

情報メディア学会 20191102

材料科学分野のための総合的記述言語 tq

物質・材料研究機構: 天野晃; 坂本浩一; 鈴木晃; 松田朝彦
筑波大学: 鈴木伸崇

構成

- 背景
- 目的
- 言語紹介
- 開発状況

背景:

材料科学基盤はマルチスケール・マルチフィジクス

スケール	技術(例)	理論(例)	主な力
核子	格子ゲージシミュレーション	量子色力学	強い力
原子	量子化学計算	量子力学	電磁気力
分子	密度汎関数法	量子力学	電磁気力
分子の集団	分子動力学	古典力学	電磁気力
組織	FEM/CA/CALPHAD	複合	電磁気力
部材	FEM	複合	電磁気力
構造物	FEM	複合	電磁気力/重力
構造物+環境	FEM/ナビエ・ストークスEq	複合/流体/粉体	電磁気力/重力

※結晶、準結晶もある

背景:

マルチスケールの例

高温超伝導開発は、ミクロ(原子レベル)な特性がマクロな特性(組織レベル)に影響を及ぼす現象を研究する。

→ 電子の振る舞い(が、フェルミ-ディラック統計かボーズ-アインシュタイン統計か)というミクロな性質が、マクロな抵抗という性質の起因となる。

背景:

マルチフィジクスの例

固体電池開発は、電磁気特性だけでなく、固体の相を保つという、別の側面においても開発要素がある。

背景: 情報学の視点から

MI (マテリアルズインテグレーション/マテリアルズインフォマティクス) を加速させるためには、マルチスケール・マルチフィジクス (つまり異なる科学分野) の知識/データ形式を同じ枠組で取り扱う必要がある

- すでに行われているチャレンジである
- サービスや実装もある
 - modelica
 - Wolfram Data Repository
- しかし、我々は、従来とは異なる手法を用いる
 - 従来はデータを直接的に統一形式に統合する方式を用いていたのに対して
 - 「データ形式」と「データ形式定義」を分離する
 - 「知識」を統一せず、「知識の記述形式」を統一する

ゴール:

あらゆる科学分野の知識を統一的/形式的に記述可能な言語、それによるデータ形式定義記述の実現

→ 「トランスレショナルリサーチ」は、どちらかといえば垂直展開を意味するが、我々は「垂直水平同時展開」を可能とするトランスレショナルリサーチの「形式的」な枠組みを必要としている、とも言える。

- 木構造とグラフ構造を同じ構文で記述する
 - データ形式（テンソル）と辞書（グラフ）を同じシンタックスで記述する必要がある。
- パーサーとソルバーを分離する
 - 知識は不整合や矛盾を含むものである。ソルバーがあると成功しない。
 - その代わり、既存ソルバーと接続できるようにする。
- 知識は抽象度を高くし、即値を含ませない（例：円周率は”円周率”であり3.14ではない）。ただし、言語としての量指定子が必要。
 - その代わり、即値（実データ）をバインドできるようにする。
- 当然、辞書を解析し単位変換等の項目間関係性を取り出したい

出口:

あらゆる目的において、「同等」なデータを「同等でない形式」のデータから抽出
／統合し、機械学習用データとして再構成する
→ 機械のためのデータフェデレーション

- 入出力関係、報酬関係も記述可能
- データ次元も記述可能
- 「同等」かどうかを判定する根拠(辞書)も記述可能
- 必要ならば学習用ネットワークも記述可能

目的:

- 知識とデータ構造を形式的に記述できる言語の開発（本日の話）
 - 1) 項参照と、2) データバインドにより、
 - 知識はトリプルに断片化。
 - 全体でグラフ構造を構成。
 - 木（テンソル）のノードにデータをバインドしてデータ構造を表現。
- 辞書形式の開発（本日の話）
 - 当該言語を用いて辞書を記述する。
- 辞書アナライザの開発
 - 辞書のグラフ構造を解析する。

tq:

① $\langle tq \rangle :: \langle label \rangle > \langle reference \rangle > \langle operator \rangle > \langle name \rangle > \langle bind \rangle$
($\langle tq \rangle > , \dots$)

e.g.

#1\$#2\$Op\$Name(#2Name2[2])

#1\$#2\$Op\$Name@#2Name2@#2Name2(#2Name2,[2]@(Length,Weight))

tq: 項参照

表現	参照部	被参照部	被参照部バインド表現
\$#1f	\$#1		\$#1f
(\$#1,#1)	\$#1	#1	(\$#1@#1,#1)
\$#1f(#1g)	\$#1	#1g	\$#1@#1g(#1g)
#1f(\$#1g)	\$#1	#1f	#1f(\$#1g@#1f)

tq: データバインド

表現	データ	データバインド表現
[1]	L, W, 22, 3, 21, 5	[1]⊙(L)
[2]	L, W, 22, 3, 21, 5	[2]⊙(L, W)
[2]([2])	L, W, 22, 3, 21, 5	[2]([2]⊙(L, W, 22, 3))
H1[2](H2[2])	L, W, 22, 3, 21, 5	H1[2](H2[2]⊙(L, W, 22, 3))

tq: 項参照とデータバインドを使ったデータの再構成

データ: Length,Weight,mm,kg,1,2 322,4,5,68

入力形式定義: (#1[2],#2[2],[3](#4[2]))

出力形式定義: \$PI\$(\$#1,Quantity(\$#4,\$#2))

出力:

```
( (Length,Quantity(1,mm)),(Weight,Quantity(2,kg)) ),  
( (Length,Quantity(322,mm)),(Weight,Quantity(4,kg)) ),  
( (Length,Quantity(5,mm)),(Weight,Quantity(68,kg)) ) )
```

※改行は便宜上

tq: トリプル表現

	P(Arrow)	S(Dom)	O(Cod)
P(S,O)	P	S	O
(S(\$#1P),#1O)	P	S	O
S(\$#1P) / (S(\$#1P),)	P	S	<無名>
S(O)	<無名>	S	O
(,)	<無名>	<無名>	<無名>

tq: 辞書表現

グラフ: $\$G\$((\text{<オブジェクトリスト>}),(\text{<トリプルリスト>}))$
↓
そのまま!! ただし、辞書トリプルの S と O には木構造を許す
辞書: $\$D\$((\text{<オブジェクトリスト>}),(\text{<トリプルリスト>}))$

e.g.
 $((\#1A, \#2B, \#3C), (f(\$ \#1, \$ \#2), \$X\$near(\$ \#2, \$ \#3),$
 $\$def\$(\underline{f}, (\underline{(a, b)}, \underline{\$eq\$ (a, b)})), \dots))$
 $((\#1A, \#2B, \#3C), (f(\$ \#1, \$ \#2), \$X\$near(\$ \#2, \$ \#3),$
 $\$def\$(\underline{f}, (\underline{(a, b)}, \underline{\$pow\$ (b, 2)})))$, ...))

開発状況: 実装済み部分

- 木構造パーサー
- ラベルによる項参照
- データバインディング
- データリストの内積化
- アンパックス (木構造の平坦化)
- クォータライズ
- リテラライズ (オペレーション回避)