えられる。この操作の根拠はひとつには名前の場合と同じくその処理が組成を変化させないと考えることだが、これは正しくない場合がある。もうひとつの根拠は化学組成を指定した問い合わせが「そのような組成を持っていた材料」についての興味を示しており、材料試験時の組成を指定しているわけでは必ずしもないと考えられるというものである。

c. 前報で挙げた例では疲労試験片の表面状態の記述が無かった。この欠落は常識的には試験の結果に影響を及ぼさない程度に十分に平滑に仕上げられたことを意味すると考えられるが、この推測は例えば試験環境の場合ほど確かではない。しかしこれを欠落したまま残すと、粗い表面を持った試験片のデータを検索結果から排除し、「十分に平滑な」試験片のデータだけを検索する事ができなくなる。この問題は定性的な粗さの記述、例えばsmoothを導入し、それを表面粗さのデフォルト値とすることなどで解決できるだろう。

4・2・2 結果に影響を及ぼさない因子 材料試験を行うに当たって試験結果に影響を及ぼさないと考えられる因子は測定あるいは制御されず、したがって不明であることがある。例えば空気中で行われた実験では正確な温度は不明な場合が多い。分析されなかった元素の含有量もこの例である。同様な理由で分かっている報告には含まれないものもある。例えば図2の材料の記述には本来必要と考えられる熱処理雰囲気⁽⁷⁾の記述がない。これは熱処理後の素材から試験片を切り出しているため、雰囲気による表面近傍の組成や組織の変化が結果に影響を及ぼさないことによる。

そのスロットの値が特性に影響を及ぼさないならば、あらゆる値を持った問い合わせに対してこのデータがマッチして良いようにも思われる。しかしある因子が本当に特性に影響を及ぼさないのか、あるいは誤ってそう考えられただけなのかを個々の場合について判断することは明らかに高度に専門的な知識を必要とする作業である。したがって取りあえずここでは、このようなスロットに対しては上の不知によるマッチングの失敗を適用するものとする。

上の雰囲気の温度や化学組成の記述の欠落は大まかには推定でき、その範囲内での変化は特性に影響を及ぼさないと考えられたとも受け取れる。その意味で前項で挙げた例との間の明確な区別が難しい場合もあるかも知れない。

5. 処理履歴の選行

データと問い合わせの材料のマッチングにおいて期待する検索結果を得るために、名前や化学組成のスロットの欠落とは別に処理前の材料の記述を参照する必要がある場合がある。それは素材に与えられた処理自身である。例えば次の材料 MAT#1,

MAT#1: [material,[[raw_material,MAT#2],

[process,[heat_treatment,HT#1]]]]

がデータベースに含まれており、問い合わせ MAT#2,

MAT#2: [material,[[name,ti-6al-4v],

[process,[hip]]]]

つまりHIPで作ったTi-6Al-4Vへの関心が与えられた場合、HIPで製造された Ti-6Al-4Vを熱処理した材料MAT#1は検索結果に出力されるべきだと普通考える。しかしスロット processのスロット値が一致しないためにM1の検索規則によってマッチングは失敗する。同様なことが素材、raw_materialのスロット値についても起こる。

このマッチングを成功させるためには、上で述べた材料の名前と化学組成と同様に処理、素材に関しても

「ある材料とのマッチングに失敗したら、処理前の材料とのマッチングを試みよう」という規則をつけ加えればよい。このように材料の属性が処理の後も保存されることはマッチングアルゴリズムの上では処理自身、化学組成、素材それに名前について共通していて都合がよい。このアルゴリズムを「処理履歴の遡行」と呼ぶことにしよう。しかしそれが持つ意味はそれぞれについて同じではない。

一方MAT#2の問い合わせがHIPしたままの材料に対する興味である場合も考えられる。ある処理を「最後に受けた」材料と「受けたことがある」材料の間のこのような区別は検索者の意図だけによる。したがってその区別は問い合わせオブジェクトにおいてだけ必要であり、この区別をする副詞句「最後に」の働きはスロットの働きとは異なるものと考えられる。したがって今まで述べてきた枠組みの中ではこの違いを検索プログラムに知らせる手だてがなく、それを行う有効な手段は現在検討中である。

6. 検索プログラムと検索実験の結果

上のふたつの章での考察をまとめると2章で述べた検索の基本的なアルゴリズムにつけ加えるべき付加的なマッチング規則は取りあえず、

1) 試験環境と試験片の表面粗さについてのデフォルト規則、および

2) 材料についての処理履歴の遡行規則
のふたつである。そこでこれらを含めたマッチングプログラムを作成した。付加的なマッチング規則はいずれも一般的な規則であるためにその実現に大きな困難はない。ある処理を「最後に」受けた材料の表現の実現は現在のところできていない。したがって検索の結果には意図する以外のデータも含まれる可能性があることになる。

このプログラムを用いて図2のデータオブジェクトとのマッチングを行わせたいくつかの問い合わせオブジェクトをマッチングの結果とともに図3に示す。すなわち、(a)では試験環境に対するデフォルト規則と処理履歴の遡行規則が正しく働いていることがわかる。(b)では図2の材料と与えられた熱処理の順序が異なるためにマッチングは失敗する。

7. 考 察

上で見たようにデータベースのデータの記述に用いられたフレームデータモデルとフレームデザインをそのまま用いることでデータの記述で許されるのと同じだけ詳細な問い合わせの記述が可能になる。これに語

```
object(data,q1,
[[specimen,
[specimen,
[[raw material,
[material,
[[name,ti-6al-4v],
[process,[hip]]]]]]],
[environment,ambient],
[property,
[yield_strength,
[[yield_strength,[range[[min,800]]]]]]]]].
```

(a) マッチングに成功する問い合わせオブジェクト

```
object(data,q2,
[[specimen,
[specimen,
[[raw material,
[material,
[[process,
[heat_treatment,
[[cooling_rate,water_quench]]]],
[raw material,
[material,
[[process,
[heat_treatment,
[[cooling_rate,air_colling]]]],
[raw material,
[material,
[[name,ti-6al-4v]]]]]]]]]]]]].
```

(b) マッチングに失敗する問い合わせオブジェクト

図3 問い合わせオブジェクトの例

彙の包含関係の定義およびデフォルト推論などの付加的な知識を加えることでデータの検索における強力なデータの識別能力が実現される。この能力は試験片、材料とその処理、試験の環境や方法、それに得られた特性のいずれについてもその数値的な細部に至るまで検索の条件にする事ができるという点では、実際に原論文を見ながら必要なデータを選別するのに匹敵する。さらに、それが自動的になされるという点ではもっと便利である。

しかしながら別な見方をすればこのことは実際に原論文の記述を見ることなくデータの選別が行われることを意味するから、検索の結果の正しさについて通常のデータベースの場合より慎重な考慮が必要であろう。ここで検索の正しさを次のふたつの条件で定義することにしよう⁽⁷⁾。

イ) 再現率: データベース中の検索者が必要とする全データ数に対する検索された必要なデータの割合。

ロ) 精度: 検索された全データ数に対する検索者が必要とするデータの割合。

本研究のデータベースシステムはデータと問い合わせの記述に十分な表現能力を与え、またそれに見合ったマッチング手続きを用意することで、この両者を同時に満足する、いわば完全な検索結果を得ようとする試みに他ならない。上で述べたようにその目標はかなりの程度達成できたと考えられが、一方いくつかの理由で検索の完全性が損なわれる場合があることも指摘しておく必要があるだろう。

まずデータと問い合わせの表現力が十分でない場合検索結果は不正確になるのはやむを得ない。元の記述にあいまいさがある場合⁽⁵⁾や、ある処理を最後に受けた材料の記述ができないことなどがこれに当たる。このような場合、フレーム定義や語彙分類の知識が許す範囲で近似的な表現を用いるしかないが、その時は再現率と精度を同時に保証することはできなくなる。例えば最後の処理の問題についての現在の処置は再現率を保証するために精度を犠牲にしている。

元の記述における記述の欠落はあいまいさの一種であるとも考えられ、上と同様に検索の完全性に影響を及ぼす。データにおいて欠落したスロットに対する不知によるマッチングの失敗の論理は検索の精度を上げる反面、再現率を下げる。デフォルト推論はこれを補い、再現率を回復する働きをする。上で述べた試験環境や試験片の表面粗さなどのように一般的なデフォルト規則の他に、例えば「アルミニウム合金の熱処理における雰囲気」の記述の欠落には「空气中」を補えのように、ある条件下だけで働くデフォルト規則も有り得る。このような限定された条件下でのデフォルトを個別に検討して必要なデフォルト規則を設けていくことは検索の再現率を高めるためにぜひ必要な作業であり、またそのようなデフォルト推論規則を書いたデフォルト規則データベースは図1にも書いたように本データベースシステムの重要な構成要素である。

デフォルト規則の適用条件も基本的にはデータや問い合わせの記述と同じ詳細さが許されるから、極めて特殊なデフォルト規則を設けることも可能である。しかしそのような個別の規則を全て用意することは多分不可能だし、3・2・2で述べたようにデフォルト値が明かでない場合もある。またある規則に少数の例外があっても再現率を高めるためにそれを設けた方がよい場合もあるかも知れない。このように検索におけるデフォルト推論は極めて微妙であり、多くのデータを収録しさまざまな問い合わせに応じる中でそれを調整して再現率と精度の最適のバランスを選ぶのが良いであろう。その意味で元の情報の記述を記述の欠落やあいまいさも含めて忠実に再現してそのまま利用者の閲覧に供する事が重要だと考えられる。

8. まとめ

上述のように強力な記述と検索の能力のために生じる検索結果の完全性についての微妙な問題に注意を払えば、フレームをデータモデルとする材料データベースは機械工学の多くの分野で材料データの有効利用に

役立つものと期待される。すなわち前報で述べたように、規格化された材料の標準的な特性のデータベースは現在の関係データベースを用いても実現が可能である。そこでこのような詳細な記述と検索を可能にするデータベースは逆に規格外の材料の特性、あるいは標準的でない特性のデータベースを構築するために有効である。材料開発やその評価のための基礎データ、低温や高温それに腐食環境などでのデータ、摩耗など多くの入り組んだ影響因子を持つ現象のデータ、多軸応力や変動する応力下でのデータなどがそれに当たるだろう。多数の実験結果の系統的な整理を必要とされる材料の信頼性評価^{(9)~(11)}においてもデータベースが用いられるが、影響因子の細かな弁別が新たな発見をもたらす可能性があるだろう。実在するデータの集録だけでなく、材料の振舞いについての理論や経験式を用いてデータベースに存在しない特性を予測する試みもある⁽⁸⁾。これらの予測はもちろん存在するデータに基づいて行われる。したがって予測システムがより幅広い材料と条件をカバーしようとしたとき、本研究のようなデータベースは予測システムに対して適切なデータを供給するための基礎データデータベースの役割を果たすものと考えられる。

文 献

- (1) 山内, 小野寺, 情報管理, 27(1984), 674.
- (2) 野上, 岩田, 「新材料開発と材料設計学」, 三島編, ソフトサイエンス社, (1985), p.444.
- (3) 小野, 機論, 投稿中
- (4) D. Eylon, et al., Metall. Trans., 14A(1983), 2497.
- (5) 小野, 人工知能学会全国大会(第3回)論文集, (1989), p.247.
- (6) R. Reiter, Artificial Intelligence, 13(1980) 81.
- (7) 山崎, 知的検索の技術, 講談社, (1972)
- (8) 石井, 二瓶, 金沢, 西島, 材料, 39(1990), 1285.
- (9) 西島, 機誌, 90(1987), 1296.
- (10) 菊川, 畑中, 三沢, 材料, 38(1989), 1151.
- (11) 平田, 岡部, 村松, 材料, 39(1990), 26.