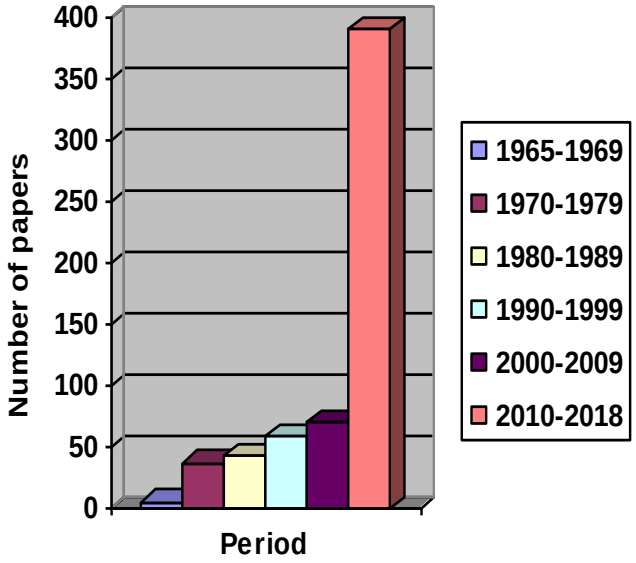
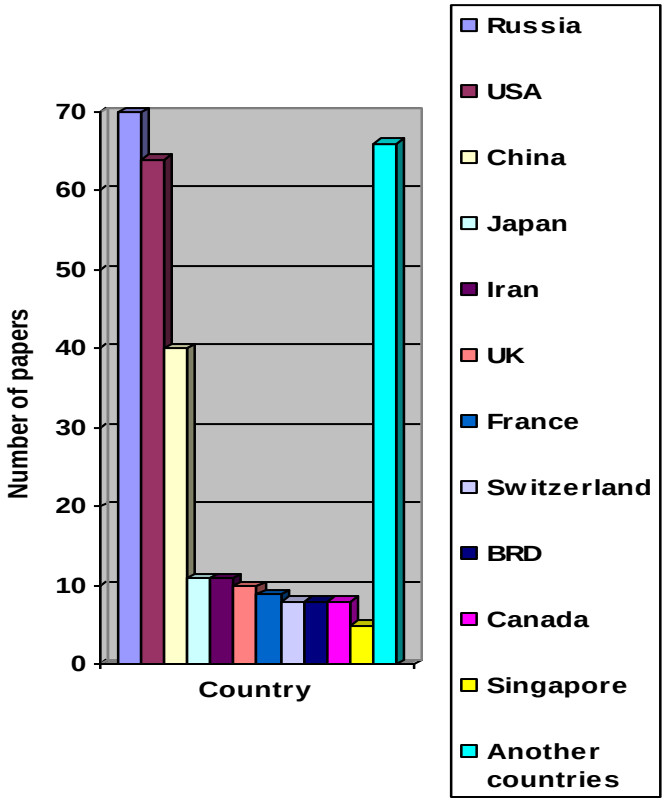
A.A.Baikov Institute of Metallurgy and Materials Science of Russian Academy of Sciences

N.N. Kiselyova

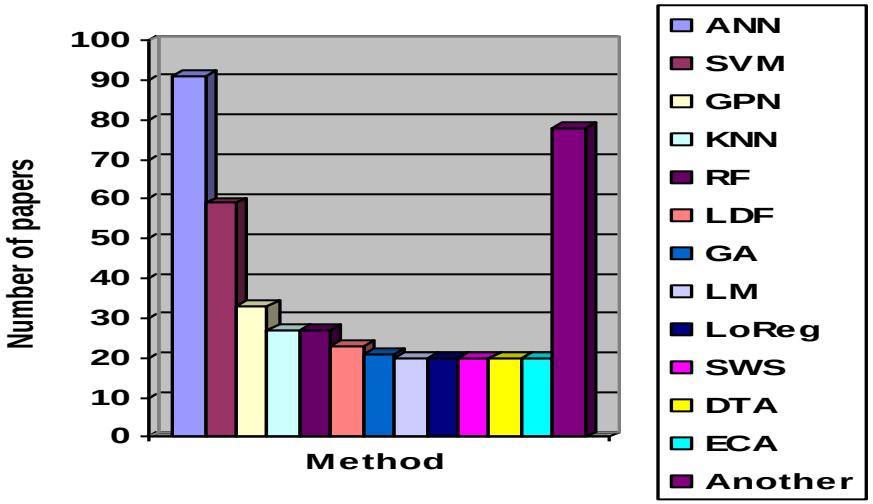
Machine Learning for Inorganic Compounds Design

Statistics of Papers with Application of Machine Learning Methods to Inorganic Materials Science



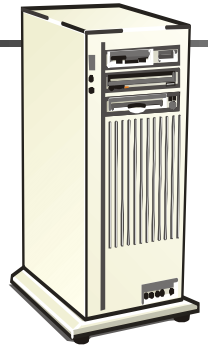


Statistics of Applications of Machine Learning Methods to Inorganic Materials Science



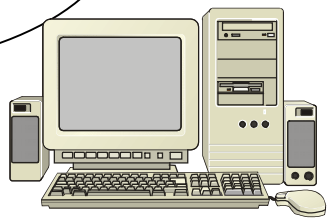
ANN – artificial neural network learning; **SVM** – support vector machine; **GPN** – concept formation using growing pyramidal networks; **KNN** – k-nearest neighbors; **RF** – random forests; **LDF** – linear Fisher discriminant; **GA** – genetic algorithm; **LM** – linear machine; **LoReg** - voting algorithm where estimations for classes are calculated with the help of voting by logical regularities system; **SWS** - statistical weighted syndromes; **DTA** - deadlock test algorithm; **ECA** - estimate calculating algorithm.

Search for classifying regularities using machine learning



Databases

Computer data analysis



Regularity

New compounds predictions

A / B	Ni	Pd	Pt
La	M	M	M
Ce	M	(M)	(M)
Pr	M	M	M
Nd	M	M	M

Some Solved Problems

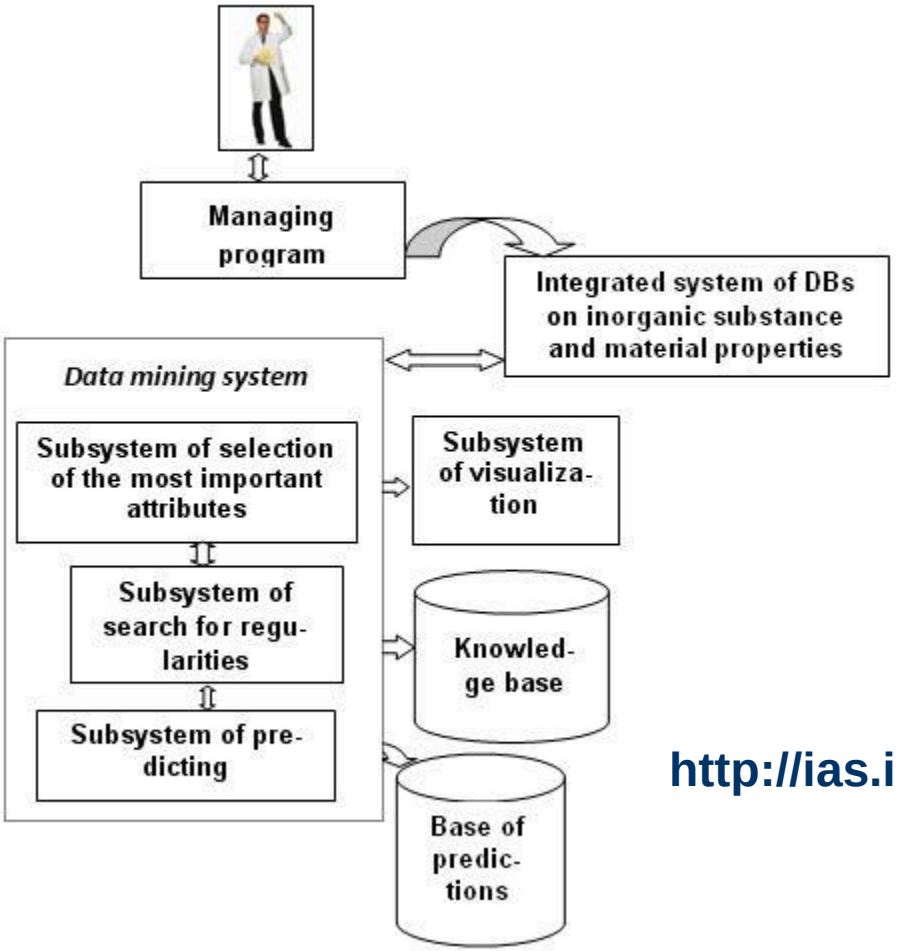
theoretic tasks of prediction of:

- inorganic system phase diagram type;
- formation of inorganic compounds of certain stoichiometric composition;
- crystal structure type of inorganic compounds;
- some of inorganic compounds properties (melting point, critical temperature of superconductivity, band gap energy, enthalpy of formation, lattice parameters, etc.);

technologic tasks of prediction of:

- mechanical properties of steels;
- acoustic properties of tellurite glasses;
- tribological behavior of aluminum–copper based composite;
- functional properties of ceramic materials,
- dielectric breakdown,
- thermoelectric properties,
- ionic conductivities, etc.

Structure of the Information-Analytical System of IMET RAS



http://ias.imet-db.ru

Procedure of Machine Learning

1) selection of the objects for machine learning,

2) formation of the attribute description (including selection of the most informative attributes and filling gaps of the attribute values also),

3) selection of the best machine learning algorithms,

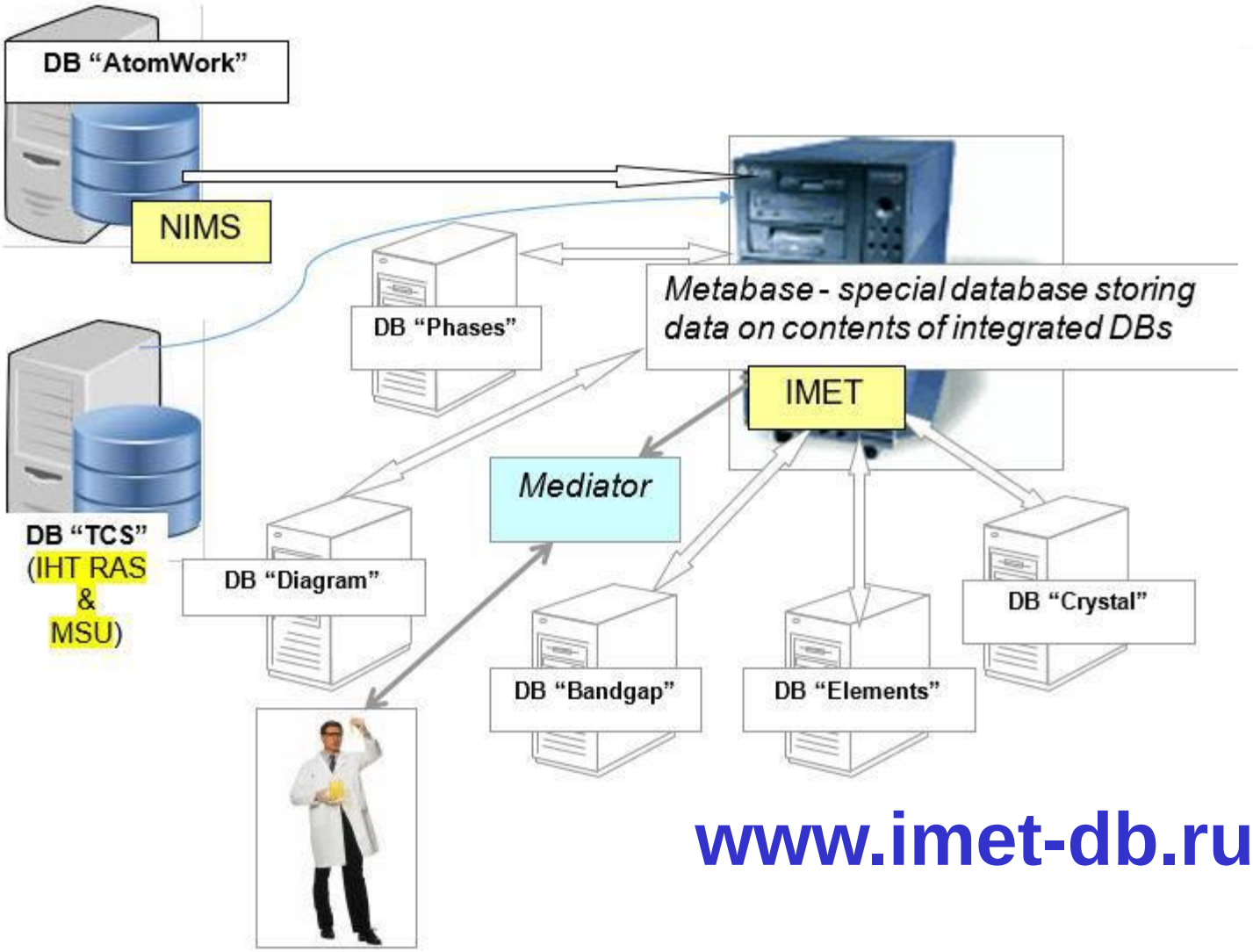
4) machine learning including application of ensembles of the algorithms and synthesis of a collective solution in the case of the several algorithms use,

5) estimation of the machine learning quality,

6) new objects prediction and the results interpretation.

Selection of the Objects for Machine Learning

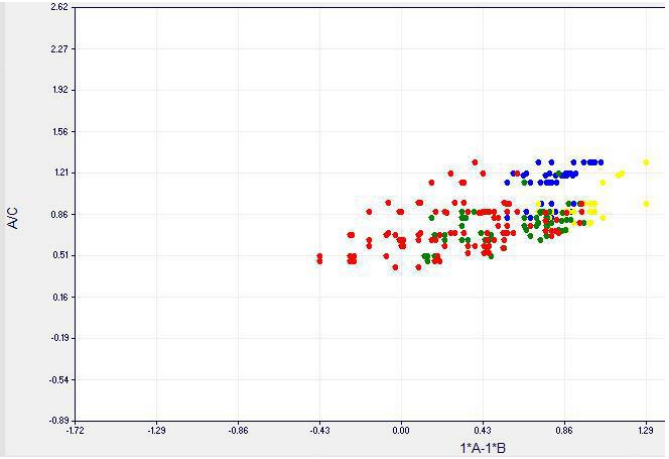
Structure of Integrated System of DBs



www.imet-db.ru

Formation of the Attribute Description

1. CCP (Convex Correcting Procedures) [*Senko O.V. An Optimal Ensemble of Predictors in Convex Correcting Procedures // Pattern Recognition and Image Analysis. 2009. V. 19. N. 3. P.465–468*];
2. GLMNET [*Friedman J., Hastie T., Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent // J. Statistical Software, 2009. V.33. N.1. P. 1-22*];
3. Improved GLMNET [*Yuan G.-X., Ho C.-H., Lin C.-J. An Improved GLMNET for L_1-regularized Logistic Regression // J. Machine Learning Research. 2012. V.13. P.1999-2030*].



Ось X	Ось Y
A - Radii ionic (Shannon)	A - Radii ionic (Shannon)
B - Radii ionic (Shannon)	B - Radii ionic (Shannon)
k1 - 1	C - Radii ionic (Shannon)
k2 - 1	k1 - 1
	k2 - 1
	k3 - 1

Характеристика	Кол-во	Цвет
nonformation	100	
another	55	
K2NiF4	46	
beta-K2SO4	31	

Task: Prediction of crystal structure type of AB_2Hal_4 compounds

Selection of Machine Learning Algorithms

- System RECOGNITION (A.A.Dorodnicyn Computer Center of Russian Academy of Sciences). It includes the methods of *Q*–nearest neighbors, Fisher's linear discriminant, linear machine, multi-level perceptron (neural network training), support vector machine, genetic algorithm, estimates calculation algorithms, LoReg (Logical Regularities), deadlock test algorithm, statistical weighted syndromes, etc.
- System of concept formation ConFor (V.M.Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine). It is based on special data structure in a computer memory named as growing pyramidal networks.

Machine Learning and Application of Ensembles of the Algorithms and Synthesis of a Collective Solution

- Bayes method,
- clustering and selection,
- decision templates,
- logical correction,
- the method of a convex stabilizer,
- the Woods dynamic method,
- committee methods, etc.

Estimation of Machine Learning Quality

- cross-validation (including LOOCV),
- ROC-analysis,
- traditional experimental method.

Average reliability – 80 %

Prediction of Formation of AB_3X_3 Compounds

A / B	Fe	Ga	In	Sn	Sb	La	Ce	Pr	Nd	Sm	Eu	Gd	Tb	Dy	Bi
X = S															
K	©	#2	1	1	#1	1		1	1	1	1	1	1	1	#2
Rb	©		#1	1	©	1			1	1	1	1	1	1	#1
Tl	1	Θ	#1	#1	#1	#2	2	#2	Θ	2	2				
X = Se															
K	#1	#1	1	©	#1		1			1	1	1	1	1	#1
Rb	1	1	1	1	©	1	1			1	1	1	1	1	#1
Ag	2	#2	Θ		#2	Θ	Θ	Θ	2	Θ	Θ		Θ	2	Θ
Cs	1	#1	1	1	©	1				1	1	1	1	1	#1
Tl	1	Θ	#2	#1	#1	2	2	2	2						#2
X = Te															
Rb	1	1	1	©	1	1	1			1	1	1	1	1	1
Ag	2	#2	#2	Θ	#2	2	2	2	2	2	2	#2	Θ	Θ	Θ
Cs	1	1	1	©	1	1	1			1	1	1	1	1	1
Tl	©	Θ	Θ	#1	#2	#2	2	2	Θ						#2

Designations: **1** - the prediction of formation of AB_3X_3 under normal conditions; **2** - the prediction of an absence of AB_3X_3 under normal conditions; # - examples, the information on which is used for machine learning; empty square - uncertain prediction; © - the prediction of formation of AB_3X_3 matches new experimental data; **Θ** - the prediction of the absence of a compound matches experimental data.

N.N. Kiselyova. Prediction of Formation of AB_3X_3 (X = S, Se, Te), Inorg. Mater., 2009, 45(10), p.1077-1080.

Prediction of $ThCr_2Si_2$ Crystal Structure Type for AB_2Si_2 Compounds

A \ B	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ru	Rh	Pd	Ag	Os	Ir	Pt	Au
Ca	+	⊕	↔	⊕	⊕	⊕	⊕	+	©	©	⊕	+	+	+	⊕
Sr		⊕		⊕	⊕	⊕	⊕	-	-		⊕			©	⊕
Y	⊕	⊕	⊕	⊕	⊕		-				∅	©	©	⊗	∅
Ba	+	⊕	+	Ω	⊕	⊕	⊕	+	⊗	+	⊕	+	⊗	⊗	⊕
La	+	⊕	⊕	⊕	⊕	⊕	⊕	⊕	©	©	⊕	©	©	⊕	⊕
Ce	↔	⊕	⊕	⊕	⊕	⊕	⊕	⊕	©	⊕	⊕	©	©	⊕	⊕
Pr	+	⊕	⊕	⊕	⊕	©	⊕	©	©	©	⊕	©	©	⊗	⊕
Nd	©	⊕	⊕	⊕	⊕	©	⊕	⊕	©	©	⊕	©	©	Ω	⊕
Pm	+	+	+	+	⊕	+	+	+	+	+	+	+	+	+	+
Sm	©	⊕	⊕	⊕	⊕	©	+	⊕	©	©	⊕	©	©	⊗	⊕
Eu	+	+	⊕	⊕	⊕	⊕	©	©	©	©	⊕	+	©	Ω	⊕
Gd	©	⊕	⊕	⊕	⊕	©	+	©	©	⊕	⊕	©	⊕	Ω	⊕
Tb	©	⊕	⊕	⊕	⊕	©	+	⊕	©	©	©	©	©	⊗	⊕
Dy	©	⊕	⊕	⊕	⊕	©	+	⊕	©	©	©	©	©	Ω	⊕
Ho	©	⊕	⊕	⊕	⊕	©	+	©	©	©	+	©	+	⊗	⊕
Er	©	⊕	⊕	⊕	⊕	©	+	⊕	©	©	+	©	©	Ω	⊕
Tm	©	©	⊕	⊕	⊕	©	+	©	©	©	©	©	+	Ω	+
Yb	©	⊕	⊕	⊕	⊕	©	+	⊕	⊕	⊕	⊕	©	©	Ω	⊕
Lu	©	©	⊕	⊕	⊕	©	+	©	©	©	+	©	+	Ω	+
Th	⊕	⊕	⊕	⊕	⊕	⊕		©	©	©	+	©	©	©	©
U	©	©		©	⊕	©		©	©	©	+	©	⊕	©	©
Np	©	©	⊕	⊕	⊕			©	©	©	+	©	⊗	⊗	⊗
Pu	©	©		©	©	©		©	©	©	+	©	⊗	⊗	©

Prediction error = 12 %

Designations: **+** - formation of a compound with the crystal structure type $ThCr_2Si_2$ is predicted; **-** - formation of a compound with the type $ThCr_2Si_2$ is not predicted; ⊕ - a compound with the type $ThCr_2Si_2$ was synthesized and appropriate information was used in the machine learning process; ↔ - a $ThCr_2Si_2$-compound does not exist under normal conditions and this fact was used in the machine learning process; © - predicted formation of a compound with the type $ThCr_2Si_2$ confirmed by experiment; ⊗ - predicted formation of a $ThCr_2Si_2$-compound is not confirmed by experiment; ∅ - predicted absence of a $ThCr_2Si_2$-compound is not confirmed by experiment; **Ω** - erroneous data in learning set; empty square - indeterminate result.

Киселева Н.Н., Савицкий Е.М. Прогнозирование образования тройных силицидов со структурой типа $ThCr_2Si_2$ // Изв.АН СССР. Неорган.материалы. 1983. Т.19. №3. С.489-491.

Prediction of Melilite Lattice Parameters

N.N. Kiselyova et al. Computer-assisted design of compounds with the melilite crystal structure type // Materialovedenie (in press)

Состав	a, A	c, A	Состав	a, A	c, A
$Ca_2MgGe_2O_7$	7.959	5.147	$Eu_2MnGe_2O_7$	8.310	5.335
$Ca_2MnSi_2O_7$	7.959	5.015	$Pb_2MnGe_2O_7$	8.472	5.140
$Ca_2FeSi_2O_7$	7.929	4.993	$Ca_2FeGe_2O_7$	8.040	5.135
$Ca_2CdSi_2O_7$	8.087	5.034	$Eu_2FeGe_2O_7$	8.281	5.313
$Pb_2MnSi_2O_7$	8.319	4.985	$Pb_2FeGe_2O_7$	8.443	5.119
$Ba_2FeSi_2O_7$	8.288	5.333	$Ca_2CoGe_2O_7$	7.947	5.161
$Pb_2FeSi_2O_7$	8.290	4.963	$Eu_2CoGe_2O_7$	8.188	5.339
$Pb_2ZnSi_2O_7$	8.178	4.995	$Pb_2CoGe_2O_7$	8.350	5.144
$Sr_2BeSi_2O_7$	7.596	5.138	$Ca_2NiGe_2O_7$	7.950	5.139
$Ba_2BeSi_2O_7$	7.777	5.328	$Sr_2NiGe_2O_7$	8.164	5.300
$Eu_2BeSi_2O_7$	7.592	5.145	$Ba_2NiGe_2O_7$	8.383	5.502
$Eu_2FeSi_2O_7$	8.102	5.150	$Eu_2NiGe_2O_7$	8.191	5.317
$Eu_2CoSi_2O_7$	8.010	5.176	$Pb_2NiGe_2O_7$	8.353	5.122
$Pb_2CoSi_2O_7$	8.197	4.989	$Ca_2CuGe_2O_7$	8.028	5.072
$Sr_2NiSi_2O_7$	8.016	5.146	$Sr_2CuGe_2O_7$	8.242	5.234
$Eu_2NiSi_2O_7$	8.012	5.154	$Eu_2CuGe_2O_7$	8.269	5.250
$Ca_2CuSi_2O_7$	7.918	4.930	$Pb_2CuGe_2O_7$	8.431	5.056
$Eu_2CuSi_2O_7$	8.091	5.087	$Eu_2ZnGe_2O_7$	8.169	5.345
$Pb_2CuSi_2O_7$	8.278	4.900	$Pb_2ZnGe_2O_7$	8.331	5.151
$Eu_2ZnSi_2O_7$	7.990	5.182	$Ca_2CdGe_2O_7$	8.197	5.176
$Eu_2CdSi_2O_7$	8.260	5.191	$Sr_2CdGe_2O_7$	8.411	5.337
$Eu_2MgGe_2O_7$	8.200	5.326	$Eu_2CdGe_2O_7$	8.438	5.354
$Ca_2MnGe_2O_7$	8.069	5.157	$Pb_2CdGe_2O_7$	8.600	5.160

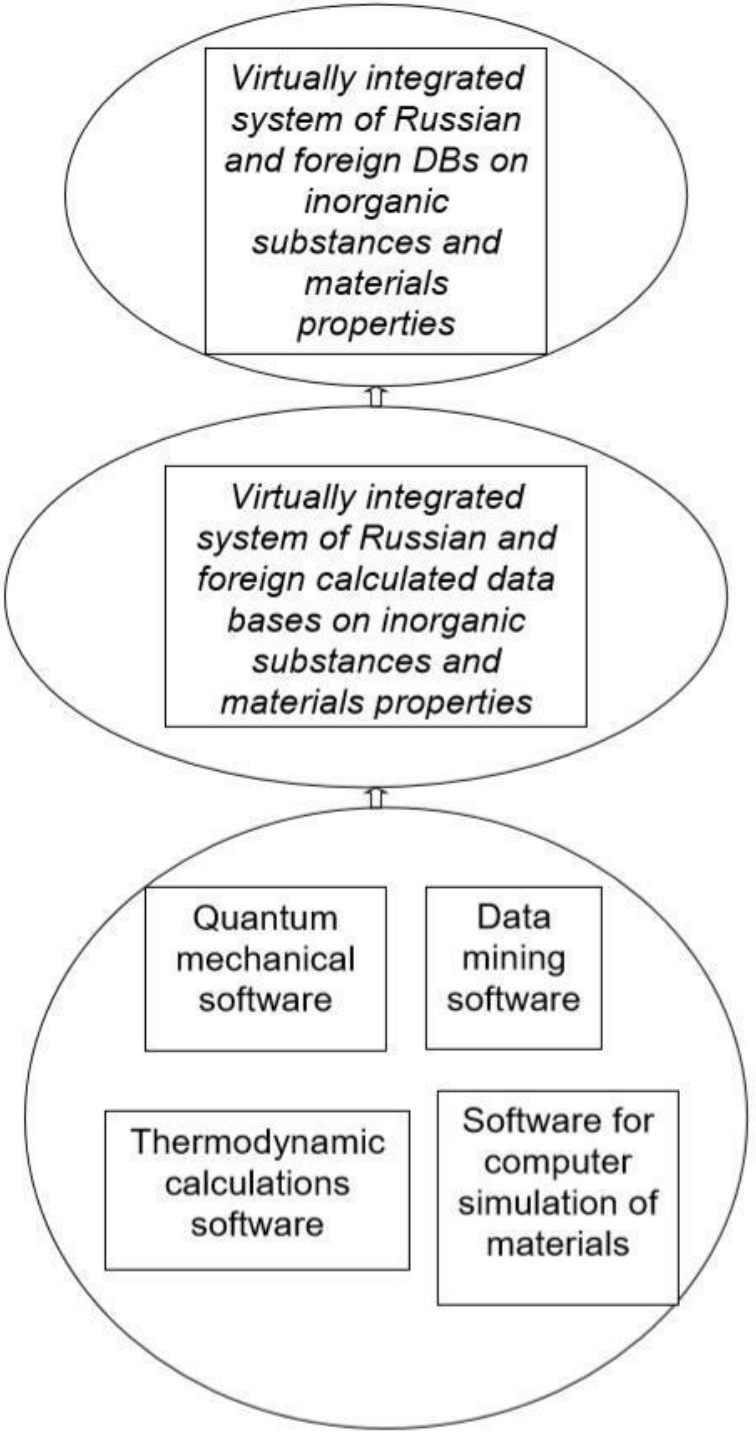
Prediction of Bandgap of New Chalcopyrites

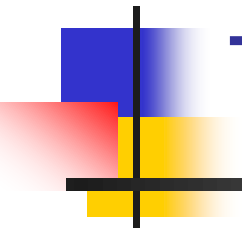
Compound	Prediction of E_g
$ZnAlS_2$	>2 eV
$ZnAlSe_2$	>2 eV
$ZnAlTe_2$	<2 eV
$AgFeS_2$	<2 eV (1.1 eV)
$AgFeTe_2$	<2 eV
$ZnGaTe_2$	<2 eV
$CdGaTe_2$	<2 eV
$HgGaTe_2$	<2 eV

N.N. Kiselyova, *et.al*. Proc. 6th Int.Conf. on Computer-Aided Design of Discrete Devices (CAD DD'07). Vol.1. 2007

Project of In-formation Infrastructure for Inorganic Materials Science

Materials data bankers of the world, unite!





Thank you!

Questions?

http://imet-db.ru/