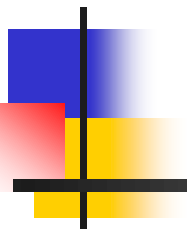


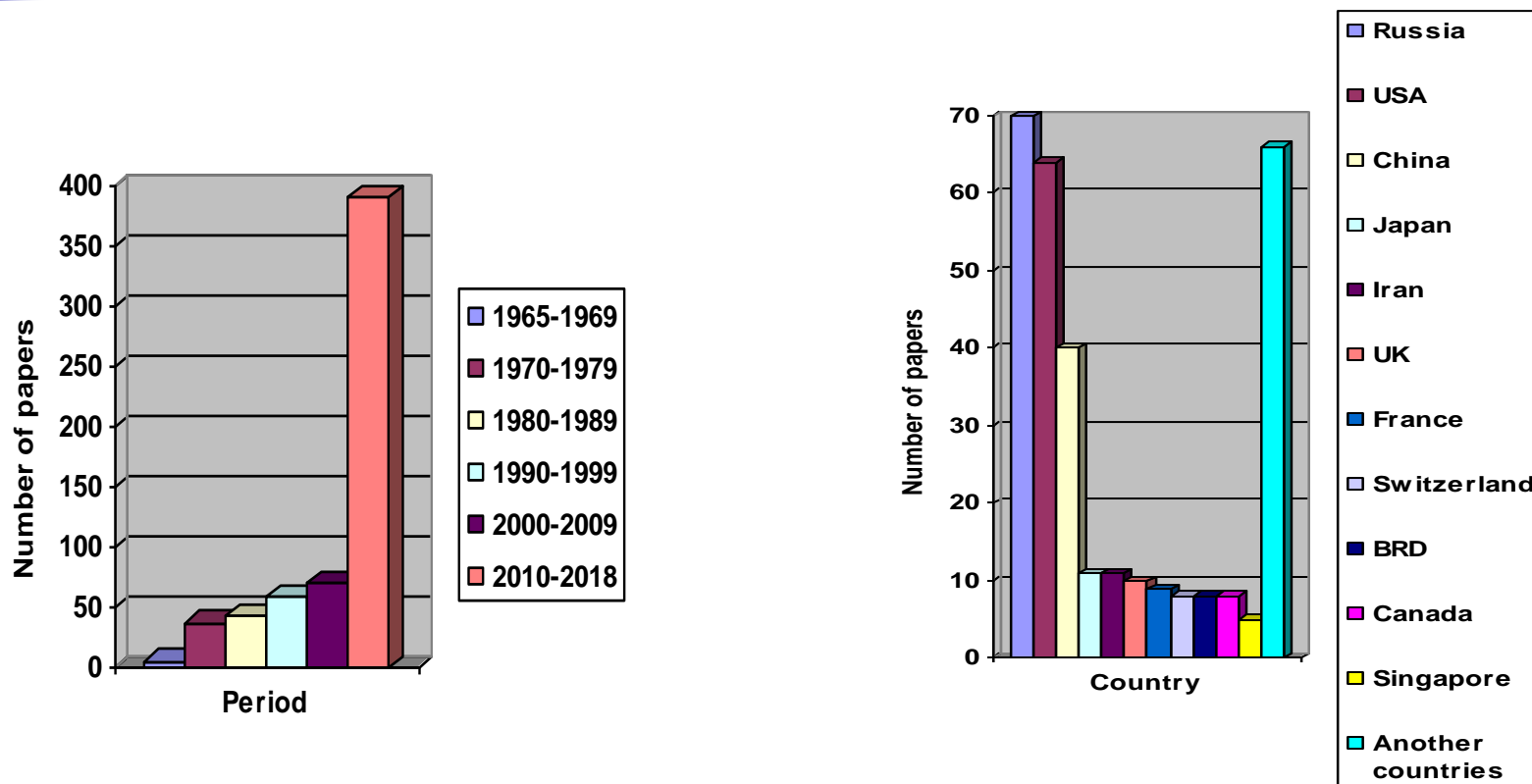


***A.A.Baikov Institute of Metallurgy and
Materials Science of Russian Academy
of Sciences***
N.N. Kiselyova

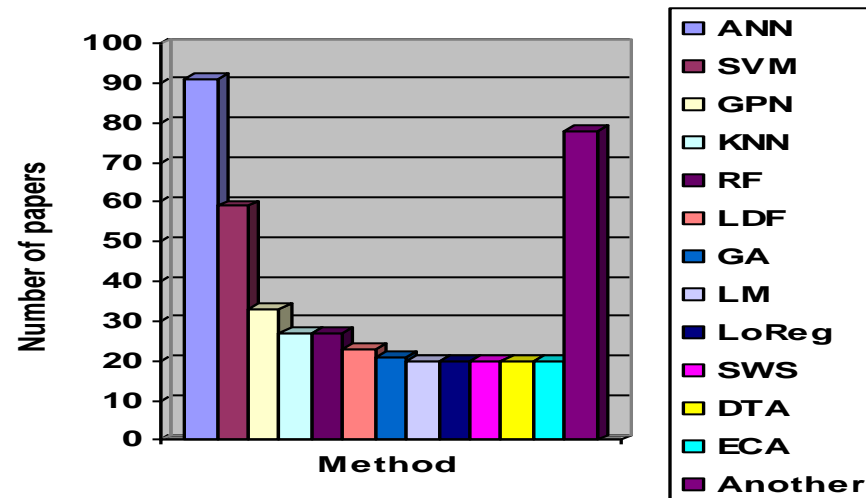


Machine Learning for Inorganic Compounds Design

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

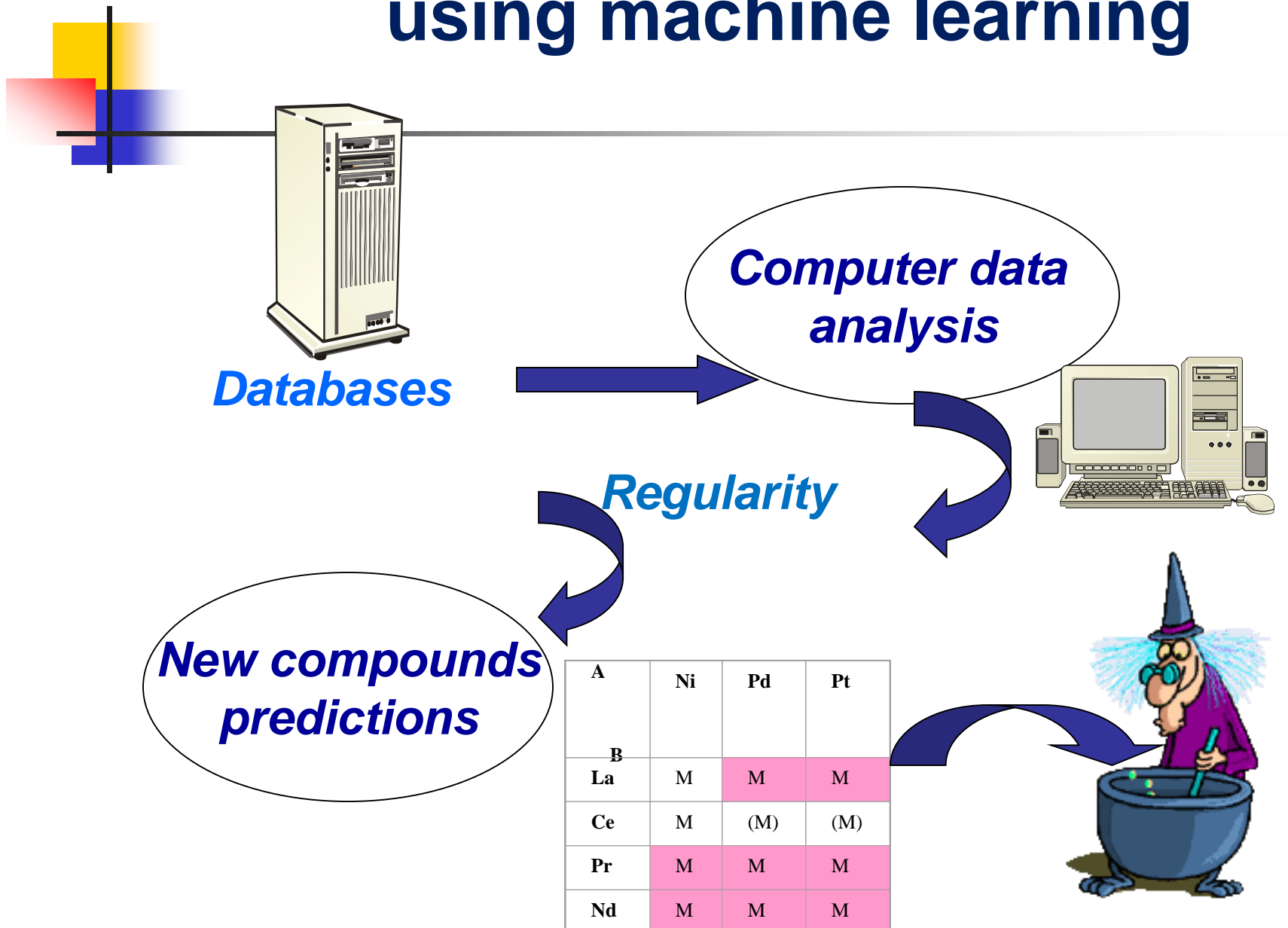


Statistics of Applications of Machine Learning Methods to Inorganic Materials Science



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





Some Solved Problems

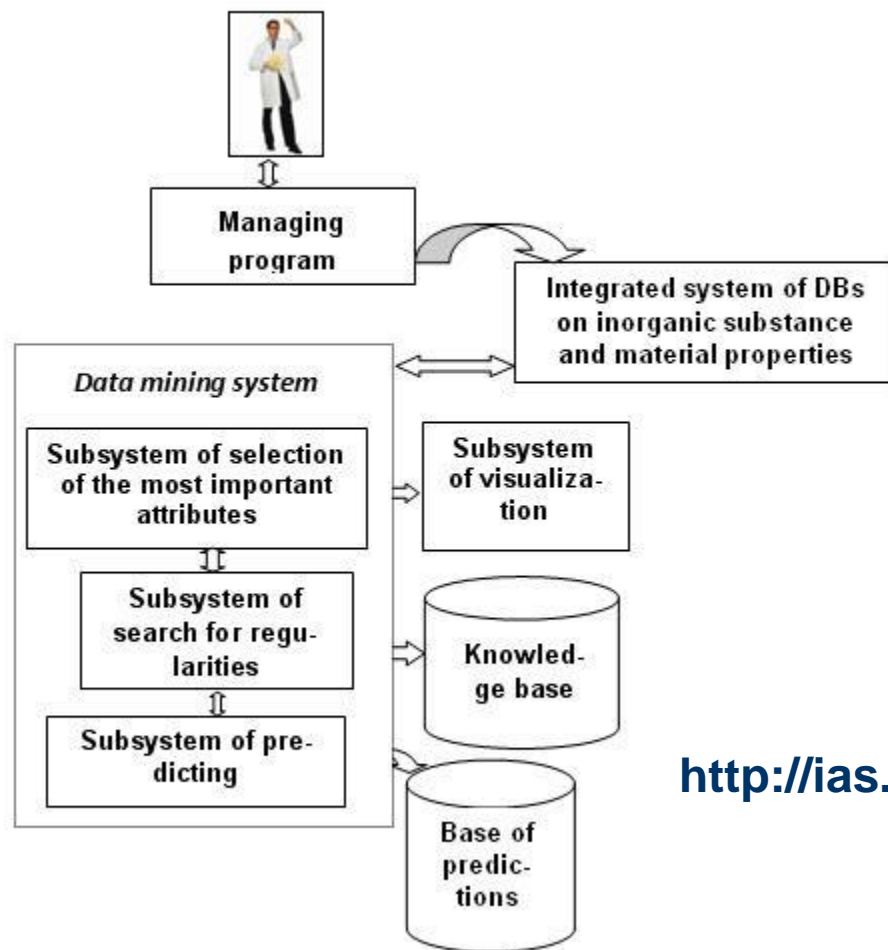
theoretic tasks of prediction of:

- 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>

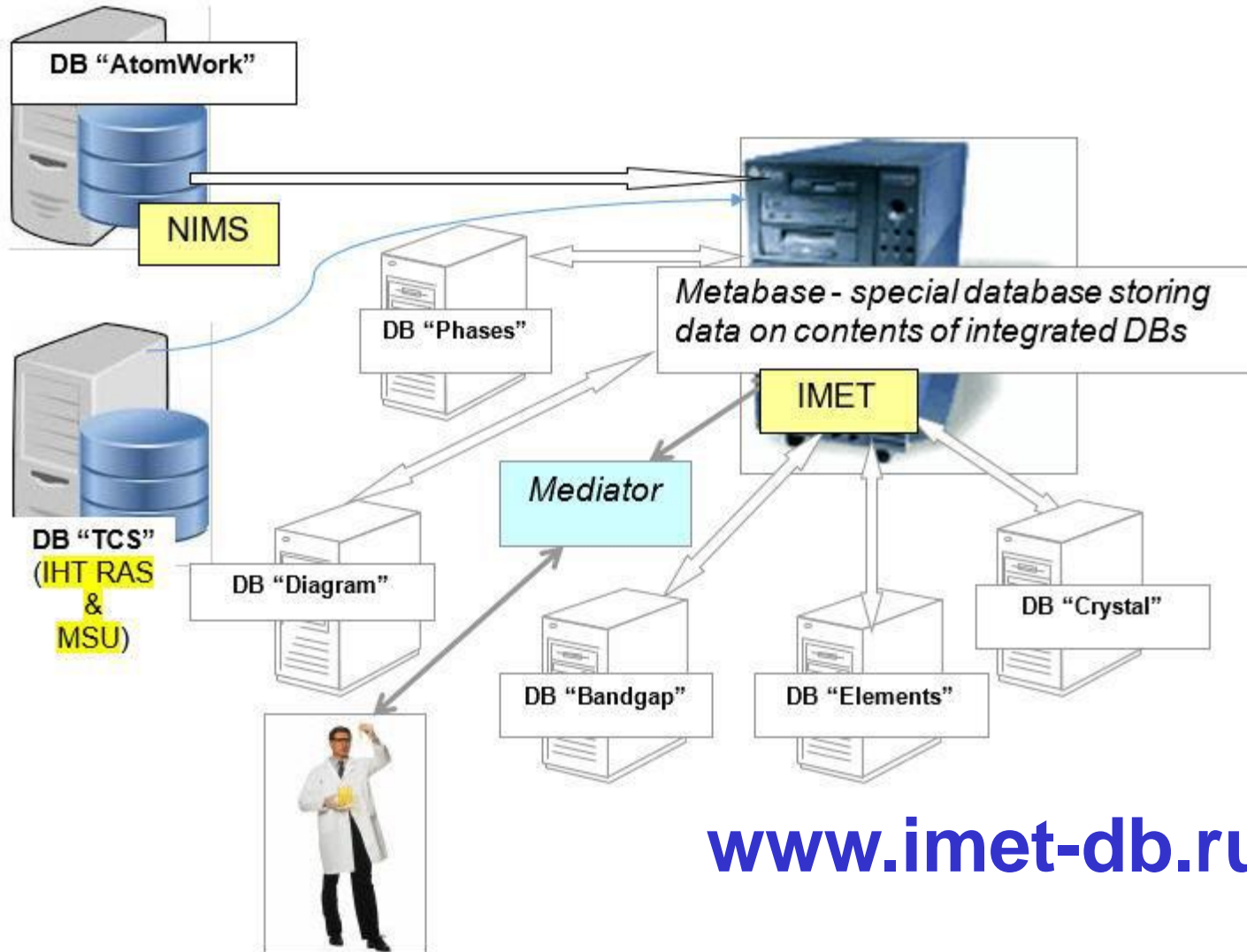
Procedure of Machine Learning



- 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.

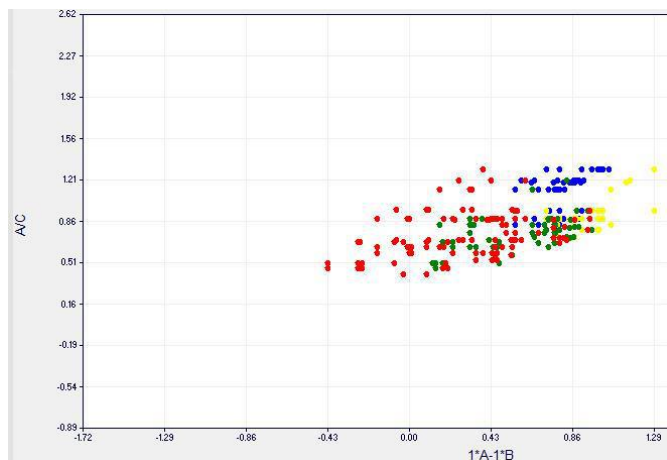
Selection of the Objects for Machine Learning

Structure of Integrated System of DBs



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 - Радиi ionic (Shannon)	A - Радиi ionic (Shannon)
B - Радиi ionic (Shannon)	B - Радиi ionic (Shannon)
k1 - 1	C - Радиi ionic (Shannon)
k2 - 1	k1 - 1
k3 - 1	k2 - 1
	k3 - 1

Характеристика	Кол-во	Цвет
nonformation	100	Red
another	55	Green
K2NiF4	46	Blue
beta-K2SO4	31	Yellow

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	Fe	Ga	In	Sn	Sb	La	Ce	Pr	Nd	Sm	Eu	Gd	Tb	Dy	Bi
B																
X = S																
K	⊙	#2	1	1	#1	1		1	1	1	1	1	1	1	1	#2
Rb	⊙		#1	1	⊙	1			1	1	1	1	1	1	1	#1
Tl	1	⊖	#1	#1	#1	#2	2	#2	⊖	2	2					
X = Se																
K	#1	#1	1	⊙	#1		1			1	1	1	1	1	1	#1
Rb	1	1	1	1	⊙	1	1			1	1	1	1	1	1	#1
Ag	2	#2	⊖		#2	⊖	⊖	⊖	2	⊖	⊖		⊖	2	⊖	
Cs	1	#1	1	1	⊙	1				1	1	1	1	1	1	#1
Tl	1	⊖	#2	#1	#1	2	2	2	2							#2
X = Te																
Rb	1	1	1	⊙	1	1	1			1	1	1	1	1	1	1
Ag	2	#2	#2	⊖	#2	2	2	2	2	2	2	#2	⊖	⊖	⊖	
Cs	1	1	1	⊙	1	1	1			1	1	1	1	1	1	1
Tl	⊙	⊖	⊖	#1	#2	#2	2	2	⊖							#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; ⊖ - 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

Prediction error = 12 %

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	⊙	⊙		⊙	⊙	⊙			⊙	⊙	⊙	+	⊙	⊗	⊗	⊙

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)

CoctAB	a, A	c, A	CoctAB	a, A	c, A
Ca₂MgGe₂O₇	7.959	5.147	Eu₂MnGe₂O₇	8.310	5.335
Ca₂MnSi₂O₇	7.959	5.015	Pb₂MnGe₂O₇	8.472	5.140
Ca₂FeSi₂O₇	7.929	4.993	Ca₂FeGe₂O₇	8.040	5.135
Ca₂CdSi₂O₇	8.087	5.034	Eu₂FeGe₂O₇	8.281	5.313
Pb₂MnSi₂O₇	8.319	4.985	Pb₂FeGe₂O₇	8.443	5.119
Ba₂FeSi₂O₇	8.288	5.333	Ca₂CoGe₂O₇	7.947	5.161
Pb₂FeSi₂O₇	8.290	4.963	Eu₂CoGe₂O₇	8.188	5.339
Pb₂ZnSi₂O₇	8.178	4.995	Pb₂CoGe₂O₇	8.350	5.144
Sr₂BeSi₂O₇	7.596	5.138	Ca₂NiGe₂O₇	7.950	5.139
Ba₂BeSi₂O₇	7.777	5.328	Sr₂NiGe₂O₇	8.164	5.300
Eu₂BeSi₂O₇	7.592	5.145	Ba₂NiGe₂O₇	8.383	5.502
Eu₂FeSi₂O₇	8.102	5.150	Eu₂NiGe₂O₇	8.191	5.317
Eu₂CoSi₂O₇	8.010	5.176	Pb₂NiGe₂O₇	8.353	5.122
Pb₂CoSi₂O₇	8.197	4.989	Ca₂CuGe₂O₇	8.028	5.072
Sr₂NiSi₂O₇	8.016	5.146	Sr₂CuGe₂O₇	8.242	5.234
Eu₂NiSi₂O₇	8.012	5.154	Eu₂CuGe₂O₇	8.269	5.250
Ca₂CuSi₂O₇	7.918	4.930	Pb₂CuGe₂O₇	8.431	5.056
Eu₂CuSi₂O₇	8.091	5.087	Eu₂ZnGe₂O₇	8.169	5.345
Pb₂CuSi₂O₇	8.278	4.900	Pb₂ZnGe₂O₇	8.331	5.151
Eu₂ZnSi₂O₇	7.990	5.182	Ca₂CdGe₂O₇	8.197	5.176
Eu₂CdSi₂O₇	8.260	5.191	Sr₂CdGe₂O₇	8.411	5.337
Eu₂MgGe₂O₇	8.200	5.326	Eu₂CdGe₂O₇	8.438	5.354
Ca₂MnGe₂O₇	8.069	5.157	Pb₂CdGe₂O₇	8.600	5.160

Prediction of Bandgap of New Chalcopyrites

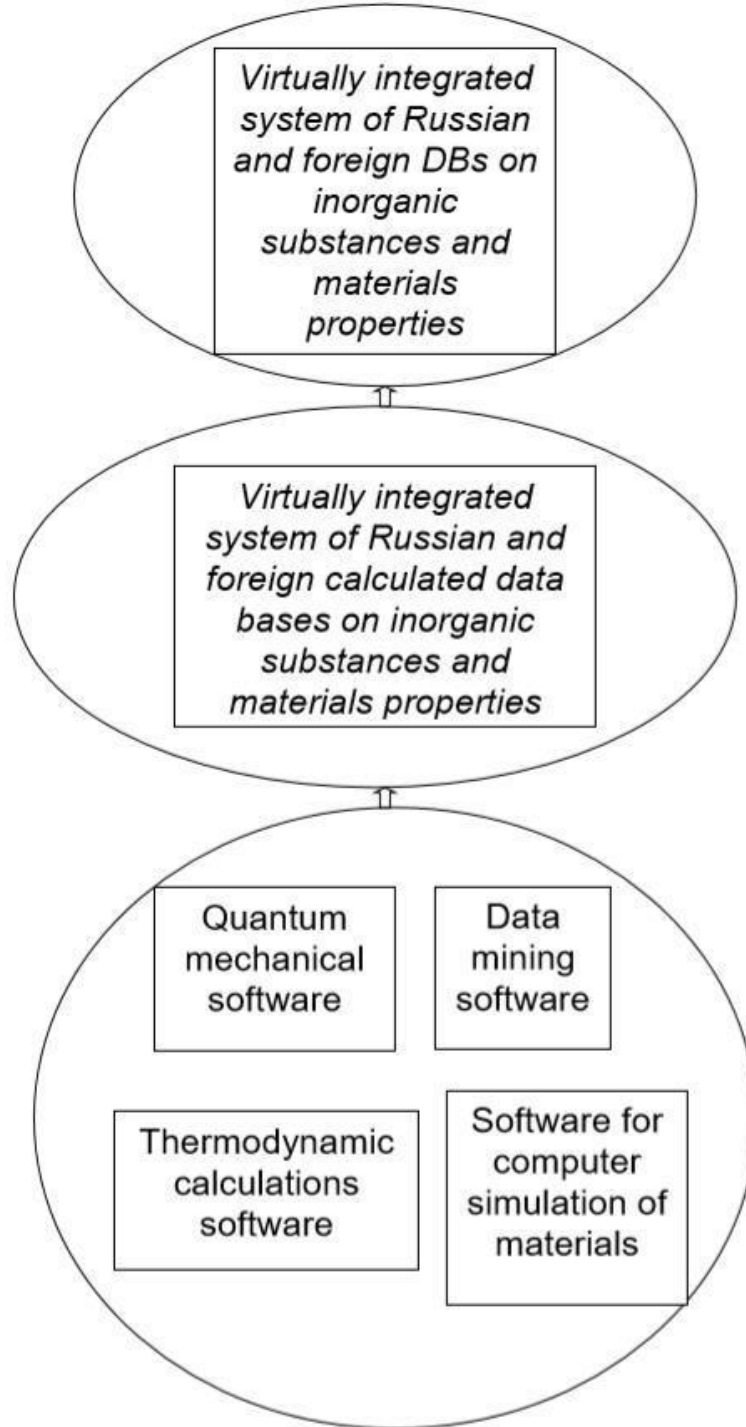
Compound	Prediction of E_g
ZnAlS ₂	>2 eV
ZnAlSe ₂	>2 eV
ZnAlTe ₂	<2 eV
AgFeS ₂	<2 eV (1.1 eV)
AgFeTe ₂	<2 eV
ZnGaTe ₂	<2 eV
CdGaTe ₂	<2 eV
HgGaTe ₂	<2 eV

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



Project of Information Infrastructure for Inorganic Materials Science

Materials data bankers of the world, unite!





Thank you!

Questions?

<http://imet-db.ru/>