

Solr PSICQUIC Webservice

35 fields

Reference27&followup=http%3A%2F%2Fcode.google.com%2Fp%2Fpsicquic%2Fwiki%2FMIqReference27

Most Visited Getting Started Latest Headlines Editor

Grouping and field grouping: brca2 AND (mouse "in vitro")

Fields

The following table shows the available standard fields that can be used in PSICQUIC searches:

Field Name	Searches on	MITAB 2.7 Columns	Example
idA	Identifier A	1	idA:P74565
idB	Identifier B	2	idB:P74565
id	Identifiers (A or B)	1,4	id:P74565
alias	Aliases (A or B)	5,6	alias:(KHDRB51 OR HCK)
identifier	Identifiers (A or B) or Alternatives (A or B) Aliases (A or B)	1,6	identifier:P74565
pubauth	Publication 1st author(s)	8	pubauth:scott
pubid	Publication Identifier(s)	9	pubid:(10837477 OR 12029088)
taxidA	Tax ID interactor A: be it the tax ID or the species name	10	taxidA:mouse
taxidB	Tax ID interactor B: be it the tax ID or species name	11	taxidB:9606
species	Species. Tax ID A and Tax ID B	10,11	species:human
type	Interaction type(s)	12	type:"physical association"
detmethod	Interaction Detection method(s)	7	detmethod:"two hybrid"
interaction_id	Interaction identifier(s)	14	interaction_id:EBI-761050
pbioroleA	Biological role A	17	pbioroleA:ancillary
pbioroleB	Biological role B	18	pbioroleB:"MI:0684"
pbiorole	Biological roles (A or B)	17,18	pbiorole:enzyme
pptypeA	Interactor type A	21	pptypeA:protein
pptypeB	Interactor type B	22	pptypeB:"gene"
pptype	Interactor types (A or B)	21,22	pbiorole:small molecule
pxrefA	Interactor xref A (or Identifier A)	23,1	pxrefA:"GO:0003824"
pxrefB	Interactor xref B (or Identifier B)	24,2	pxrefB:"GO:0003824"
pxref	Interactor xrefs (A or B or Identifier A or Identifier B)	23,24,1,2	pxref:catalytic activity
xref	Interaction xrefs (or Interaction identifiers)	25,14	xref:nuclear pore
annot	Interaction annotations and tags	26	annot:internally curated
update	Update date	32	update:[20100101 TO 20120101]
negative	Negative interaction boolean	36	negative:true
complex	Complex expansion	16	complex:spoke expanded OR complex:"MI:1060"
ftypeA	Feature type of participant A	37	ftypeA:sufficient to bind
ftypeB	Feature type of participant B	38	ftypeB:mutation
ftype	Feature type of participant A or B	37,38	ftype:binding site
pmethodA	Participant identification method A	41	pmethodA:western blot
pmethodB	Participant identification method B	42	pmethodB:sequence tag identification
pmethod	Participant identification methods (A or B)	41,42	pmethod:immunostaining
sto	Stoichiometry (A or B). Only true or false, just to be able to filter interaction having stoichiometry available	39,40	sto:true
param	Interaction parameters. Only true or false, just to be able to filter interaction having parameters available	30	param:true

Enter a comment: Hint: You can use [Wiki Syntax](#).

Indexing Component

Lucene 2.3.2 to 3.6.0

New Schema

resolved issue with previous text analyzer (split between
'.' + duplicate terms) **<miql_field>**, we store the unique
set of store db, value, txt, db:value

added copy fields (indexed, unstored) to generate
composite fields (id, identifier, species)

added indexed string fields (unstored) for faceting and
sorting (<db>:<value>) **<miql_field>_s**

added stored fields (not-indexed) to flexibly retrieve
original fields **<mitab_field>_o** using the solr option
fl=x,y,z

UUID for distributed searching

```
<fieldType name="miglTextOld" class="solr.TextField" positionIncrementGap="100" indexed="true" stored="true" omitNorms="true"
termVectors="true" required="true" multiValued="true">
  <analyzer>
    <tokenizer class="solr.ClassicTokenizerFactory"/>      <!-- StandardTokenizer's behavior pre-Solr 3.1 -->
    <filter class="solr.ClassicFilterFactory"/>             <!-- Removes dots from acronyms and 's from the end of tokens -->
    <filter class="solr.LowerCaseFilterFactory"/>           <!-- "I.B.M.", "Solr" ==> "i.b.m.", "solr" -->
    <filter class="solr.StopFilterFactory" words="stopwords.txt" ignoreCase="true" enablePositionIncrements="false"/> <!--
Discards common words: "a", "an", "and", "are", "as", "at", "be", "but", "by" -->
    <!-- <filter class="solr.RemoveDuplicatesTokenFilterFactory"/> -->
    <!-- <filter class="solr.SynonymFilterFactory" synonyms="syn.txt" ignoreCase="true" expand="false"/> -->
  </analyzer>
</fieldType>
```

taxid:4932(yeast)|taxid:4932("Saccharomyces cerevisiae (Baker's yeast)")|taxid:1111708(syn3)|taxid:1111708("Synechocystis sp. (strain PCC 6803 / Kazusa)")

```
org.apache.solr.analysis.ClassicTokenizerFactory {luceneMatchVersion=LUCENE_36}
```

[illegible]

Indexing Component

No	Rank ▼	Field	Text
1	17890	idA	innatedb
2	17890	idA	ensembl
3	531	idA	ensg00000120738
4	531	idA	idbg-47191
5	291	idA	ensmusg00000028717
6	291	idA	idbg-175763
7	209	idA	idbg-57543
8	209	idA	ensg00000173039
9	160	idA	idbg-89258
10	160	idA	ensg00000198793
11	148	idA	idbg-196509
12	148	idA	ensg00000125347
13	148	idA	idbg-42125
14	148	idA	ensmusg00000041515
15	143	idA	idbg-40102
16	143	idA	ensg00000175104
17	140	idA	idbg-63225
18	140	idA	ensg00000126456
19	135	idA	idbg-77617
20	135	idA	ensg00000115415
21	123	idA	idbg-79573
22	123	idA	ensg00000118260
23	122	idA	ensg00000101966

No	Rank ▼	Field	Text
1	14	idA	uniprotkb
2	8	idA	uniprotkb:o43521
3	8	idA	o43521
4	4	idA	q07812
5	4	idA	uniprotkb:q07812
6	1	idA	uniprotkb:q07820
7	1	idA	q92843
8	1	idA	uniprotkb:q92843
9	1	idA	q07820

Indexing Component

1	17875	detmethod	mi
2	17875	detmethod	psi
3	11128	detmethod	tag
4	11127	detmethod	anti
5	11124	detmethod	coimmunoprecipitation
6	11124	detmethod	0007
7	11124	detmethod	MI:0007
8	2270	detmethod	assay
9	1777	detmethod	down
10	1777	detmethod	pull
11	1246	detmethod	0096
12	1246	detmethod	MI:0096
13	964	detmethod	mobility
14	964	detmethod	electrophoretic
15	808	detmethod	supershift
16	808	detmethod	MI:0412
17	808	detmethod	0412
18	803	detmethod	hybrid
19	799	detmethod	two
20	787	detmethod	0018
21	787	detmethod	MI:0018
22	776	detmethod	affinity
23	757	detmethod	kinase
24	574	detmethod	technology
25	547	detmethod	chromatography
26	534	detmethod	MI:0004
27	534	detmethod	0004
28	528	detmethod	qst
29	528	detmethod	0059
30	528	detmethod	MI:0059

No	Rank ▼	Field	Text
1	14	detmethod	psi-mi
2	8	detmethod	surface
3	8	detmethod	mi:0107
4	8	detmethod	psi-mi:mi:0107
5	8	detmethod	resonance
6	8	detmethod	plasmon
7	4	detmethod	anti
8	4	detmethod	coimmunoprecipitation
9	4	detmethod	psi-mi:mi:0007
10	4	detmethod	mi:0007
11	4	detmethod	tag
12	2	detmethod	x-ray
13	2	detmethod	mi:0114
14	2	detmethod	psi-mi:mi:0114
15	2	detmethod	crystallography

SOLR indexing

- ✧ Read **MITAB 2.5**, **MITAB 2.6** or **MITAB 2.7**
(calimochó)
- ✧ Spring batch 2.1.8.RELEASE : **chunk processing**
- ✧ Indexing : solrj 3.0.6
- ✧ Same SOLR schema based on **MIQL 2.7** (<http://code.google.com/p/psicquic/wiki/MiqlReference27>)
- ✧ Source code : <http://psicquic.googlecode.com/svn/trunk/psicquic-solr>

How to start process

- Start SOLR server
- `./indexMitab27.sh ${mitab_file} "http://localhost:8983/solr"`
- Start PSICQUIC webservice (not done yet)

http://localhost:8983/solr/select/?q=%3A*&version=2.2&start=0&rows=10&indent=on&fl=idA_o,idB_o

```
<response>
  <lst name="responseHeader">
    <int name="status">0</int>
    <int name="QTime">0</int>
    <lst name="params">
      <str name="fl">idA_o,idB_o</str>
      <str name="indent">on</str>
      <str name="start">0</str>
      <str name="q">*:*</str>
      <str name="version">2.2</str>
      <str name="rows">10</str>
    </lst>
  </lst>
  <result name="response" numFound="3235" start="0">
    <doc>
      <arr name="idA_o">
        <str>uniprotkb:P73045</str>
      </arr>
      <arr name="idB_o">
        <str>intact:EBI-1579103</str>
      </arr>
    </doc>
    <doc>
      <arr name="idA_o">
        <str>intact:EBI-1579109</str>
      </arr>
      <arr name="idB_o">
        <str>uniprotkb:P73824</str>
      </arr>
    </doc>
    <doc>
      <arr name="idA_o">
        <str>uniprotkb:Q59978</str>
      </arr>
      <arr name="idB_o">
```

```
<response>
  <lst name="responseHeader">
    <int name="status">0</int>
    <int name="QTime">18</int>
    <lst name="params">
      <str name="facet">>true</str>
      <str name="indent">on</str>
      <str name="start">0</str>
      <str name="q">*:*</str>
      <str name="facet.field">detmethod_s</str>
      <str name="version">2.2</str>
      <str name="rows">0</str>
    </lst>
  </lst>
  <result name="response" numFound="14" start="0">
    <lst name="facet_counts">
      <lst name="facet_queries">
        <lst name="facet_fields">
          <lst name="detmethod_s">
            <int name="psi-mi:MI:0107">8</int>
            <int name="psi-mi:MI:0007">4</int>
            <int name="psi-mi:MI:0114">2</int>
          </lst>
        </lst>
      </lst>
      <lst name="facet_dates">
      </lst>
      <lst name="facet_ranges">
      </lst>
    </response>
```

http://localhost:8983/solr/select/?q=%3A*&version=2.2&start=0&rows=0&indent=on&facet=true&facet.field=detmethod_s

Next steps

- Test indexing with larger files
- Update PSICQUIC webservice source code
(<https://psicquic.googlecode.com/svn/trunk/psicquic-solr>)
 - SOLR behind the scene
 - Support MITAB 2.6 and MITAB 2.7 + current formats
 - Update methods to allow faceting (group by)
- Update **PSICQUIC-view** and **IMEX-view**
- Export data in MITAB 2.7!