ontochem
IT SOLUTIONS

OC|processor

Understanding linguistic Markush expressions in chemical patents

OntoChem IT Solutions GmbH
Blücherstr. 24
06120 Halle (Saale)
Germany
Tel. +49 345 4780472
Fax: +49 345 4780471
mail: info@ontochem.com

ontochem
IT SOLUTIONS

OntoChem is creating customized semantic **chemistry search engines:**

ocminer

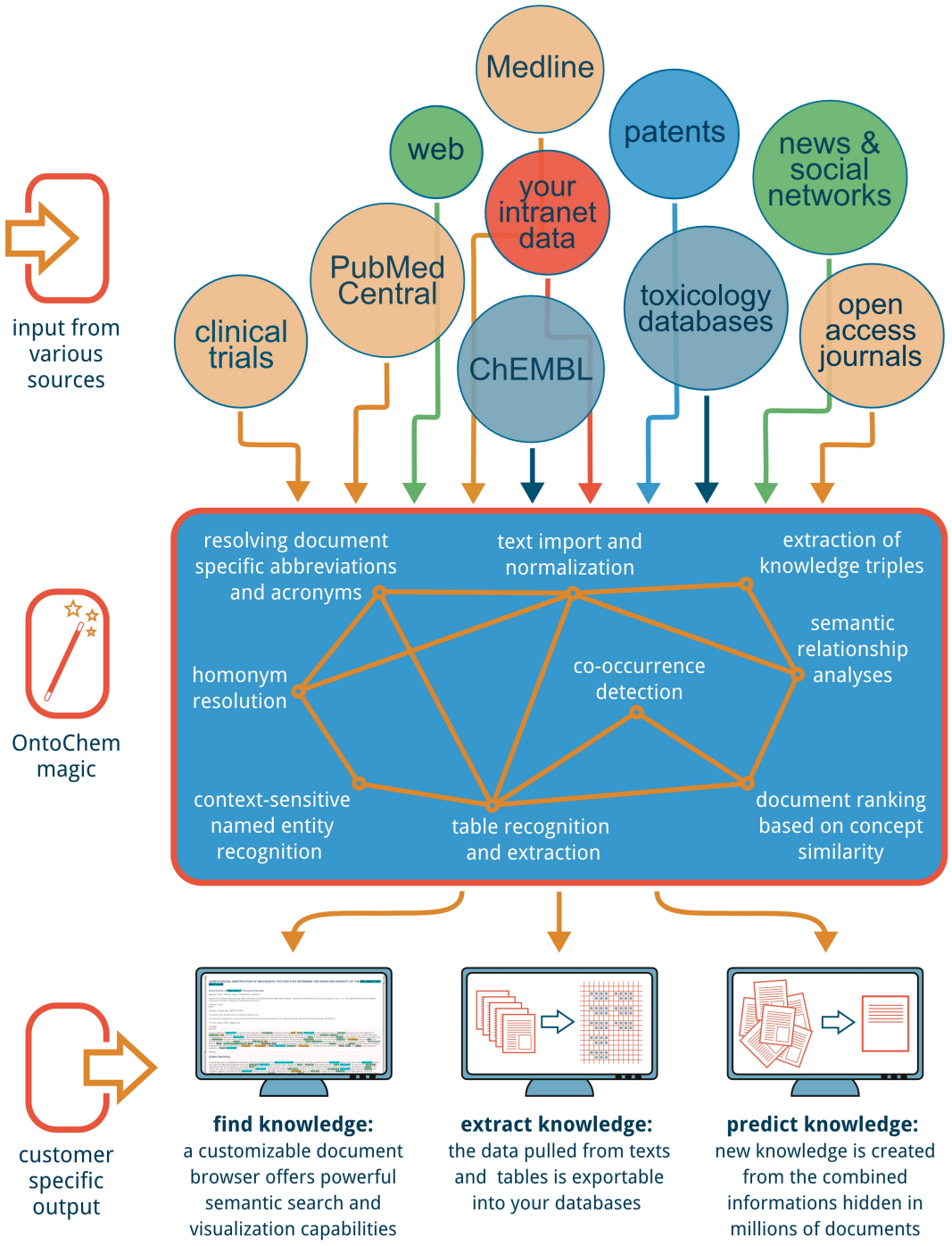
http://demo.ocminer.com

chemanalyser
close your information gap

https://www.chemanalyser.com

SciWalker
Find. Extract. Predict.

http://sciwalker.com/



Agenda

ontochem
IT SOLUTIONS

Understanding linguistic Markush expressions in chemical patents

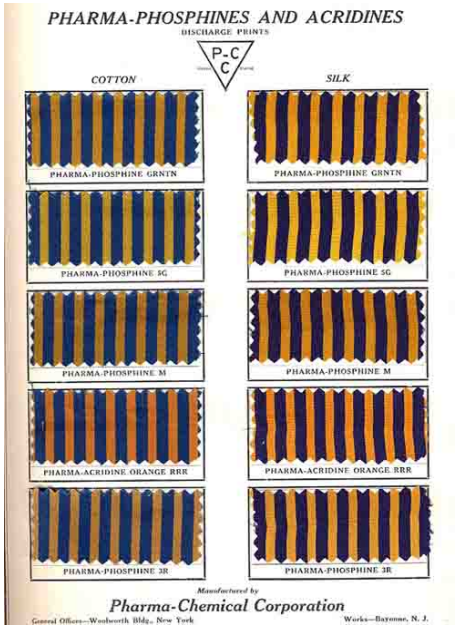
- **Markush:** foundation & future
- **Markush:** phenotypes
- Semantic extraction of Markush data **from text**
- Semantic extraction of Markush data **from tables**

Markush: Foundation

ontochem
IT SOLUTIONS

Eugene A. Markush

born in Budapest, Hungary, 1888, migrated to USA 1913,
founded Pharma-Chemical Corporation (NJ), 1919 sold to Bayer, died in 1968



US 1,506,316, filed January 9, 1923
"Pyrazolone Dye and Process of Making the Same"
after a ruling in 1925 by the US Patent Office

Claims: 95

1. The process for the manufacture of dyes which comprises coupling with a halogen-substituted pyrazolone, a diazotized unsulphonated material selected from the group consisting of aniline, homologues of 100 aniline and halogen substitution products of aniline.

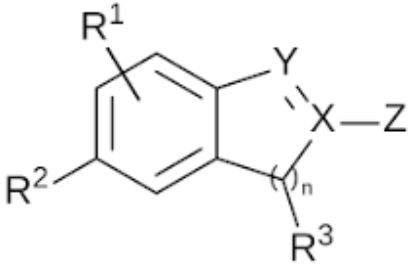
Markush: Phenotypes

ontochem IT SOLUTIONS

A "Markush" claim recites a list of **alternatively useable species**. …

A Markush claim is commonly formatted as: "*selected from the group consisting of A, B, and C;*" however, the phrase "Markush claim" means any claim that recites a list of alternatively useable species **regardless of format**.

https://www.uspto.gov/web/offices/pac/mpep/s2173.html



Embodiments of the present compound include compounds of the formula (1) shown in Table 1, compounds of the formula (2) shown in Table 2, compounds of the formula (2A) shown in Table 3, compounds of the formula (2B) shown in Table 4 and compounds of the formula (2C) shown in Table 5. These compounds are the compounds disclosed in International Publication WO 2013/018928 and can be produced by the methods described in this publication.

TABLE 1

The present compound	R^1	R^2	R^3	R^4	R^5	R^6	A^1	A^2	A^3	n
1	Et	H	H	H	CF_3	H	NMe	CH	N	0
2	Et	H	H	H	CF_3	H	NMe	CH	N	1
3	Et	H	H	H	CF_3	H	NMe	CH	N	2
4	Et	H	CF_3	H	CF_3	H	NMe	CH	N	0
5	Et	H	CF_3	H	CF_3	H	NMe	CH	N	2
6	Et	H	H	H	CF_2CF_3	H	NMe	CH	N	0
7	Et	H	H	H	CF_2CF_3	H	NMe	CH	N	1
8	Et	H	H	H	CF_2CF_3	H	NMe	CH	N	2

Text
a compound of formula I, present in the amount of 10-20 wt%

Text
a compound selected from a group of 2-chloro-, 2-bromo-, 2-iodo-pyridine

Text
a compound of formula I, where R^1 is selected from alkyl, cycloalkyl, aryl or heteroaryl

Markush: Phenotypes

ontochem
IT SOLUTIONS

Julius Rebek & Stuart Kauffman

PCT/US1994/004314 RANDOM CHEMISTRY FOR THE GENERATION OF NEW COMPOUNDS

7. The method of claim 1 wherein the different organic molecules of the starting group are selected from the group consisting of alkanes, alkenes, alkynes, arenes, alcohols, ethers, amines, aldehydes, ketones, acids, esters, amides, cyclic compounds, heterocyclic compounds, organometallic compounds, hetero-atom bearing compounds, amino acids, nucleotides, and mixtures thereof.

Markush: Future

ontochem
IT SOLUTIONS

Limiting Markush uses is ongoing:

USPTO is trying to limit Markush:

> *For example, if the language of a claim, given its broadest reasonable interpretation, is such that a person of ordinary skill in the relevant art would read it with* ***more than one*** *reasonable* ***interpretation****, then a rejection … is appropriate.*
>
> *…However, if such a practice renders the claims* ***indefinite*** *or if it results in undue multiplicity, an appropriate rejection should be made.*

https://www.uspto.gov/web/offices/pac/mpep/s2173.html

Chinese patent office is increasingly granting patents with specific examples only

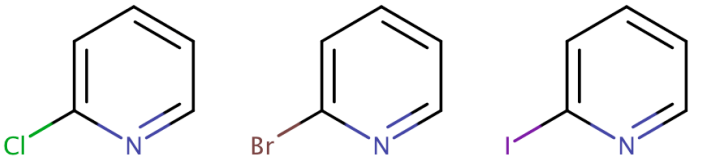
Protection of non-exemplified Markush covered claims / compounds / compositions is weak

Small Molecule Markush

ontochem IT SOLUTIONS

data mining easy:

2-chloro-, 2-bromo-, or 2-iodo-pyridine



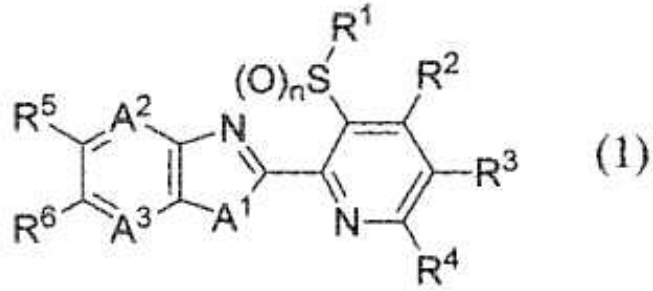
data mining impossible:

Embodiments of the present compound include compounds of the formula (1) shown in Table 1, compounds of the formula (2) shown in Table 2, compounds of the formula (2A) shown in Table 3, compounds of the formula (2B) shown in Table 4 and compounds of the formula (2C) shown in Table 5. These compounds are the compounds disclosed in International Publication WO 2013/018928 and can be produced by the methods described in this publication.

TABLE 1

The present compound	R^1	R^2	R^3	R^4	R^5	R^6	A^1	A^2	A^3	n
1	Et	H	H	H	CF_3	H	NMe	CH	N	0
2	Et	H	H	H	CF_3	H	NMe	CH	N	1
3	Et	H	H	H	CF_3	H	NMe	CH	N	2
4	Et	H	CF_3	H	CF_3	H	NMe	CH	N	0
5	Et	H	CF_3	H	CF_3	H	NMe	CH	N	2
6	Et	H	H	H	CF_2CF_3	H	NMe	CH	N	0
7	Et	H	H	H	CF_2CF_3	H	NMe	CH	N	1
8	Et	H	H	H	CF_2CF_3	H	NMe	CH	N	2

US-20160000081-A1



(1)

more about small molecule organic chemistry Markush: J.M. Barnard, P.M. Wright, "Towards in-house searching of Markush structures from patents" World Patent Information 31 (2009) 97–103

Markush Areas

ontochem
IT SOLUTIONS

breadth of today's Markush type expressions:

> *Inventions in metallurgy, refractories, ceramics, pharmacy, pharmacology and biology are most frequently claimed under the Markush formula but purely mechanical features or process steps may also be claimed by using the Markush style of claiming.*
>
> https://www.uspto.gov/web/offices/pac/mpep/s2173.html

collection of small organic compounds

formulations (tablets, cosmetics, nutrition …)

dyes (ink jet dye compositions …)

alloys (magnets, steel …)

glasses & ceramics

liquid crystals (monitors …)

…

Markush in Text: Formulations

ontochem IT SOLUTIONS

Glass example: EP-1906937-B1

Markush type claims

1. A pharmaceutical composition comprising:

3-(6-(1-(2,2-difluorobenzo[d][1,3]dioxol-5-yl)cyclopropanecarboxamido)-3-methylpyridin-2-yl)benzoic acid (Compound 1) Form I in an amount of 100 mg or 200 mg;

and

one or more excipients selected from a filler, a disintegrant, a surfactant, a diluent, a binder, a glidant, a lubricant, a colorant, and a fragrance, or any combination thereof,

wherein the pharmaceutical composition is in tablet form.

2. The pharmaceutical composition of claim 1, wherein the pharmaceutical composition comprises a filler which is microcrystalline cellulose and is present in an amount ranging from 10 wt% to 60 wt% by weight of the composition.

3. The pharmaceutical composition of claim 2, wherein the microcrystalline cellulose is present in an amount of 20 wt % to 45 wt% by weight of the composition.

example

In another embodiment, the invention provides a pharmaceutical composition comprising the following components:

High Shear Granule Blend	**mg**
Compound 1 Form I or Form II	100
Microcrystalline cellulose	33.3
Mannitol	21.7
Croscarmellose Sodium	3.3
Polyvinylpyrrolidone	4.4
Sodium Lauryl Sulfate	1.1
Core Tablet Composition (100 mg dose, 197 mg image)	**mg**
High Shear Granule Blend	163.9
Microcrystalline cellulose	27.6
Croscarmellose Sodium	3.9
Magnesium Stearate	2.0

Markush in Text: Glass Compositions

ontochem IT SOLUTIONS

US-20060039426-A1: **Novel ytterbium-phosphate glass**

PDF

We claim:

1. A ytterbium-phosphate glass comprising: 60 to 75 mole percent P_2O_5, 10 to 30 mole percent Yb_2O_3; and from 0 to 30 mole percent of a combination of two or more of the following X_2O_3, R_2O, and MO wherein;

X_2O_3 is present in an amount of from 0 to 26 mole percent and X is selected from the group consisting of Al, B, La, Sc, Y, and mixtures thereof;

R_2O is present in an amount of from 0 to 26 mole percent and R is selected from the group consisting of Li, Na, K, and mixtures thereof; and

MO is present in an amount of from 0 to 26 mole percent and M is selected from the group consisting of Mg, Ca, Sr, Ba, Zn, and mixtures thereof.

Technical Challenges & Tools

ontochem
IT SOLUTIONS

Converting terms to meaning:

1. recognizing claim lists:

   R_2O is present in an amount of from 0 to 26 mole percent and R is selected from the group consisting of Li, Na, K, and mixtures thereof;
   recognizing syntax structures

2. short terms:

   does "K" mean potassium?
   context sensitive annotation:
   using "grey lists" - confidence increased from 0 by context

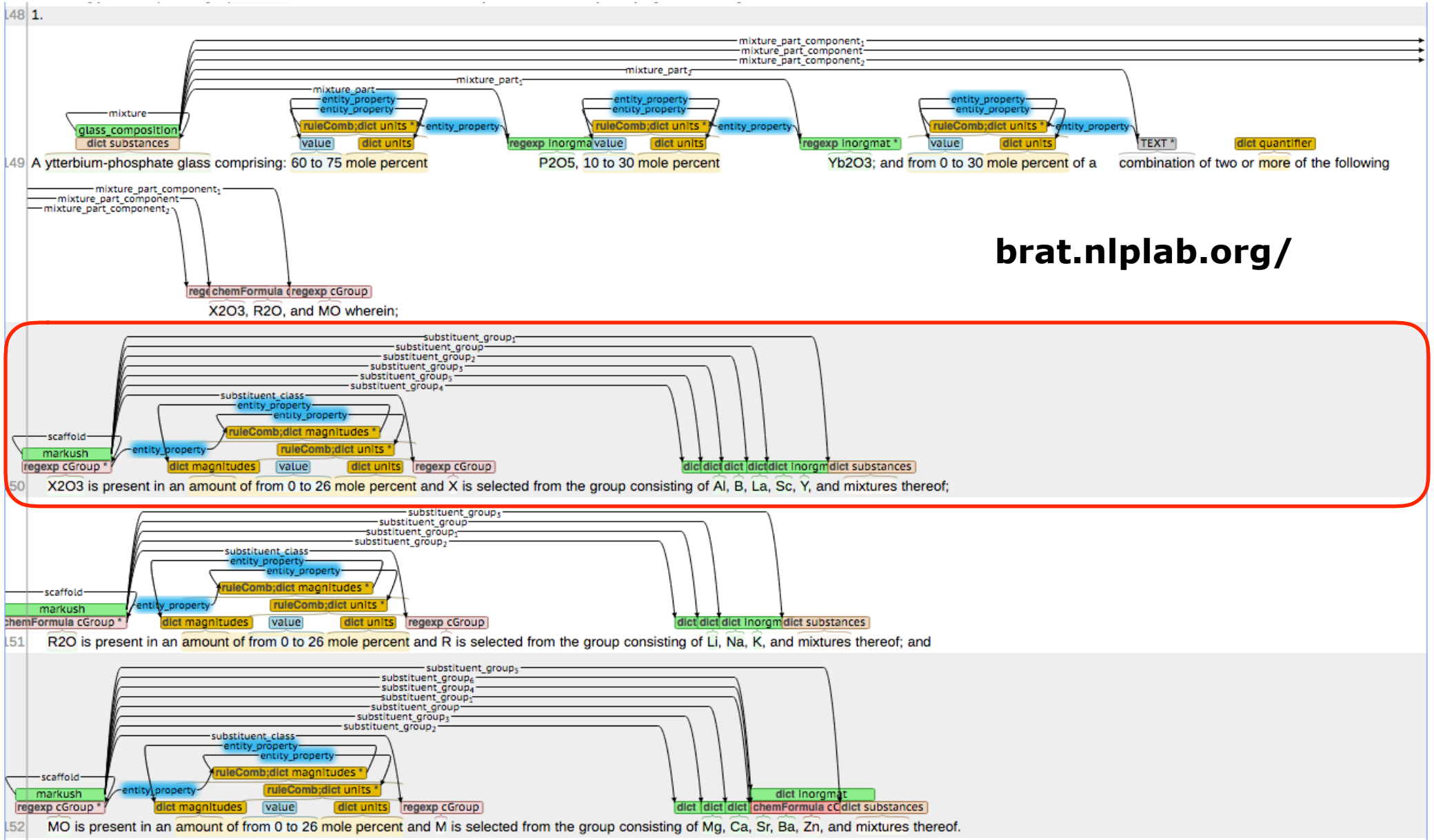
3. unusual Markush chemistry terms:

   e.g. alloys: $Sm_2(Co,Cu,Fe,Zr)_{17}$ or $Ca_{1-x-z}La_xSr_zFe_{2n-y}Co_y$

   e.g. liquid crystals: 3-HH-*-B-2

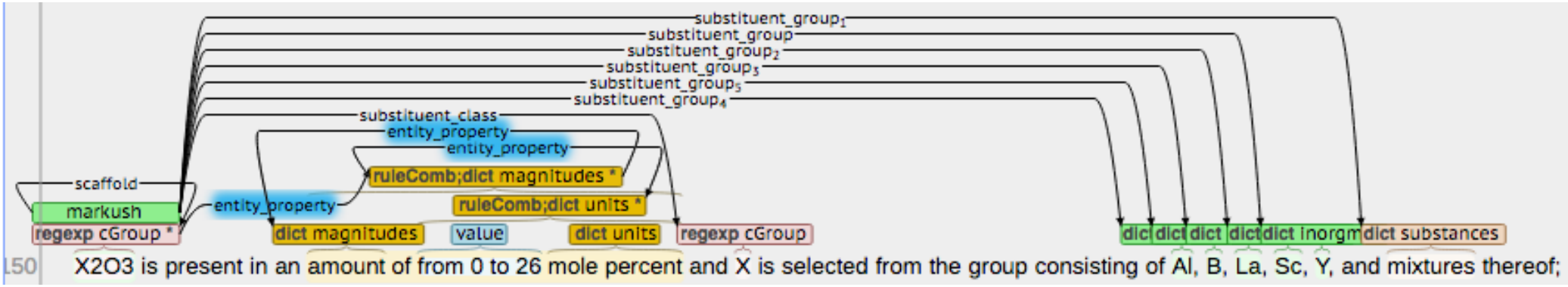
Markush Extraction: Compositions

ontochem IT SOLUTIONS



brat.nlplab.org/

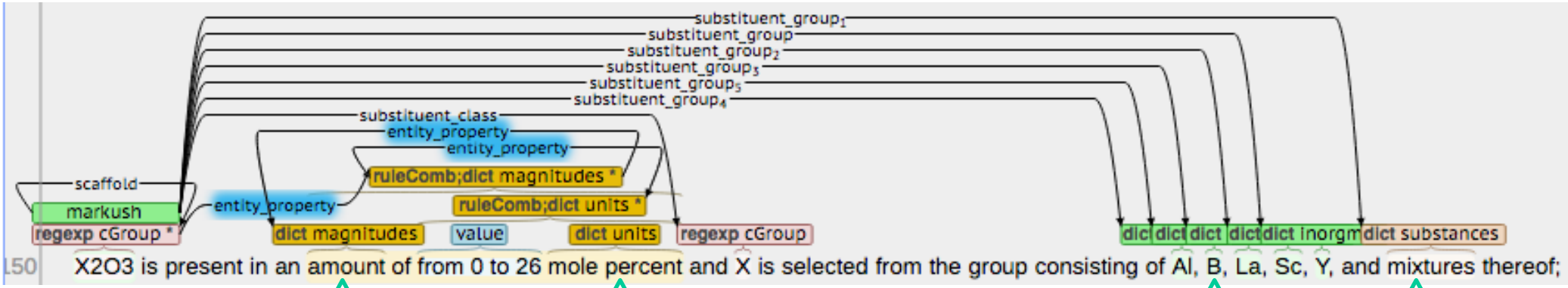
Markush Extraction: Steps



1. *dictionary based named entity recognition*

Markush Extraction: Steps

ontochem
IT SOLUTIONS



magnitude unit B substances

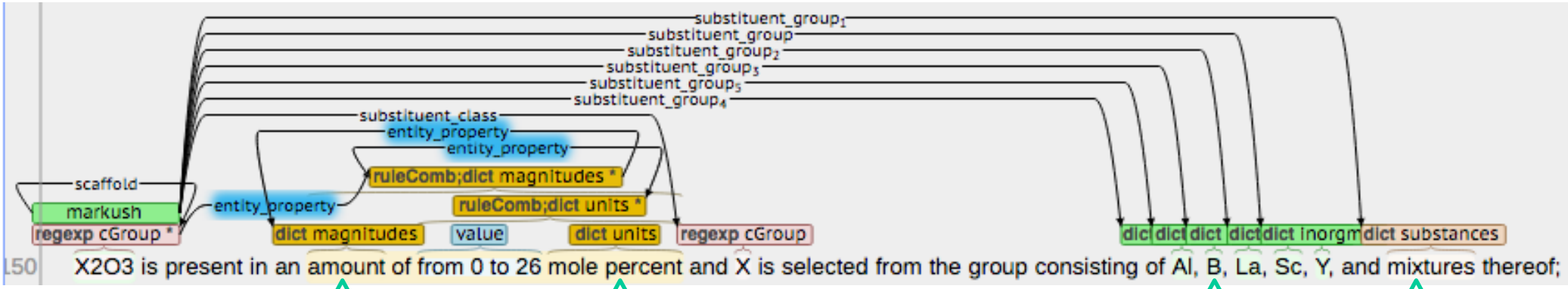
confidence =.0

inorganic materials

1. dictionary based named entity recognition

Markush Extraction: Steps

ontochem
IT SOLUTIONS



magnitude unit Al B La Sc Y substances

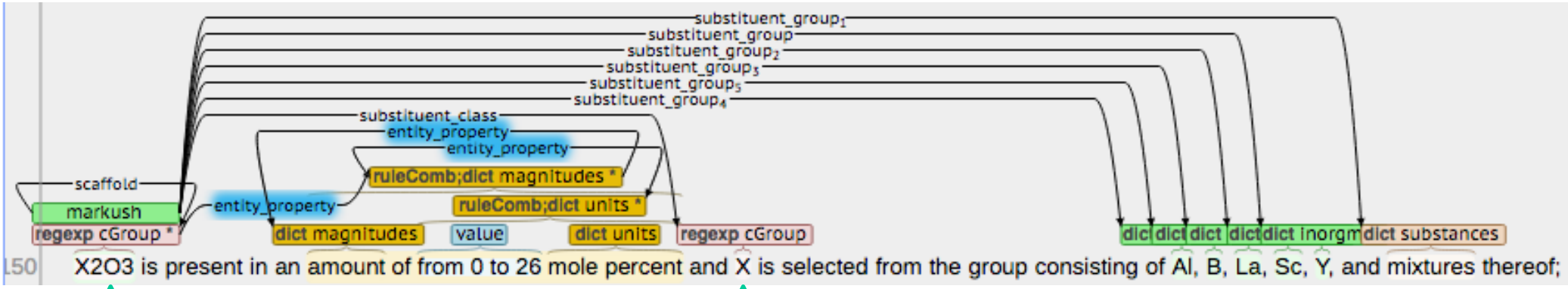
confidence =.3

inorganic materials

1. dictionary based named entity recognition

Markush Extraction: Steps

ontochem
IT SOLUTIONS

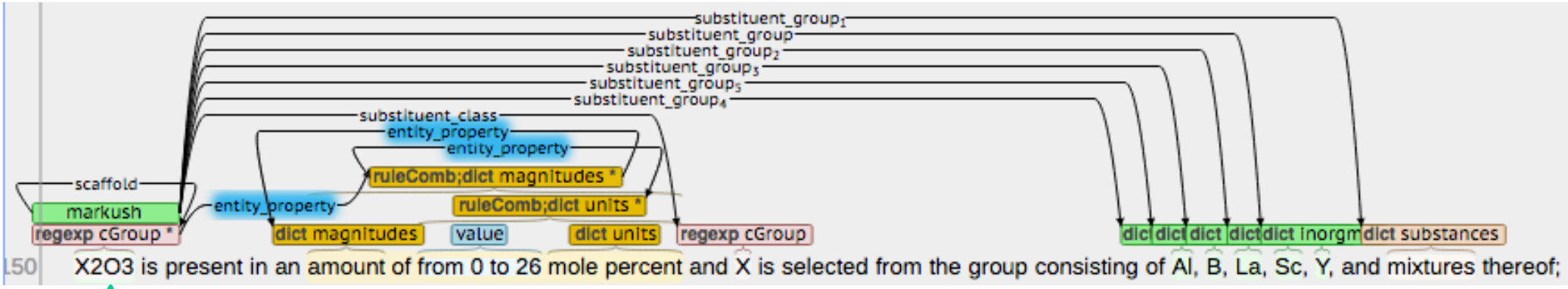


chemGroup magnitude unit chemGroup Al B La Sc Y substances

1. *dictionary based named entity recognition*
2. *regular expressions for potential chemistry*

Markush Extraction: Steps

ontochem
IT SOLUTIONS



chemGroup

magnitude 0-26 unit

chemGroup

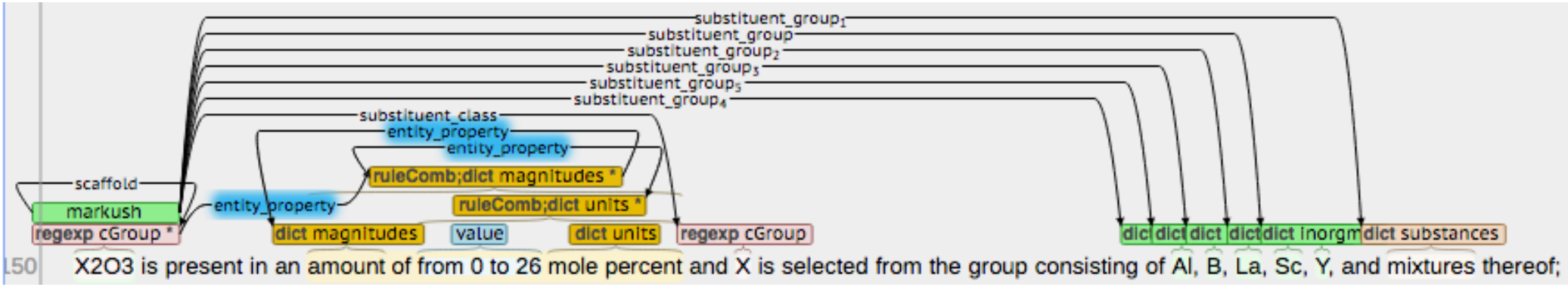
value + unit

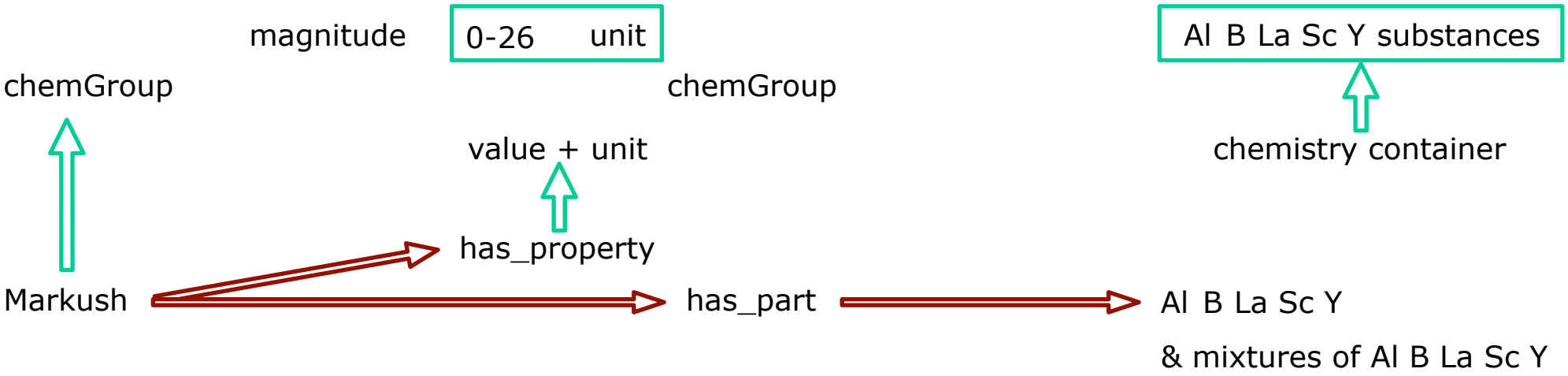
Al B La Sc Y substances

chemistry container

1. *dictionary based named entity recognition*
2. *regular expressions*
3. *rule combiner*

Markush Extraction: Steps

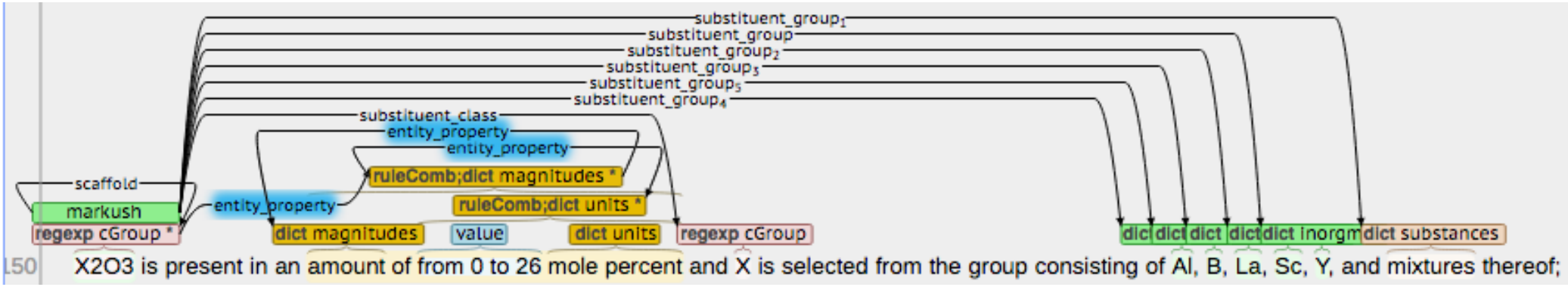




1. *dictionary based named entity recognition*
2. *regular expressions*
3. *rule combiner*
4. *syntax pattern matcher*

Markus Extraction: Steps

ontochem IT SOLUTIONS



magnitude 0-26 unit

Al B La Sc Y substances

chemGroup

chemGroup

chemistry container

value + unit

has_property

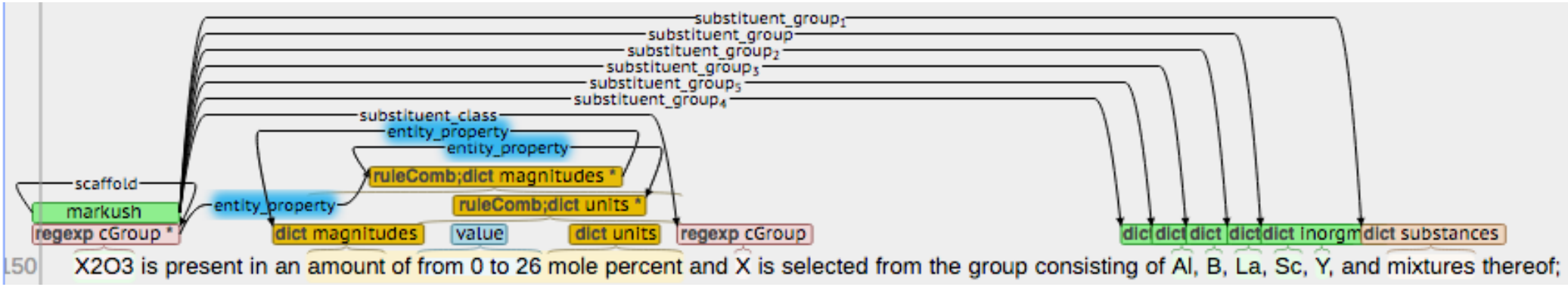
Markus has_part Al B La Sc Y

& mixtures of Al B La Sc Y

1. *dictionary based named entity recognition*
2. *regular expressions*
3. *rule combiner*
4. *syntax pattern matcher*
5. *Frame matcher* is_part_of **Markus**

Markush Extraction: Result

ontochem
IT SOLUTIONS



Pat_document-id	ITEM_ID	RELATION	CONCEPT_ID	MODULE	SOURCE_CONTEXT	SECTION	ROLE	ENTITY	CONCEPT	PROPERTY	UNIT	VALUE
US-20060039426-A1	57	[markush]	232000000014	relMatcher: 232000000014:1	X2O3 is present in an amount of from 0 to 26 mole percent and X is selected from the group consisting of Al, B, La, Sc, Y, and mixtures thereof;	claims						
	57-000						scaffold	X2O3	chemGroup	percentage	mole percent	from 0 to 26
	57-001						substituent class	X	chemGroup			
	57-002						substituent group	Al	inorgmat| 229910052782: aluminium			
	57-003						substituent group	B	inorgmat| 229910052796: boron			
	57-004						substituent group	La	inorgmat| 229910052746:l anthanum			
	57-005						substituent group	Sc	inorgmat| 229910052706:s candium			
	57-006						substituent group	Y	inorgmat| 229910052727:y ttrium			
	57-007						substituent group	mixtures	substances| 239000000203: mixtures			

Markush in Tables: LC Compositions

ontochem
IT SOLUTIONS

text mining made possible:

US 7,081,279 B2: Chisso Corporation, liquid crystals

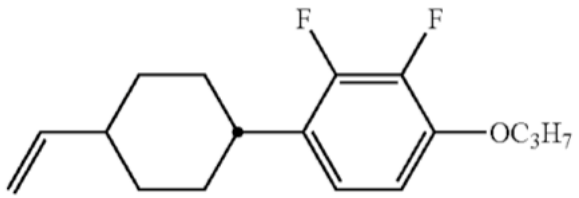
re-inventing chemical nomenclature:

Example 27

Composition Example 26

V-HB(2F,3F)-O3	(No. 11)	12.0%
V2-HB(2F,3F)-O2	(No. 8)	12.0%
V-HHB(2F,3F)-O2	(No. 31)	8.0%
V2-HHB(2F,3F)-O1	(No. 33)	8.0%
3-HH-EMe		12.0%
5-HH-EMe		5.0%
3-HEB-O2		6.0%
3-HEB-O4		8.0%
4-HEB-O2		6.0%
5-HEB-O1		6.0%
5-HEB-O2		4.0%
3-HHB-1		13.0%

compound NO.11



Patent Table Statistics

ontochem
IT SOLUTIONS

tables in US patent documents (2001-2017 August 2017)

- documents 8,436,770
- detected 10,747,205
- parsed 10,732,857
- normalized 10,601,073

table structure:

one tgroup	3,340,523
multiple tgroups	7,392,334

table content:

compositions	877,029
IC50/EC50	106,882
R-groups	102,018

Table Normalization & Correction

ontochem IT SOLUTIONS

PDF

OASIS: 5 tgroup: 1 theader (empty) + 5 tbody

TABLE 3

Inhibition (IC_{50}) of cellular proliferation in A431, SKBR3 and SW620 cells.

	Cellular Growth Inhibition IC_{50} (μM)[a]											
	A431				SKBR3				SW620			
Compound	Oxic[b]	Deact.[c]	Anoxic[d]	HCR[e]	Oxic[b]	Deact.[c]	Anoxic[d]	HCR[e]	Oxic[b]	Deact.[c]	Anoxic[d]	HCR[e]
11	0.040		0.040	1.0	0.081		0.141	0.6	3.43		2.89	1.2
12	0.010		0.023	0.4	0.044		0.028	1.6	3.81		4.01	1.0
13	0.430		0.784	0.5	1.04		0.664	1.6	27.4		29.0	1.5
14	0.015		0.009	1.7	0.028		0.027	1.1	2.02		2.46	0.8
16	1.26		1.30	1.0	3.19		5.51	0.6	18.6		19.6	0.9
17	0.462	12	0.444	1.0	1.52	19	1.84	0.8	112.5	33	111.5	1.0
18	2.34	59	0.401	5.8	3.75	44	1.29	2.8	92.2	27	79.8	1.2
19	2.64	66	0.201	13.1	6.33	78	0.483	13.1	156.4	46	28.0	5.6
20	2.09	53	0.078	26.8	3.93	49	0.217	18.1	63.3	18	20.2	3.1
21	11.0	277	13.7	1.3	26.9	332	30.0	0.9	422.6	123	365.1	1.2
22	5.86	147	0.564	10.4	11.0	136	1.64	6.7	281.3	82	123.1	2.3
23	0.670	65	0.025	26.9	2.47	56	0.068	36.1	100.8	27	13.8	7.3
27	255.3	203	76.1	3.4	506.8	159	74.3	7.6	684.3	37	584.7	1.2

Footnotes for Table 3

[a]Dose-response curves were determined at 5 concentrations. Cells received a 24 hour exposure to test compounds before being continuously washed with drug-free media. The IC_{50} (μM) values are the concentrations required to inhibit cell growth by 50%, as determined from the dose-response curves. Values are the average of between three and nine independent determinations (% CV < 20 in all cases).

[b]Experiment performed entirely under oxic conditions.

[c]Fold reduction in oxic cellular growth inhibition relative to the parent kinase inhibitor.

[d]The first 4 hours of the 24 hour drug exposure was performed under anoxic conditions.

[e]Hypoxic Cytotoxicity Ratio = fold increase in cellular growth inhibition for cells receiving 4 hours of anoxia relative to cells that received only oxic conditions.

normalized+corrected OC-XHTML

1 thead, 1 tbody, 1 tfoot, 1 anchor text

The compounds of Table 3 were tested for their ability to inhibit the proliferation of three human carcinoma cell lines, selected to provide a comparison with literature precedent: (Tsou et al, J Med Chem, 2001, 44, 2719-2734) A431 (epidermoid), which overexpresses erbB1; SKBR3 (breast), which overexpresses erbB2 and to a lesser extent, erbB1; and SW620 (colon), which serves as a control line not expressing erbB1 or erbB2 to any significant extent. The cells were exposed to test compounds for either 24 hours under oxic conditions or for 4 hours under anoxia followed by 20 hours under oxic conditions. They were then washed free of drug and incubated for a further 4 days, before being stained for cell survival with sulforhodamine B.

TABLE 3: Inhibition (IC50) of cellular proliferation in A431, SKBR3 and SW620 cells.

	Cellular Growth Inhibition IC_{50} (μM)[a]											
	A431				SKBR3				SW620			
Compound	Oxic[b]	Deact.[c]	Anoxic[d]	HCR[e]	Oxic[b]	Deact.[c]	Anoxic[d]	HCR[e]	Oxic[b]	Deact.[c]	Anoxic[d]	HCR[e]
11	0.040		0.040	1.0	0.081		0.141	0.6	3.43		2.89	1.2
12	0.010		0.023	0.4	0.044		0.028	1.6	3.81		4.01	1.0
13	0.430		0.784	0.5	1.04		0.664	1.6	27.4		29.0	1.5
14	0.015		0.009	1.7	0.028		0.027	1.1	2.02		2.46	0.8
16	1.26		1.30	1.0	3.19		5.51	0.6	18.6		19.6	0.9
17	0.462	12	0.444	1.0	1.52	19	1.84	0.8	112.5	33	111.5	1.0
18	2.34	59	0.401	5.8	3.75	44	1.29	2.8	92.2	27	79.8	1.2
19	2.64	66	0.201	13.1	6.33	78	0.483	13.1	156.4	46	28.0	5.6
20	2.09	53	0.078	26.8	3.93	49	0.217	18.1	63.3	18	20.2	3.1
21	11.0	277	13.7	1.3	26.9	332	30.0	0.9	422.6	123	365.1	1.2
22	5.86	147	0.564	10.4	11.0	136	1.64	6.7	281.3	82	123.1	2.3
23	0.670	65	0.025	26.9	2.47	56	0.068	36.1	100.8	27	13.8	7.3
27	255.3	203	76.1	3.4	506.8	159	74.3	7.6	684.3	37	584.7	1.2
Footnotes for Table 3												
[a]Dose-response curves were determined at 5 concentrations. Cells received a 24 hour exposure to test compounds before being continuously washed with drug-free media. The IC_{50} (μM) values are the concentrations required to inhibit cell growth by 50%, as determined from the dose-response curves. Values are the average of between three and nine independent determinations (% CV < 20 in all cases).												
[b]Experiment performed entirely under oxic conditions.												
[c]Fold reduction in oxic cellular growth inhibition relative to the parent kinase inhibitor.												
[d]The first 4 hours of the 24 hour drug exposure was performed under anoxic conditions.												
[e]Hypoxic Cytotoxicity Ratio = fold increase in cellular growth inhibition for cells receiving 4 hours of anoxia relative to cells that received only oxic conditions.												

Markush: Compositions from Tables

ontochem IT SOLUTIONS

US-20060039426-A1: **Novel ytterbium-phosphate glass**

PDF

TABLE 5

Component	21	22	23	24	25	26	27	28	29	30
P_2O_5	64.0	64.0	64.0	64.0	64.0	64.0	64.0	64.0	64.0	62.0
Yb_2O_3	15.0	15.0	15.0	15.0	14.0	15.0	15.0	15.0	15.0	15.0
BaO	18.0	10.0	13.0	13.0	10.0	13.0	17.0	17.0	17.0	17.0
SrO	0.0	0.0	0.0	11.0	0.0	0.0	0.0	0.0	0.0	0.0
CaO	0.0	0.0	3.0	0.0	3.0	3.0	3.0	3.0	3.0	3.0
ZnO	3.0	11.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Y_2O_3	0.0	0.0	5.0	0.0	1.0	5.0	1.0	1.0	1.0	1.0
B_2O_3	0.0	0.0	0.0	5.0	10.0	0.0	0.0	0.0	0.0	0.0
Er_2O_3	10.0	0.0	0.0	0.0	0.0	5.0	10.0	15.0	0.0	0.0
Nd_2O_3	0.0	10.0	0.0	0.0	20.0	0.0	0.0	5.0	15.0	12.5
Ho_2O_3	0.0	0.0	10.0	0.0	0.0	0.0	0.0	0.0	5.0	0.0
Tm_2O_3	0.0	0.0	0.0	10.0	0.0	0.0	0.0	0.0	0.0	5.0

OC|miner OC_XHTML

TABLE 5

Component	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
P_2O_5	64.0	64.0	64.0	64.0	64.0	64.0	64.0	64.0	64.0	62.0
Yb_2O_3	15.0	15.0	15.0	15.0	14.0	15.0	15.0	15.0	15.0	15.0
BaO	18.0	10.0	13.0	13.0	10.0	13.0	17.0	17.0	17.0	17.0
[illegible]	0.0	0.0	0.0	11.0	0.0	0.0	0.0	0.0	0.0	0.0
[illegible]	0.0	0.0	3.0	0.0	3.0	3.0	3.0	3.0	3.0	3.0
ZnO	3.0	11.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Y_2O_3	0.0	0.0	5.0	0.0	1.0	5.0	1.0	1.0	1.0	1.0
B_2O_3	0.0	0.0	0.0	5.0	10.0	0.0	0.0	0.0	0.0	0.0
Er_2O_3	10.0	0.0	0.0	0.0	0.0	5.0	10.0	15.0	0.0	0.0
Nd_2O_3	0.0	10.0	0.0	0.0	20.0	0.0	0.0	5.0	15.0	12.5
Ho_2O_3	0.0	0.0	10.0	0.0	0.0	0.0	0.0	0.0	5.0	0.0
Tm_2O_3	0.0	0.0	0.0	10.0	0.0	0.0	0.0	0.0	0.0	5.0

Markush: Compositions from Tables

ontochem IT SOLUTIONS

US-20060039426-A1: **Novel ytterbium-phosphate glass**

annotated XML

TABLE 5

Component	21	22	23	24	25	26	27	28	29	30
P_2O_5	64.0	64.0	64.0	64.0	64.0	64.0	64.0	64.0	64.0	62.0
Yb_2O_3	15.0	15.0	15.0	15.0	14.0	15.0	15.0	15.0	15.0	15.0
BaO	18.0	10.0	13.0	13.0	10.0	13.0	17.0	17.0	17.0	17.0
SrO	0.0	0.0	0.0	11.0	0.0	0.0	0.0	0.0	0.0	0.0
CaO	0.0	0.0	3.0	0.0	3.0	3.0	3.0	3.0	3.0	3.0
ZnO	3.0	11.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Y_2O_3	0.0	0.0	5.0	0.0	1.0	5.0	1.0	1.0	1.0	1.0
B_2O_3	0.0	0.0	0.0	5.0	10.0	0.0	0.0	0.0	0.0	0.0
Er_2O_3	10.0	0.0	0.0	0.0	0.0	5.0	10.0	15.0	0.0	0.0
Nd_2O_3	0.0	10.0	0.0	0.0	20.0	0.0	0.0	5.0	15.0	12.5
Ho_2O_3	0.0	0.0	10.0	0.0	0.0	0.0	0.0	0.0	5.0	0.0
Tm_2O_3	0.0	0.0	0.0	10.0	0.0	0.0	0.0	0.0	0.0	5.0

OC|miner OC_XHTML

TABLE 5

Component	21	22	23	24	25	26	27	28	29	30
P_2O_5	64.0	64.0	64.0	64.0	64.0	64.0	64.0	64.0	64.0	62.0
Yb_2O_3	15.0	15.0	15.0	15.0	14.0	15.0	15.0	15.0	15.0	15.0
BaO	18.0	10.0	13.0	13.0	10.0	13.0	17.0	17.0	17.0	17.0
SrO	0.0	0.0	0.0	11.0	0.0	0.0	0.0	0.0	0.0	0.0
CaO	0.0	0.0	3.0	0.0	3.0	3.0	3.0	3.0	3.0	3.0
ZnO	3.0	11.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Y_2O_3	0.0	0.0	5.0	0.0	1.0	5.0	1.0	1.0	1.0	1.0
B_2O_3	0.0	0.0	0.0	5.0	10.0	0.0	0.0	0.0	0.0	0.0
Er_2O_3	10.0	0.0	0.0	0.0	0.0	5.0	10.0	15.0	0.0	0.0
Nd_2O_3	0.0	10.0	0.0	0.0	20.0	0.0	0.0	5.0	15.0	12.5
Ho_2O_3	0.0	0.0	10.0	0.0	0.0	0.0	0.0	0.0	5.0	0.0
Tm_2O_3	0.0	0.0	0.0	10.0	0.0	0.0	0.0	0.0	0.0	5.0

Markush: Compositions from Tables

ontochem IT SOLUTIONS

US-20060039426-A1: **Novel ytterbium-phosphate glass**

annotated XML

TABLE 5

Component	21	22	23	24	25	26	27	28	29	30
P_2O_5	64.0	64.0	64.0	64.0	64.0	64.0	64.0	64.0	64.0	62.0
Yb_2O_3	15.0	15.0	15.0	15.0	14.0	15.0	15.0	15.0	15.0	15.0
BaO	18.0	10.0	13.0	13.0	10.0	13.0	17.0	17.0	17.0	17.0
SrO	0.0	0.0	0.0	11.0	0.0	0.0	0.0	0.0	0.0	0.0
CaO	0.0	0.0	3.0	0.0	3.0	3.0	3.0	3.0	3.0	3.0
ZnO	3.0	11.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Y_2O_3	0.0	0.0	5.0	0.0	1.0	5.0	1.0	1.0	1.0	1.0
B_2O_3	0.0	0.0	0.0	5.0	10.0	0.0	0.0	0.0	0.0	0.0
Er_2O_3	10.0	0.0	0.0	0.0	0.0	5.0	10.0	15.0	0.0	0.0
Nd_2O_3	0.0	10.0	0.0	0.0	20.0	0.0	0.0	5.0	15.0	12.5
Ho_2O_3	0.0	0.0	10.0	0.0	0.0	0.0	0.0	0.0	5.0	0.0
Tm_2O_3	0.0	0.0	0.0	10.0	0.0	0.0	0.0	0.0	0.0	5.0

extracted **US-20060039426-A1_table_5.csv**:

MODULE	SECTION	ENTITY	NAME	CONCEPT	PROPERTY	VALUE
TableMixtureParts Extractor	description					
		mixture	22	substances		
		mixture part	P2O5	inorgmat	percentage	64
		mixture part	Yb2O3	inorgmat	percentage	15
		mixture part	BaO	inorgmatl229910014179:BaO	percentage	10
		mixture part	SrO	inorgmatl229910045055:SrO	percentage	0
		mixture part	CaO	chemCompoundl 190000021704:calcium monoxide	percentage	0
		mixture part	ZnO	chemCompoundl 190000022155:zinc monoxide	percentage	11
		mixture part	Y2O3	inorgmatl 229910010292:Yttrium(III) oxide	percentage	0
		mixture part	B2O3	inorgmatl 229910011255:B2O3	percentage	0
		mixture part	Er2O3	inorgmatl 229910000125:erbium(III) oxide	percentage	0
		mixture part	Nd2O3	inorgmatl 229910014145:Nd2–O3	percentage	10
		mixture part	Ho2O3	inorgmatl 229910000133:holmium(III) oxide	percentage	0
		mixture part	Tm2O3	inorgmat	percentage	0

Patent Classification

ontochem
IT SOLUTIONS

Markush Classification Pipeline:

- UIMA (**U**nstructured **I**nformation **M**anagement **A**rchitecture)

  tokenization, annotation, feature extraction

- UIMA + Machine Learning = ClearTK

  classification using SVM (Support Vector Machine) + Random Forrests

- Increase word frequencies: stemming

  "stems", "stemmer", "stemming", "stemmed" => "stem"

  "fishing", "fished", "fisher" => "fish"

- Increase word frequencies: Named Entity normalization

  "SiO2", "Silicon oxide", "Quartz", "Quartz silica" => "silicium dioxide"

  "Al2-O3", "aluminium(III) oxide", "aloxide" => "Al2O3"

Table Classification & Extraction

ontochem IT SOLUTIONS

identifying the right table: **anchoring text**

US20150246842A1	**Table 3** **anchoring text:**	The following glasses in Table 3 were prepared at very low total iron oxide levels with cerium oxide contents from 0.50 wt. % down to 0 wt. %. The data includes the calculated batch data after the batch has lost water and carbon dioxide as a result of the glassmaking process, the data for the three glasses that were analyzed for composition, viscosity data, liquidus data, and spectral data.

TABLE 8

Component	HIRA Lithium glass Campaign A	LIRA Lithium glass Campaign B
SiO_2	59-63 wt. %	60-63 wt. %
Na_2O	10-13 wt. %	10-12 wt. %
Li_2O	4-5.5 wt. %	4-5.5 wt. %
Al_2O_3	15-23 wt. %	17-19 wt. %
ZrO_2	2-5 wt. %	3.5-5 wt. %
$(Al_2O_3 + ZrO_2)$	19-25 wt. %	21.5-24 wt. %
FeO	0.02-0.05 wt. %	0.001-0.010 wt. %
FeO/Fe_2O_3	0.2-0.4	0.005-0.15
CeO_2	0.00	wt. %
Fe_2O_3 (total iron)	800-1200 ppm	50-less than 800

annotated:

TABLE 8

Component		HIRA Lithium glass Campaign A	LIRA Lithium glass Campaign B	
SiO_2	59-63	wt. %	60-63	wt. %
Na_2O	10-13	wt. %	10-12	wt. %
Li_2O	4-5.5	wt. %	4-5.5	wt. %
Al_2O_3	15-23	wt. %	17-19	wt. %
ZrO_2	2-5	wt. %	3.5-5	wt. %
$(Al_2O_3 + ZrO_2)$	19-25	wt. %	21.5-24	wt. %
FeO	0.02-0.05	wt. %	0.001-0.010	wt. %
FeO/Fe_2O_3		0.2-0.4	0.005-0.15	
CeO_2		0.00	0.02-0.45	wt. %
Fe_2O_3 (total iron)	800-1200	ppm	50-less than 800	ppm

Table Extraction Problems

ontochem IT SOLUTIONS

problems in current Markush data extraction:

- lacking standardization in inorganic chemistry
- lacking formalization of conditional data
- overdetermined data
- QA tools

TABLE 8

Component	HIRA Lithium glass Campaign A	LIRA Lithium glass Campaign B
SiO_2	59-63 wt. %	60-63 wt. %
Na_2O	10-13 wt. %	10-12 wt. %
Li_2O	4-5.5 wt. %	4-5.5 wt. %
Al_2O_3	15-23 wt. %	17-19 wt. %
ZrO_2	2-5 wt. %	3.5-5 wt. %
$(Al_2O_3 + ZrO_2)$	19-25 wt. %	21.5-24 wt. %
FeO	0.02-0.05 wt. %	0.001-0.010 wt. %
FeO/Fe_2O_3	0.2-0.4	0.005-0.15
CeO_2	0.00	0.02-0.45 wt. %
Fe_2O_3 (total iron)	800-1200 ppm	50-less than 800 ppm

annotated:

TABLE 8

Component		HIRA Lithium glass Campaign A	LIRA Lithium glass Campaign B
SiO_2	59-63	wt. %	60-63 wt. %
Na_2O	10-13	wt. %	10-12 wt. %
Li_2O	4-5.5	wt. %	4-5.5 wt. %
Al_2O_3	15-23	wt. %	17-19 wt. %
ZrO_2	2-5	wt. %	3.5-5 wt. %
$(Al_2O_3 + ZrO_2)$	19-25	wt. %	21.5-24 wt. %
FeO	0.02-0.05	wt. %	0.001-0.010 wt. %
FeO/Fe_2O_3		0.2-0.4	0.005-0.15
CeO_2		0.00	0.02-0.45 wt. %
Fe_2O_3 (total iron)	800-1200	ppm	50-less than 800 ppm

condition

Alloys Data Extraction

ontochem IT SOLUTIONS

Relationship extraction

ELSEVIER

Journal of Magnetism and Magnetic Materials 242–245 (2002) 1277–1283

Journal of magnetism and magnetic materials

www.elsevier.com/locate/jmmm

Invited paper

Nanocrystalline high performance permanent magnets

O. Gutfleisch*, A. Bollero, A. Handstein, D. Hinz, A. Kirchner, A. Yan, K.-H. Müller, L. Schultz

Institute of Solid State and Materials Research, IFW Dresden, P.O. Box 270016, 01171 Dresden, Germany

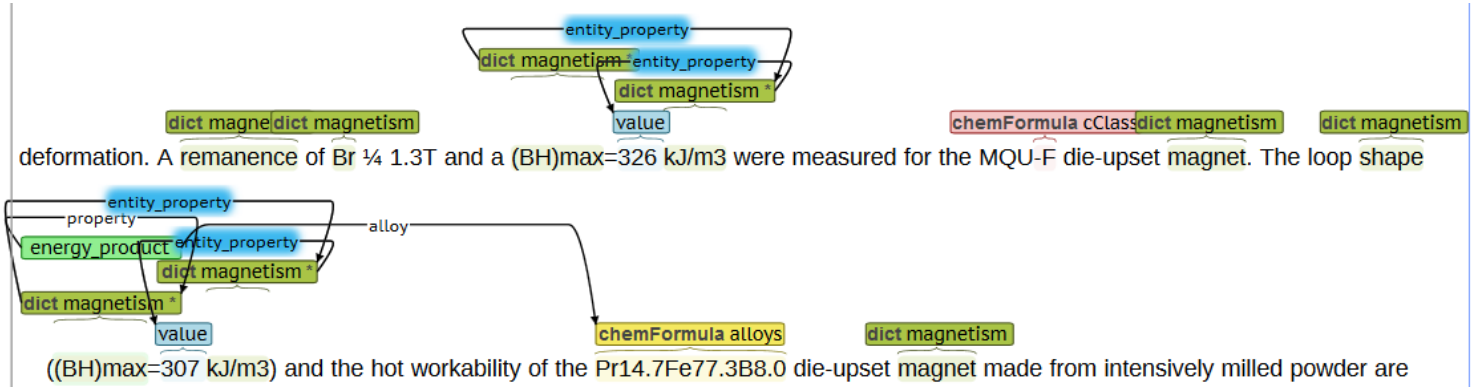
Abstract

Recent developments in nanocrystalline rare earth–transition metal magnets are reviewed and emphasis is placed on research work at IFW Dresden. Principal synthesis methods include high energy ball milling, melt spinning and hydrogen assisted methods such as reactive milling and hydrogenation-disproportionation-desorption-recombination. These techniques are applied to NdFeB-, PrFeB- and SmCo-type systems with the aim to produce high remanence magnets with high coercivity. Concepts of maximizing the energy density in nanostructured magnets by either inducing a texture via anisotropic HDDR or hot deformation or enhancing the remanence via magnetic exchange coupling are evaluated.

Keywords: Permanent magnets; Nanocrystalline materials; Exchange coupling; Texture; Hydrogen absorption

"8th Multiscale Materials Modeling" international conference MMM 2016, Dijon, France

grains after hot deformation. A remanence of $B_r = 1.3\,T$ and a $(BH)_{max} = 326\,kJ/m^3$ were measured for the MQU-F die-upset magnet. The loop shape ($(BH)_{max} = 307\,kJ/m^3$) and the hot workability of the $Pr_{14.7}Fe_{77.3}B_{8.0}$ die-upset magnet made from intensively milled powder are excellent and it can be expected that



Extracted Data: alloys + magnetism

ontochem IT SOLUTIONS

CONCEPT	SOURCE_CONTEXT	CREATED_BY_MODULE	PUB_DATE	BIB_DESCRIPTION	alloy	alloy_ELEMENTS	FORMULA	property	unit	value
[coercivity]	Nd–Fe–B magnets (μ0Hc	relMatcher;frameEnricher	2014	Acta Materialia 82, 336-343.	Nd–Fe–B	Nd;Fe;B		μ0Hc	T	1.2
[coercivity]	Nd–Fe–B magnets is reduced .....t high temperature (μ0Hc	relMatcher;frameEnricher	2012	Scripta Materialia 67 (6), 536-541.	Nd–Fe–B	Nd;Fe;B		μ0Hc		
[coercivity]	μ0Hc(T) in (Ba1−xLax)-(Fe12−xCox)O19	relMatcher	2002	Physica B: Physics of Condensed Matter 319 (1), 127-132.	(Ba1−xLax)-(Fe12−xCox)O19	La;Ba;Co;Fe;O		μ0Hc		
[coercivity]	μ0Hc ≈ 2.7 T) have been recently achieved in heavy rare earth free NdFeB	ruleComb;frameEnricher	2012	Acta Materialia 60 (9), 3783-3788.	NdFeB	Nd;Fe;B		μ0Hc	T	2.7
[coercivity]	NdFeB/Ta films (μ0Hc	ruleComb;frameEnricher	2012	Acta Materialia 60 (9), 3783-3788.	NdFeB	Nd;Fe;B		μ0Hc	T	2.7
[coercivity]	Nd–Fe–B sintered magnets with a grain size of ~1 μm and μ0Hc	relMatcher	2014	Acta Materialia 82, 336-343.	Nd–Fe–B	Nd;Fe;B		μ0Hc	T	2
[coercivity]	μ0Hc(T) J..... FePt/Fe3O4	TableCreator	2011	Handbook of Magnetic Materials 19, 291-407.	Fe3O4	Fe;O	Fe72.36O27.64	μ0Hc		
[coercivity]	μ0Hc(T) J..... Bulk Fe61.5Pt38.5	TableCreator	2011	Handbook of Magnetic Materials 19, 291-407.	Fe61.5Pt38.5	Pt;Fe	Pt38.5Fe61.5	μ0Hc		
[coercivity]	μ0Hc(T) J..... Electrodeposited Fe37Pt38O25	TableCreator	2011	Handbook of Magnetic Materials 19, 291-407.	Fe37Pt38O25	Pt;Fe;O	Pt38Fe37O25	μ0Hc		
[coercivity]	μ0Hc(T) J..... Nanocrystalline Fe50Pt50	TableCreator	2011	Handbook of Magnetic Materials 19, 291-407.	FePt	Pt;Fe		μ0Hc		
[coercivity]	μ0Hc(T) J..... Fe50Pt50	TableCreator	2011	Handbook of Magnetic Materials 19, 291-407.	FePt	Pt;Fe		μ0Hc		
[coercivity]	μ0Hc(T) J..... Nanocrystalline Fe60Pt40	TableCreator	2011	Handbook of Magnetic Materials 19, 291-407.	Fe3Pt	Pt;Fe	Pt53.8Fe46.2	μ0Hc		
[coercivity]	μ0Hc(T) J..... Bulk Fe61.5Pt38.5 alloy	TableCreator	2011	Handbook of Magnetic Materials 19, 291-407.	alloy			μ0Hc		
[coercivity]	μ0Hc(T) J..... FePt	TableCreator	2011	Handbook of Magnetic Materials 19, 291-407.	FePt	Pt;Fe		μ0Hc		
[coercivity]	Fe80Ta8C12 films showed an amo.....s ≈ 1.7 T and a low μ0Hc	ruleComb;frameEnricher	2011	Handbook of Magnetic Materials 19, 291-407.	Fe80Ta8C12	Ta;Fe;C	Ta8Fe80C12	μ0Hc	mT	0.03
[coercivity]	Fe80Ta8C12 films showed an amorphous structure with low saturation magnetisation (Js ≈ 0.75 T) and coercivity μ0Hc	ruleComb;frameEnricher	2011	Handbook of Magnetic Materials 19, 291-407.	Fe80Ta8C12	Ta;Fe;C	Ta8Fe80C12	μ0Hc	mT	1.8
[coercivity]	SmFeN powder: μ0Hc	ruleComb;frameEnricher	2012	Scripta Materialia 67 (2), 153-156.	SmFeN	Sm;Fe;N		μ0Hc	T	1.2
[coercivity]	Sm–Fe–N magnets D. Prabhua .....large coercivity of μ0Hc	ruleComb;frameEnricher	2012	Scripta Materialia 67 (2), 153-156.	Sm–Fe–N	Sm;Fe;N		μ0Hc	T	2.75
[coercivity]	Sm–Fe–N bonded magnet with the maximum energy product (BH)max = 158 kJ m−3 and μ0Hc	ruleComb;frameEnricher	2012	Scripta Materialia 67 (2), 153-156.	Sm–Fe–N	Sm;Fe;N		μ0Hc	T	0.8
[coercivity]	Sm–Fe–N bonded magnet with μ0Hc	ruleComb;frameEnricher	2012	Scripta Materialia 67 (2), 153-156.	Sm–Fe–N	Sm;Fe;N		μ0Hc	T	1.9
[coercivity]	Sm2Fe17Nx compacts at 435 °C results in a substantial enhancement of coercivity to μ0Hc	ruleComb;frameEnricher	2012	Scripta Materialia 67 (2), 153-156.	Sm2Fe17Nx	Sm;Fe;N		μ0Hc	T	2.75
[coercivity]	U2Fe13Si4 U2Fe..... μ0Hc	TableCreator	2002	Physica B: Physics of Condensed Matter 319 (1), 208-219.	U2Fe13Si4	U;Fe;Si		μ0Hc	T	3.1

Resulting Software & Tools:

ontochem
IT SOLUTIONS

OC|processor UIMA document processing pipeline, including modules:

name-2-chemistry

recognizing chemical terms: dictionary terms, IUPAC names, formulas, alloys, polymers, liquid crystals, peptides, DNA & RNA sequences, MolPuzzler

rule combiner

rule based entity recognition, e.g. "mdm2 antibody", "microcrystalline cellulose"

pattern matcher

recognizes syntax pattern, e.g. "a group of A, B, C"

document-2-table&image

extracts and corrects tables and images

frame enricher

extracts complex relationships

relationship extractor

extracts defined relationships

Thanks to...

ontochem
IT SOLUTIONS

OIS team:

Claudia Bobach, Timo Böhme, Matthias Irmer, Tino Müller, Konstantin Kruse

Financial Support: