

소재 연구데이터 표준어휘 사전 작성 가이드

Guide for Building the Standard Vocabulary
Dictionary of Materials Research Data

2024.7.5.

소재 연구데이터 표준화 전문위원회

목 차

어휘 사전의 작성 규칙 -----	2
(부록) 허용 unit 일람표 -----	5
(부록) 소재 데이터의 예 (가상의 데이터) -----	8

어휘 사전의 작성 규칙

1. 본 소재 데이터 표준어휘 사전은 json 포맷을 사용하여 소재 데이터의 구조와 소재 데이터에서 사용되는 keyword 들을 정의한다.
2. 각 데이터 keyword 들은 아래의 정보를 담고 있다.

- eng_definition : keyword의 영문 설명
- alias : keyword와 동일한 의미로 사용되는 용어
- data_type : keyword에 해당하는 데이터의 형태
- data_unit : keyword에 해당하는 데이터의 단위
- data_example : keyword에 해당하는 데이터의 예시

(중요) 이들 명칭을 이용하여 parsing한 어휘 사전의 정보를 바탕으로 플랫폼이 구축된다. 따라서, 위 명칭들은 어휘 사전에서 오류없이 사용되어야 한다.

3. 국제표준어휘와의 연동을 위해 keyword는 모두 영문어휘를 사용한다.
4. 본 사전에 사용되는 모든 keyword는 고유명사를 제외하고 소문자로 작성한다.

예) chemical information; Czochralski growth method

(※) 소재 연구데이터 플랫폼에서는 keyword 명칭을 바탕으로 내부변수를 생성하여 사용한다. 내부변수는 표준어휘 keyword의 단어와 철자를 그대로 사용하되 모든 단어의 첫 자는 대문자로 하고 띄어쓰기 없이 붙여서 생성한다.

예) “structural property” → StructuralProperty

“Czochralski growth method” → CzochralskiGrowthMethod

5. keyword의 괄호로 표시된 영역은 실제 DB에서는 해당 keyword로 대체된다. 같은 성격의 반복되는 키워드는 뒤에 “_(n)”을 붙여 표시한다. 실제 DB에서는 “(n)”은 자연수로 대체된다.

예)

본 사전 내의 keyword	실제 DB 데이터 예
(process)	“chemical synthesis”
(analysis)	“optical microscopy”
(unit)	“g/mole”
materials_(n)	“materials_1”
precursor_(n)	“precursor_3”

6. data_type으로는 아래의 것들이 있다.

- * string : 문자열 데이터
- * numeric : 수치 데이터
- * numeric array : 테이블형 수치 데이터
- * list : [data_1, data_2, data_3,...,data_n]
- * dictionary : {key_1:data_1, key_2:data_2, , key_n:data_n}

7. 모든 numeric 데이터의 data_unit은 “(부록) 사용 unit 일람표”에 근거하여 표시한다.

8. 모든 numeric 데이터는 value와 uncertainty를 가진다. uncertainty는 absolute uncertainty를 value와 같은 unit의 값으로 제공한다.

예))

```
▼ melting temperature : {  
  eng_definition : temperature at the materials melt  
  alias : value  
  ▼ value : {  
    eng_definition : data value  
    data_type : numeric  
    data_unit : K  
    data_example : 2500  
  }  
  ▼ uncertainty : {  
    eng_definition : uncertainty of the value  
    data_type : numeric  
    data_unit : same as value  
  }  
}
```

9. numeric array 데이터는 “numeric data expression” section의 “numeric array data/format”의 포맷으로 기술한다.

10. dictionary data_type인 경우에는 각 key와 해당 데이터 형태를 정의하여야 한다.

예))

```
▼ composition : {  
  eng_definition : constituent of materials  
  alias : value  
  ▼ value : {  
    eng_definition : data value  
    data_type : dictionary  
    {constituent:constituent_quantity,...,unit:string,uncertainty:  
      uncertainty value}  
    data_example : {Pt:56.0, Ni:34.7, unit:at.%, uncertainty:0.1};  
                   {Fe2O3:97.8, Y2O3:2.2, unit:wt.%, uncertainty:0.2}  
  }  
}
```

11. data_example의 예시들은 세미콜론 (;) 으로 구분한다. 복수 개의 alias 역시 세미콜론 (;)으로 구분한다.

예)) annealed INCONEL625 plate; Al7075-T6; Al10Si5Mg; Fe; TiO2 on CNT

12. array로 제시되는 data_example은 제시된 list 중 하나를 데이터로 갖는다. 따라서 platform에서는 data_example들을 바탕으로 selection menu를 구현한다.

예))

```
▼ data example [4]
  0 : interstitial
  1 : substitutional
  2 : vacancy
  3 : cluster
```

13. unit는 superscript만을 사용하여 표시한다.

(예))

$$\begin{array}{ll} V/m \rightarrow V \, m^{-1} & m^2/s^2V \rightarrow m^2 \, s^{-2} \, V^{-1} \\ m/s \rightarrow m \, s^{-1} & m^3/C \rightarrow m^3 \, C^{-1} \\ A/m^2 \rightarrow A \, m^{-2} & m^{1/3}/C \rightarrow m^{1/3} \, C^{-1} \\ mm/year \rightarrow mm \, year^{-1} & \end{array}$$

14. 특수기호와 수식은 latex 표기법을 따른다.

https://en.wikipedia.org/wiki/Help:Displaying_a_formula#Alphabets_and_typefaces

(예))

\alpha	\mu	\epsilon	\theta	\lambda	\sigma	\cdot	\log	\text{\AA}
α	μ	ϵ	θ	λ	σ	\cdot	log	Å

(부록) 허용 unit 일람표

7 SI base unit을 바탕으로 하되, 재료과학 분야에서 널리 사용되는 단위들의 사용을 허용한다.

red = 7 SI base units

Time		Length		Mass	
year	year	m	meter	kg	kilogram
d	day	cm	centi meter	g	gram
h	hour	mm	milli meter	mg	mili gram
min	minute	{Wmu m}	micro meter	m_{e}	electron mass
s	second	nm	nano meter	Da	dalton
ms	mili second	WANGSTROM	Angstrom		
ns	nano second				
fs	femto second				

Electricity		Thermodynamic Temperature / Energy / Power		Amount	
A	Ampere	eV	electron volt	mol	mole
mA	milli Ampere	keV	kilo electron volt		
{Wmu A}	micro Ampere	meV	mili electron volt		
nA	nano Ampere	J	Joule		
pA	pico Ampere	kJ	kilo Joule		
V	volt	pJ	pico Joule		
kV	kilo volt	W	watt		
mV	millivolt	K	kelvin degree		
C	coulomb	kW	kilo watt		
{Wmu C}	micro Coulomb	MeV	mega electron volt		
pC	pico Coulomb	mW	mili watt		
e	charge of an electron				
S	conductance (Siemens)				
Ohm	resistance				

Luminous Intensity		Force / Pressure		Others	
cd	candela	N	Newton	at. %	atomic per cent
		{Wmu N}	micro Newton	wt. %	weight per cent
		kgf	kilo gram force	ppm	part per million
		gf	gram force	%	per cent
		Pa	pascal	HR	Hardness scale
		MPa	mega pascal	HB	
		GPa	giga pascal	HV	
		Torr	Torr	HK	
		bar	Bar	Jones	specific detectivity
		atm	atmosphere	Hz	Hertz
				kHz	kilo Hertz
				revolution	revolution, rotation
				degree	geometric angle
				L	liter
				mL	milli liter
				cycle	number of repeat
				T	tesla
				{Wmu L}	micro liter
				GHz	giga Hertz

(부록) 소재 데이터의 예 (가상의 데이터)

```
▼ 0 {4}
  ► meta {5}
  ► material_1 {5}
  ► material_2 {2}
  ► system {4}
```

각각의 데이터는 meta, materials_n, system 데이터 군으로 구성된다.

```
▼ meta {5}
  data ID : DOI:111.222.333.444
  data name : Pt3Ni particle on graphite support data 1
  ▼ contributor {2}
    name : Jung Hoon Lee
    affiliation : Computational Science Center, KIST
  data generation date : 2020-04-28
  note on data : information of material_2 comes from Aldrich
                  catalog\n10.1021/nl401881z
  ► material_1 {5}
  ► material_2 {2}
  ► system {4}
```

탄소를 지지체로 하는 Pt-Ni 나노입자 촉매 시스템의 데이터. meta 정보군 (meta) 과 두 개의 재료 정보군 (material_1 및 materials_2) 그리고 촉매시스템 정보군 (system) 으로 구성되어 있다.

Meta 정보군은 이 데이터에 관한 정보로서 data ID, data name, contributor, data generation date, note on data로 구성된다.


```

▼ meta {5}
  data ID : DOI:111.222.333.444
  data name : Pt3Ni particle on graphite support data 1
  ► contributor {2}
  data generation date : 2020-04-28
  note on      : information of material_2 comes from Aldrich
  data        : catalog\n10.1021/nl401881z
▼ material_1 {4}
  name : Pt3Ni
  ► chemical information {2}
  ► process [3]
  ► property {1}
▼ material_2 {2}
  name : C
  ► chemical information {1}
► system {4}

```

material_1과 material_2는 각각 Pt3Ni 나노입자 물질의 정보와 탄소 지지체 정보를 가진다. 각각의 재료는 화학 정보 (chemical information), 합성 공정 (process) 그리고 재료의 특성 (property)를 가지고 있다. 위의 예에서 materials_1은 화학 정보, 공정정보 그리고 재료 특성을 데이터로 갖고 있고, material_2는 화학 정보 만을 갖고 있다.

```

▼ material_1 {4}
  name : Pt3Ni
  ▼ chemical information {2}
    ► composition {3}
    ► crystallography {3}
  ► process [3]
  ► property {1}
▼ material_2 {2}
  name : C
  ▼ chemical information {1}
    ► composition {2}

```

material_1의 화학 정보에는 조성 (composition)과 결정구조 (crystallography) 정보를 가지고 있고, materials_2는 조성 정보 만을 가지고 있다.

```

▼ material_1 {4}
  name : Pt3Ni
  ▼ chemical information {2}
    ▼ composition {3}
      Pt   : 75
      Ni   : 25
      unit : at.%
    ► crystallography {3}
  ► process [3]
  ► property {1}
▼ material_2 {2}
  name : C
  ▼ chemical information {1}
    ▼ composition {2}
      C    : 100
      unit : at.%

```

material_1의 조성은 Pt 75 at.%, Ni 25 at.% 이다. material_2의 화학조성은 100 at.% 탄소이다.

```

▼ material_1 {4}
  name : Pt3Ni
  ▼ chemical information {2}
    ► composition {3}
    ▼ crystallography {3}
      Bravis lattice : cubic
      ▼ lattice parameter {1}
        a : 3.897
      ▼ measurement [2]
        ▼ 0 {1}
          ▼ X-ray diffraction {2}
            instrument : PANalytical X'Pert PRO Alpha-1
                        diffractometer
            image : image_xrd.png
          ▼ 1 {1}
            ▼ TEM {2}
              instrument : JEOL ARM200F microscope
              image : image_tem.png
        ► process [3]
        ► property {1}
  ▼ material_2 {2}
    name : C
    ▼ chemical information {1}
      ► composition {2}
  ► system {4}

```

material_1의 결정구조는 cubic이며 lattice parameter는 3.897 Angstrom이다. 결정구조와 lattice parameter는 X-ray diffraction과 TEM으로 측정하였다. 다수의 측정조건은 measurement의 array로 저장된다. 측정의 조건과 미세조직 사진 혹은 spectrum은 각 분석데이터의 공통어휘에 맞추어 추가된다.

```

▼ material_1 {4}
  name : Pt3Ni
  ► chemical information {2}
  ▼ process [3]
    ▼ 0 {1}
      ► solvothermal {8}
    ▼ 1 {1}
      ► heat treatment {10}
    ▼ 2 {1}
      ► centrifugation {5}
  ► property {1}
▼ material_2 {2}
  name : C
  ▼ chemical information {1}
    ► composition {2}
  ► system {4}

```

material_1은 solvothermal 공정으로 합성한 뒤 열처리 공정과 원심분리 공정을 거쳐 만들어 진다. 각 공정 object는 재료공정 공통어휘에 맞추어 작성되며, 공정 순서대로 process의 array로 저장된다.

```

▼ material_1 {4}
  name : Pt3Ni
  ► chemical information {2}
  ► process [3]
  ▼ property {1}
    ▼ structural property {3}
      ▼ morphology {1}
        ▼ particle {3}
          shape : octahedron
          ▼ size {1}
            ▼ diameter {2}
              value : 9
              uncertainty : 0.03
            ▼ measurement {1}
              ► TEM {2}
          ▼ 1D defect {5}
            type : edge dislocation
            Burgers vector : 1/3[110]
            ▼ density {2}
              value : 100000000
              uncertainty : 1000
            ▼ velocity {2}
              value : 9.2
              uncertainty : 0.05
            ▼ measurement {1}
              ► TEM {2}
          ► impurity {3}

```

material_1은 두 가지의 구조적 특성 데이터, 형상 (morphology)과 1차원 결함 (1D defect: dislocation) 및 impurity 데이터를 갖는다. 형상은 입자로서 octahedron 모양의 지름 9nm이며 TEM으로 관찰한 결과이다. 1D defect는 edge dislocation으로서 버거스벡터 1/3[100], 밀도 $1e8/m^2$, 이동속도 9.2 m/sec 이다. 역시 TEM으로 관찰한 결과이다. 측정값의 불확도가 있는 경우에는 불확도 정보를 동반한다.

```

▼ material_1 {4}
  name : Pt3Ni
  ► chemical information {2}
  ► process [3]
  ▼ property {1}
    ▼ structural property {3}
      ► morphology {1}
      ► 1D defect {5}
      ▼ impurity {3}
        0 : 0.34
        unit : ppm
        uncertainty : 0.001
  ▼ material_2 {2}
    name : C
    ▼ chemical information {1}
      ► composition {2}
  ► system {4}

```

Materials_1의 impurity는 산소 0.34 ppm이 존재하며, 측정값의 불확도는 0.001 ppm이다.

```

▼ meta {5}
  data ID : DOI:111.222.333.444
  data name : Pt3Ni particle on graphite support data 1
  ► contributor {2}
  data generation date : 2020-04-28
  note on data : information of material_2 comes from Aldrich catalog\n10.1021/nl401881z
  ► material_1 {4}
  ► material_2 {2}
▼ system {4}
  description : Pt3Ni/C catalyst system for CO2 reduction
  ► configuration {3}
  ► process [3]
  ► performance {1}

```

material system 정보는 description, configuration, process 그리고 performance로 구성된다. description에서는 materials system에 대한 간략한 설명을 포함한다. configuration은 material_1와 material_2가 어떻게 material system을 구성하는지를 기술한다. process는 재료의 공정과 마찬가지로 material system을 구축하는 공정을 순차적으로 기술한다. performance는 재료 시스템의 성능을 기술한다. configuration과 performance 정보군은 관찰하고자 하는 재료의 활용조건에 따라 고유한 구조와 어휘를 가진다.

```

▼ system {4}
  description : Pt3Ni/C catalyst system for CO2 reduction
  ▼ configuration {3}
    active material : Material_1
    amount of active material : 0.003
    support material : Material_2
  ► process [3]
  ► performance {1}

```

이 시스템의 active material은 Material_1 (Pt3Ni) 이고 support material은 Material_2 (carbon) 이다. 이 시스템에서 active material 대 support material의 비율은 0.003이다.

```

▼ system {4}
  description : Pt3Ni/C catalyst system for CO2 reduction
  ► configuration {3}
  ▼ process [3]
    ▼ 0 {1}
      ► sonication {4}
    ▼ 1 {1}
      ► centrifugation {4}
    ▼ 2 {1}
      ► rinsing {3}
  ► performance {1}

```

이 시스템의 제조를 위해 sonication, centrifugation 그리고 rinsing 공정이 사용되었다. 각 공정의 데이터는 소재공정 공통어휘에 맞추어 구성되었다. (아래 참조)

```

▼ process [3]
  ▼ 0 {1}
    ▼ sonication {4}
      ▼ material mix {2}
        Material_1 : 0.02
        Material_2 : 0.03
        solvent : toluene
        temperature : 298
        time : 3
      ▼ 1 {1}
        ▼ centrifugation {4}
          revolutions per minute : 12000
          time : 10
          temperature : 333
          ▼ additive_1 {2}
            name : acetic acid
            amount : 0.02
        ▼ 2 {1}
          ▼ rinsing {3}
            temperature : 343
            time : 0.5
          ▼ additive_1 {1}
            name : ethanol

```



```

▼ system {4}
  description : Pt3Ni/C catalyst system for CO2 reduction
  ► configuration {3}
  ► process [3]
  ▼ performance {1}
    ▼ electrochemical {1}
      ▼ ORR {3}
        ► area-specific activity {3}
        ► mass-specific activity {3}
        ► Faradaic efficiency {3}

```

촉매로서의 electrochemical 성능의 측정은 ORR 데이터를 제공한다. area-specific activity, mass-specific activity 그리고 Faradaic efficiency 데이터를 제공하며 각 데이터는 데이터 값과 측정 조건 데이터를 가진다. 측정조건 데이터는 분석방법 공통어휘에 맞추어 작성된다. (아래 참조)

```

▼ performance {1}
  ▼ electrochemical {1}
    ▼ ORR {3}
      ▼ area-specific activity {3}
        value : 49
        uncertainty : 0.1
      ▼ measurement {1}
        ▼ electrochemical activity {5}
          instrument : CHI 730C potentiostat
          temperature : 298
          voltage : 0.9
          graph image : image_electro_chem.png
          analysis method : potentiostat
      ▼ mass-specific activity {3}
        value : 0.4
        uncertainty : 0.001
      ▼ measurement {1}
        ▼ electrochemical activity {5}
          instrument : CHI 730C potentiostat
          temperature : 298
          voltage : 0.9
          graph image : image_electro_chem.png
          analysis method : potentiostat
      ▼ Faradaic efficiency {3}
        value : 90
        uncertainty : 0.5
      ▼ measurement {1}
        ▼ gas chromatography {6}
          instrument : CHI 730C potentiostat
          flow rate : 5
          temperature : 298
          voltage : 0.9
          graph image : image_electro_chem.png
          analysis method : gas chromatography

```