

Accelerating Materials Genome Ontology Development

Alden Dima

Benjamin Long

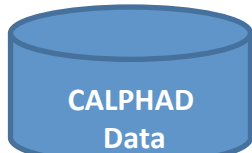
Sharief Youssef

Information Technology Laboratory

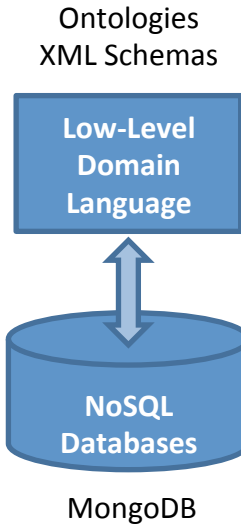
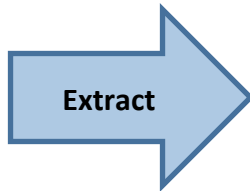
National Institute of

Standards and Technology

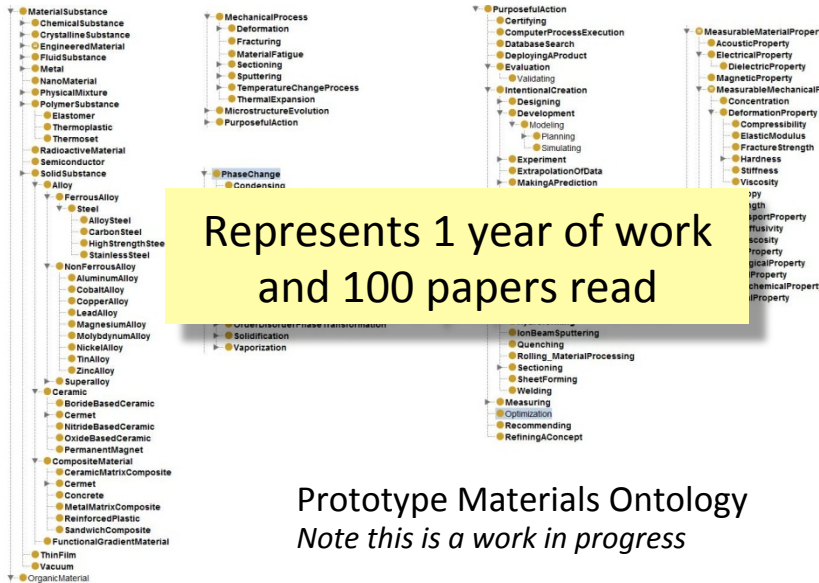
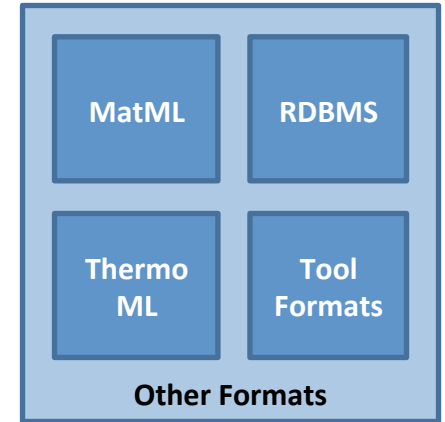
Informatomics Approach



Lab Equipment
Computation
Text files
Spreadsheets
Diagrams
Articles
Databases

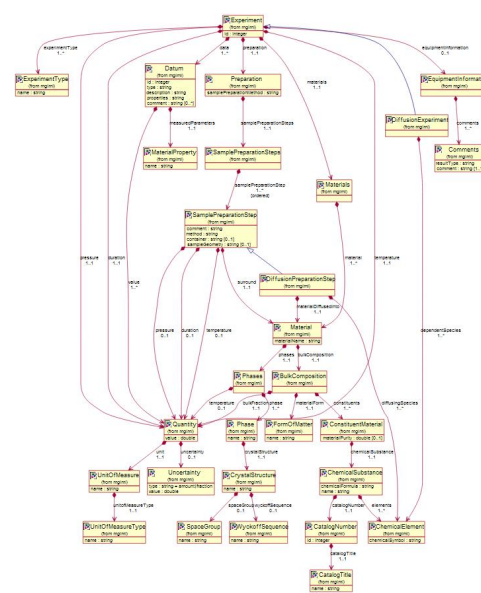


XSLT
Reasoners



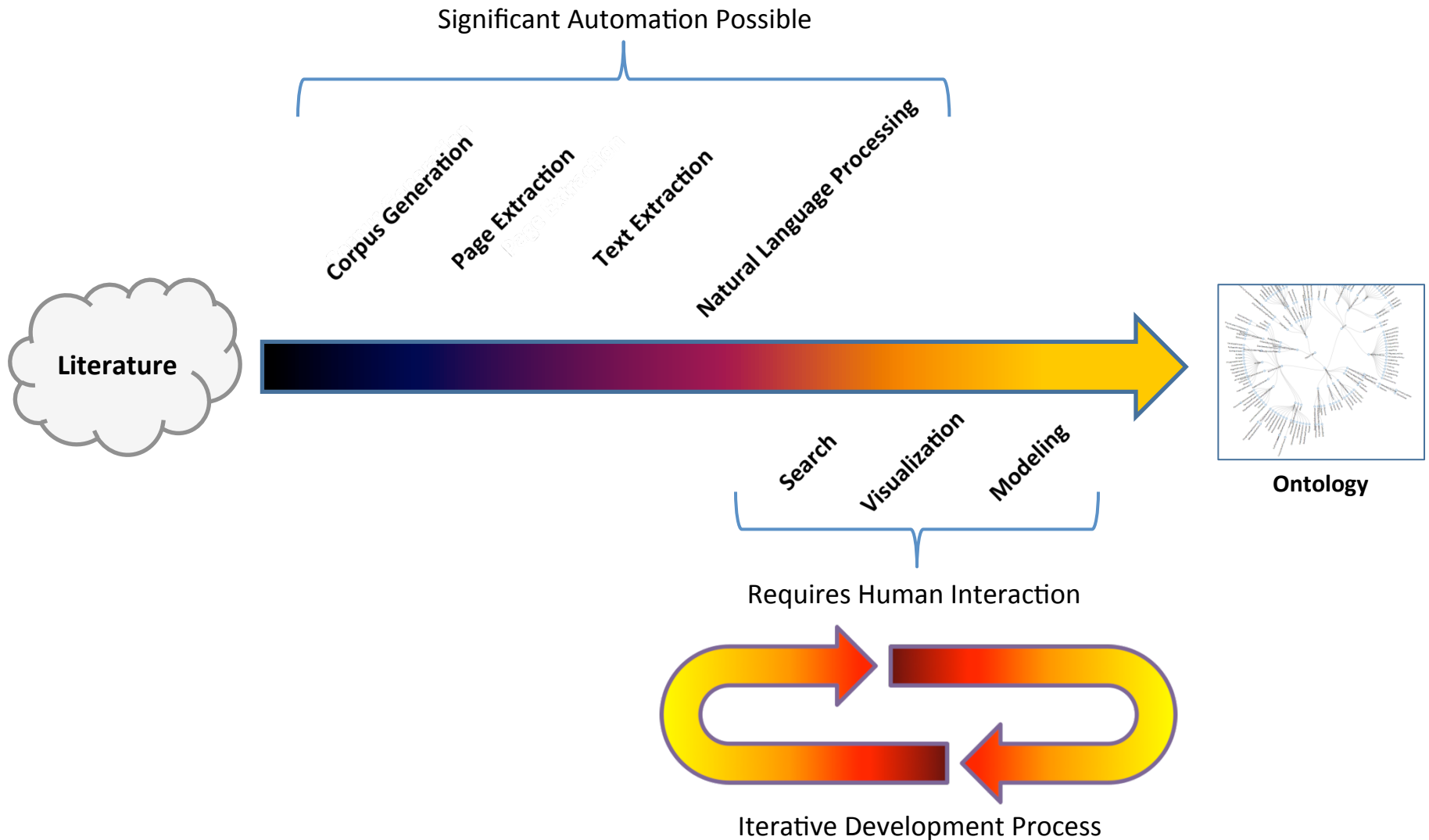
Represents 1 year of work
and 100 papers read

Section of XML Schema



Prototype Materials Ontology
Note this is a work in progress

Informatics for Ontology Acceleration



Facilitating Materials Ontology Development via NLP and Machine Learning

Starting with a corpus of 5893 PDFs

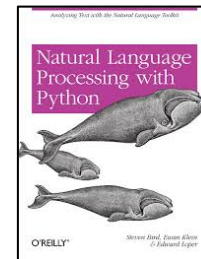


2011, 2012



2009, 2010, 2011, 2012

NIST Diffusion Data Center



radimrehurek.com/gensim
www.nltk.org

www.matscitech.org www.tms.org patapsco.nist.gov/diffusion

5893 PDFs - 5.8 GB

63K pages }
3.8M lines } 53GB

3.4M lines extracted

20M tokens

temperatur	66132
materi	53364
process	47990
diffus	45893
result	43681
surfac	33328

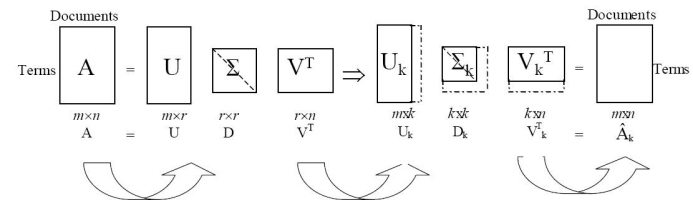
Extract text^{1,2}

Identify key stems
by frequency³

Generate concordances for key stems³

tivity method. Lattice diffusion coefficients and grain boundary diffusion coefficients were measured. Plots of the data for all the alloys, the diffusion coefficients at temperatures 1150 'C and higher, were only accepted. From these data, the parameters of temperature dependence of the diffusion coefficients were determined. For those alloys the self-diffusion coefficients were also determined at temperatures 1150 'C and higher, the following relation coefficients at temperatures below 1150 'C were

Use Latent Semantic Indexing to group similar concordance entries⁴

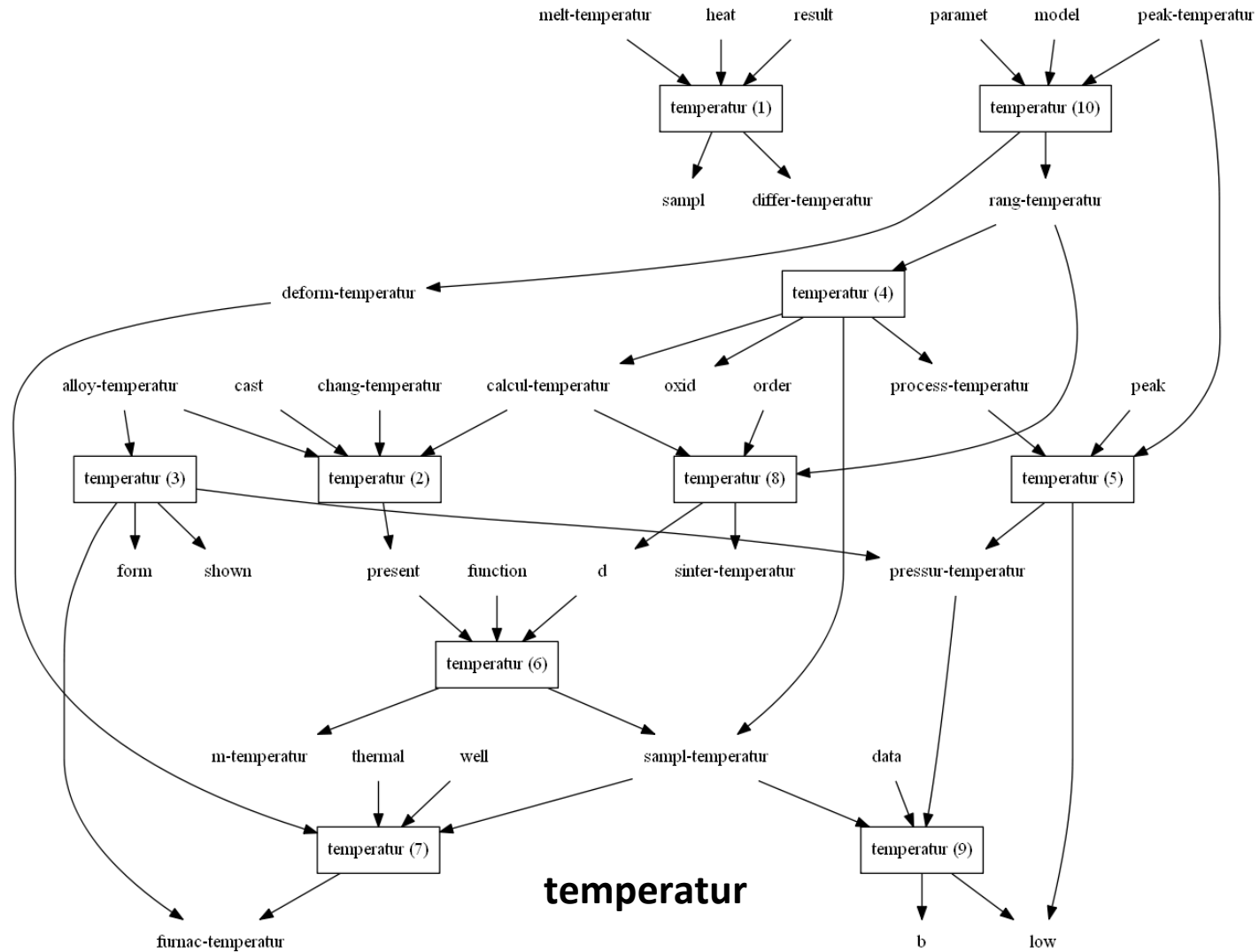


<http://liqianguo.wordpress.com/2011/06/09/latent-semantic-analysis/>

Cosine Distances Between Stems

	temperatur	diffusi	diffusion	equilib	equilibr	equilibria	equilibrium	entropi	enthalpi	energi	crystal	compos	composi	composit	thermo	thermodi	thermodynam	pressur	densit	densiti	liquid	solution	defect	gradient
temperatur	1.00	0.83	0.95	-0.04	0.99	-0.11	-0.18	0.01	0.01	-0.09	-0.06	0.18	0.44	-0.11	0.08	-0.06	0.02	0.06	-0.12	-0.10	-0.04	0.45	-0.12	0.50
diffusi	0.83	1.00	0.85	0.06	0.81	0.00	-0.14	0.20	0.19	0.37	-0.13	0.19	0.51	-0.11	0.18	-0.03	0.02	0.02	-0.16	-0.14	0.15	0.47	0.03	0.46
diffusion	0.95	0.85	1.00	0.26	0.96	0.18	-0.09	0.09	0.09	-0.03	-0.04	0.38	0.66	-0.14	0.33	0.16	0.21	0.19	-0.21	-0.16	-0.05	0.42	-0.02	0.45
equilib	-0.04	0.06	0.26	1.00	0.05	0.94	0.51	0.12	0.12	-0.10	-0.03	0.54	0.77	-0.01	0.88	0.85	0.71	0.47	-0.16	-0.11	0.07	0.02	0.18	0.09
equilibr	0.99	0.81	0.96	0.05	1.00	-0.04	-0.09	-0.03	-0.03	-0.11	-0.04	0.22	0.50	-0.11	0.15	0.02	0.09	0.11	-0.14	-0.13	-0.04	0.43	-0.11	0.54
equilibria	-0.11	0.00	0.18	0.94	-0.04	1.00	0.33	0.11	0.12	-0.07	0.01	0.53	0.72	0.04	0.93	0.91	0.81	0.18	-0.16	-0.09	0.04	0.07	0.11	-0.14
equilibrium	-0.18	-0.14	-0.09	0.51	-0.09	0.33	1.00	0.01	-0.02	-0.12	-0.04	-0.01	0.40	0.02	0.39	0.48	0.36	0.43	0.03	-0.11	0.08	-0.28	0.25	0.62
entropi	0.01	0.20	0.09	0.12	-0.03	0.11	0.01	1.00	1.00	-0.07	-0.06	0.06	0.39	-0.04	0.09	-0.08	-0.16	0.05	0.06	0.02	-0.07	-0.02	0.75	0.02
enthalpi	0.01	0.19	0.09	0.12	-0.03	0.12	-0.02	1.00	1.00	-0.10	-0.06	0.06	0.39	-0.04	0.09	-0.07	-0.15	0.00	0.05	0.02	-0.04	0.01	0.72	-0.02
energi	-0.09	0.37	-0.03	-0.10	-0.11	-0.07	-0.12	-0.07	-0.10	1.00	-0.04	0.03	-0.09	-0.05	0.03	-0.08	-0.04	-0.07	-0.06	-0.05	-0.02	-0.10	0.00	0.01
crystal	-0.06	-0.13	-0.04	-0.03	-0.04	0.01	-0.04	-0.06	-0.06	-0.04	1.00	0.71	0.15	-0.05	0.00	-0.05	-0.04	-0.06	-0.04	-0.03	-0.03	0.04	0.50	0.05
compos	0.18	0.19	0.38	0.54	0.22	0.53	-0.01	0.06	0.06	0.03	0.71	1.00	0.62	-0.05	0.49	0.33	0.25	0.33	-0.22	-0.11	-0.01	0.16	0.46	0.06
composi	0.44	0.51	0.66	0.77	0.50	0.72	0.40	0.39	0.39	-0.09	0.15	0.62	1.00	-0.01	0.79	0.64	0.55	0.29	-0.21	-0.21	0.01	0.19	0.43	0.41
composit	-0.11	-0.11	-0.14	-0.01	-0.11	0.04	0.02	-0.04	-0.04	-0.05	-0.05	-0.05	-0.01	1.00	0.18	0.15	-0.11	-0.05	0.00	-0.03	-0.07	-0.17	0.00	0.18
thermo	0.08	0.18	0.33	0.88	0.15	0.93	0.39	0.09	0.09	0.03	0.00	0.49	0.79	0.18	1.00	0.95	0.86	0.12	-0.07	-0.04	-0.13	0.05	0.14	0.10
thermodi	-0.06	-0.03	0.16	0.85	0.02	0.91	0.48	-0.08	-0.07	-0.08	-0.05	0.33	0.64	0.15	0.95	1.00	0.94	0.05	-0.02	-0.01	-0.11	0.00	-0.03	0.03
thermodynam	0.02	0.02	0.21	0.71	0.09	0.81	0.36	-0.16	-0.15	-0.04	-0.04	0.25	0.55	-0.11	0.86	0.94	1.00	-0.14	0.03	0.04	-0.15	0.09	-0.15	-0.08
pressur	0.06	0.02	0.19	0.47	0.11	0.18	0.43	0.05	0.00	-0.07	-0.06	0.33	0.29	-0.05	0.12	0.05	-0.14	1.00	-0.15	-0.12	-0.06	-0.27	0.26	0.42
densit	-0.12	-0.16	-0.21	-0.16	-0.14	-0.16	0.03	0.06	0.05	-0.06	-0.04	-0.22	-0.21	0.00	-0.07	-0.02	0.03	-0.15	1.00	0.97	0.00	0.40	0.15	0.03
densiti	-0.10	-0.14	-0.16	-0.11	-0.13	-0.09	-0.11	0.02	0.02	-0.05	-0.03	-0.11	-0.21	-0.03	-0.04	-0.01	0.04	-0.12	0.97	1.00	0.01	0.48	0.10	-0.10
liquid	-0.04	0.15	-0.05	0.07	-0.04	0.04	0.08	-0.07	-0.04	-0.02	-0.03	-0.01	0.01	-0.07	-0.13	-0.11	-0.15	-0.06	0.00	0.01	1.00	0.63	-0.16	0.02
solution	0.45	0.47	0.42	0.02	0.43	0.07	-0.28	-0.02	0.01	-0.10	0.04	0.16	0.19	-0.17	0.05	0.00	0.09	-0.27	0.40	0.48	0.63	1.00	-0.14	0.00
defect	-0.12	0.03	-0.02	0.18	-0.11	0.11	0.25	0.75	0.72	0.00	0.50	0.46	0.43	0.00	0.14	-0.03	-0.15	0.26	0.15	0.10	-0.16	-0.14	1.00	0.26
gradient	0.50	0.46	0.45	0.09	0.54	-0.14	0.62	0.02	-0.02	0.01	0.05	0.06	0.41	0.18	0.10	0.03	-0.08	0.42	0.03	-0.10	0.02	0.00	0.26	1.00

Topic Models



LSI-based Query

```
dima@h058084:~/nlp$ python findsimilarcompound.py findsimilar.json
Enter query: melt-temperatur heat result
0.82 of ambient air and melt temperature. The heating technology itself has
0.82 mixtures heated above the melting temperature in high purity Ar for
0.82 are heated up to melting temperatures by the laser and then
0.82 Robamat heating units. The melt temperature was kept constant at 720
0.82 White balance ) Aperture Melt temperature 1300'F 25 in.Hg Results and
0.81 considerable increase of the melting temperature. Figure 2. Heating chamber of
0.81 Table I. Results of melting temperature , enthalpy and crystallinity percentage
0.81 Heating the billet to a temperature above its melting point results
0.81 heating , resulting a homogeneous temperature distribution , both on the
0.81 resulted at the optimum melt temperature ( 80'C ) , die
0.78 00OR22725. leads to higher melting temperatures resulting in greater thermal efficiency
0.67 sensible heat at smelting operating temperatures. These results can only be
0.67 of the melt , the temperature and the resulting viscosity determine
0.66 metal by the empirical melting temperature rule.9 The niobium result is
0.66 II. The resulting bcc melting temperature is seen to be higher
0.66 heating at the anode operation temperature and the resulting volume expansion
0.60 technology provides lower peak flame temperatures and a more uniform heat
0.60 Total Ox en enrichment Melt temperature =- Avail2008 Unit Actual Design
0.60 heat to provide acceptable workplace temperatures and limit worker exposure to
0.59 is poured at a precise temperature into a heated mold of
0.59 scales to form low melting temperature silicates , leading to severe
0.59 into the melt and the temperature was held for 30 min.
0.59 Heat and heated from room temperature to completely molten with Mass
0.59 intensive shearing , the melt temperature is extremely uniform throughout the
0.59 during foaming. At high melt temperatures , complete loss of the
0.59 of BF slag particles Basicity Temperature Destiny Special heat Enthalpy Heat
0.59 equipment was calibrated on melting temperatures and enthalpies of pure elements
0.59 argon , and the melt temperature of 700 'C prior to
0.59 of heat vacancies concentration on temperature have been for the first
0.59 of heat vacancies concentration on temperature have been for the first
0.59 equal to bulk eutectic melting temperature 219.8'C [ 7 ] .
0.59 T , is the melting temperature and T is adiabatic temperature.
0.59 REFERENCES FOR TABLE 2 Melting temperatures are taken from the Handbook
0.59 or alloys with high melting temperature. In comparison to IGC ,
0.59 or alloys with high melting temperature. In comparison to IGC ,
0.59 P ) is the melting temperature at pressure P , regardless
0.59 heated at a relatively high temperature , e ( 'injection temperature
0.59 mixture was heated to predetermined temperature and it was held at
0.59 the results from the wire-bridge temperature validation experiments using pure Cu
0.59 mixture was heated to pre-set temperature , an aluminum plate (
0.59 profile , and the melting temperature of the pure metals. This
0.59 held far below their melting temperature , the and 13 )
0.59 beginning such as the melting temperature , pouring , filling ,
0.59 minutes after melting at a temperature of 1500'C , Then the
0.59 thin enough that the melting temperature throughout was elevated enough by
0.59 focused laser leads to high temperature and the reduction reaction resulting
0.59 . Heat TABLE 14. FABRICATION TEMPERATUR ES FOR SELS-D ALLOYst74 )
0.59 rmal Conductivity Melting Temperature Vaporization Temperature
0.59 CaO ( 9 ) Melting temperature
0.59 Sn-Pb eutectic alloy. The melting temperatures
Enter query: |
```

“melt-temperatur heat result”

Semantic Medline (SM)

1. What is it?

1. SM is a semantic layer on top of PubMed
2. SM supports semantic applications of semantic search, semantic visualization, and automatic summarization

SM

PubMed

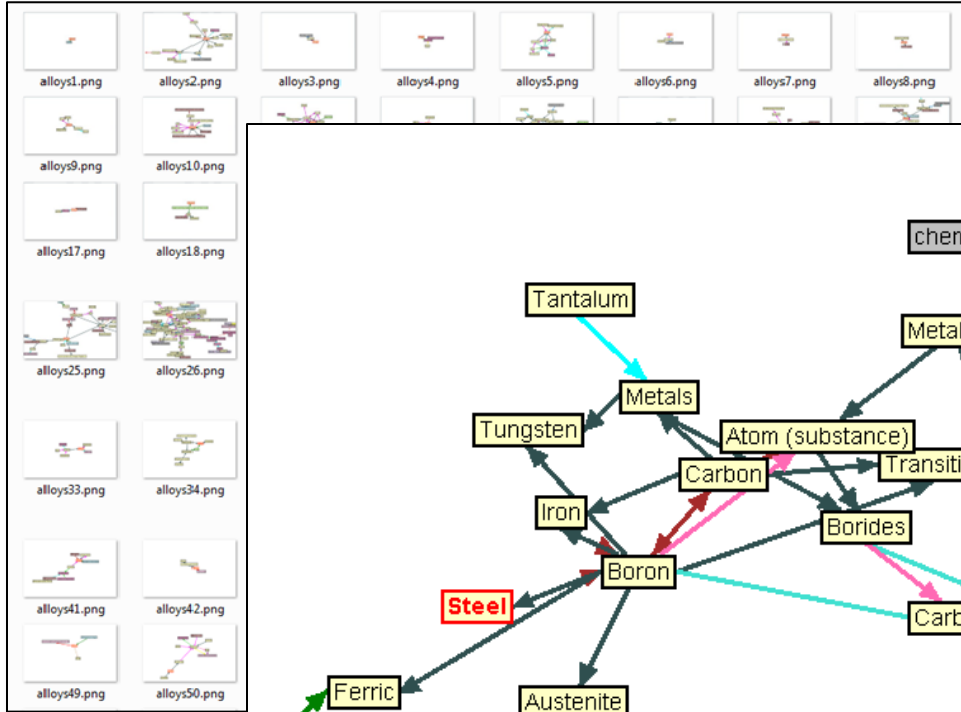
2. Collaborators: NIH/NLM

3. How does it work?

1. Supported by DB of corpus-based semantic information
2. Uses search to access semantic-info which supports semantic application operations

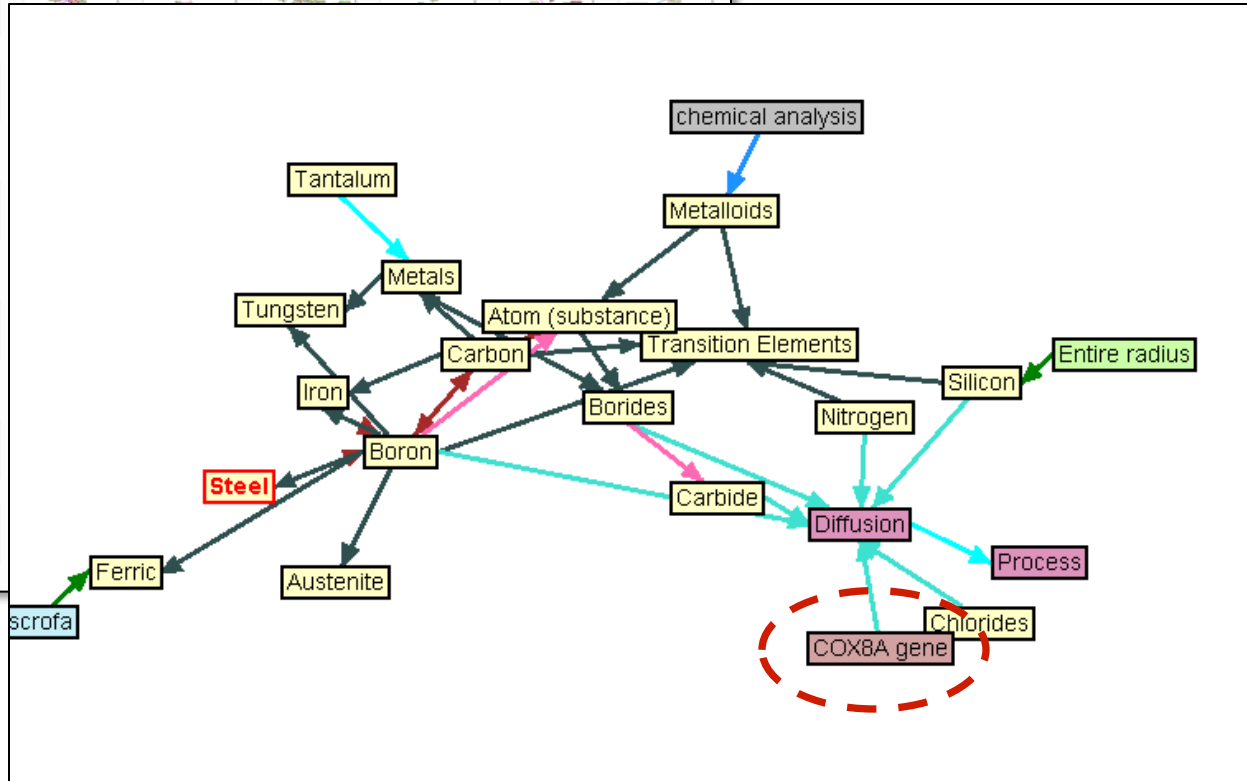
4. **GOAL:** *Retarget* SM from **biomedical** domain to **MS** domain.

Ab Initio Ontology – Semantic Medline

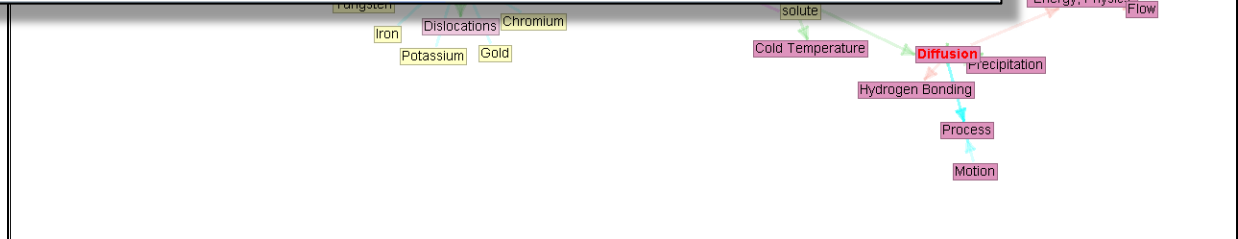


Iterative Process for Ontology Evolution

1. Start with Existing Ontology



plus
y Coverage



Ontology Acceleration - Recap

