# AUG Email 2012-Jul

### Java Based Proxy for the i2b2 Web Client [07-27-2012]

**From:** Dat Q. Phan [Dat.Q.Phan.uth.tmc] **Sent:** Friday, July 27, 2012 4:39 PM **To:** McMurry, Andrew J.; Dan Connolly **Cc:** i2b2 AUG Members **Subject:** RE: Java Based Proxy for the i2b2 Webclient

I haven't had much time to implement this to Java but I used a work around which works for now.

I simply enabled my Java web application to run PHP using the following method:

http://php-java-bridge.sourceforge.net/doc/tomcat6.php

Placing the index.php file within the Java web app and modifying the webclient to point to the new path of the index.php resolved the issues I had with the web app timing out even when the user was actively using the webclient. Not the prettiest solution but it works for now until I have time to implement a more permanent solution which is most likely what Andy suggested.

I appreciate the response.

Thanks,

Johnny

**From:** McMurry, Andrew J. [Andrew_McMurry.hms.harvard] **Sent:** Monday, July 16, 2012 11:05 AM **To:** Dan Connolly **Cc:** Dat Q. Phan; i2b2 AUG Members **Subject:** Re: Java Based Proxy for the i2b2 Webclient

SHRINE uses a java based proxy, and it is a direct port of the PHP proxy.

I believe it still works for the i2b2 case as well. Feel free to extend or modify as you need.

Hope this helps,

andy

Softwarehttps://open.med.harvard.edu/display/SHRINE/Software

Source https://open.med.harvard.edu/svn/shrine/releases/1.11/code/proxy/

Tomcat WAR http://repo.open.med.harvard.edu/nexus/content/repositories/releases/net/shrine/shrine-proxy/1.11.1/

On Jul 16, 2012, at 11:40 AM, Dan Connolly wrote:

I don't have any direct experience to relay, but I wonder... Have you considered one of the existing PHP implementations for the JVM? http://stackoverflow.com/questions/3326261/is-there-a-php-implementation-that-targets-the-jvm-akin-to-jruby-and-jython On the other hand, the i2b2 PHP proxy is fairly simple, and it may well beless trouble to just translate it to Java. I think all it does is some whitelist/blacklistchecking and then proxy the request, using a curl library.

On Fri, 2012-06-08 at 16:52 +0000, Dat Q. Phan wrote:

Hi, We are running the i2b2 webclient v1.6.04 under a Java Web Application off of Tomcat. The problem is we need to convert the php based proxy file (index.php) to a java based proxy so we can run it under Tomcat along with our Java Web Application. By doing this, we are hoping to resolve our issues that we have with inactive timeouts. Has anyone done this and would like to share this with me? Thanks, Johnny PhanProgrammer Analyst UTHealthSchool of Biomedical Informatics Dat.Q.Phan.uth.tmc

### NLP and AUG Agendas [07-17-2012]

**From:** Churchill, Susanne E. **Sent:** Tuesday, July 17, 2012 2:45 PM **To:** i2b2 AUG Members **Subject:** NLP and AUG Agendas

Dear Colleagues,

The mostly final agendas for the upcoming meetings have just been posted to www.i2b2.org/work/aug.html. Poster titles are subject to author discretion. We'll have very good participation, current registration stands at 88.

If there are any further registrations or poster submissions, please let me know. Otherwise, see you all in Boston next week. We're currently in the grips of a severe heat wave so please plan accordingly (all meeting venues thoroughly ac'd). But the good news is that lobster prices are on the floor.

Let me know if you have any questions.

Regards,

Susanne

## Query-in-Query of Group 1 intersect Group 2 gave results bigger than A and B [07-17-2012]

**From:** dwhend0.gmail [dwhend0.gmail] **On Behalf Of** Darren W Henderson **Sent:** Tuesday, July 17, 2012 11:34 AM **To:** Dan Connolly **Cc:** i2b2 AUG Members; Tamara McMahon **Subject:** Re: Query-in-Query of Group 1 intersect Group 2 gave results bigger than A and B

Dan, I read your description of the problem and read through the generated SQL. I think the functionality that you or your users are expecting is closer to what is provided by placing the patient result set from a previous query into a group panel, as opposed to actually bringing in the query. When you place the patient result set into the i2b2 panels you would get the intersection between the two sets, these 54 people you're looking for in your example. The added benefit is that patient result sets are actually stored in a collection table, so this query would run much faster than rerunning two queries independently.

When you drag a previous query into a panel, it's like placing all the panels that were in the original query into a panel as a container. When you drag the 2nd query into the next panel it's adding all those panels to the query as well. Additionally it would seem that these new repeats of the queries inherit the timing of the new query, instead of maintaining their previous timing. So in your example, your number of patients may be off from what you expected because the query timing was left as 'ANY' in the run with the queries in each panel. I can confirm that the new query was any timing with both the XML and generated sql you sent along, but you'll need to go check if either query with query_master_id equal to 7577 or 6704 were same visit queries. If they were, then that's why you're number is so different from what you expect.

Additionally, after digging into the generated SQL that you sent along, I think your example actually discovered some minor inefficiencies and bugs in the query generator for this type of query. For example, within the batch of transactions the result set is inserted into the #DX temporary table, the data goes unused as the next transaction simply pulls again from the QUERY_GLOBAL_TEMP , and then the #DX table is deleted from with no predicate:

insert into BlueHerondata.DX ( patient_num ) select * from ( select distinct patient_num from BlueHerondata.QUERY_GLOBAL_TEMP where panel_count = 7 ) q
;
insert into BlueHerondata.MASTER_QUERY_GLOBAL_TEMP(master_id, patient_num, level_no) select 'masterid:7577', patient_num, 0 from BlueHerondata.QUERY_GLOBAL_TEMP where patient_num IN ( select patient_num from BlueHerondata.QUERY_GLOBAL_TEMP where panel_count = 7 )
;
delete BlueHerondata.DX

On a larger dataset this insert and delete could take a signficant amount of time, and requires additional temporary db space for the table, and the insert and delete transactions require log space as well, and this is a waste of I/O time on the server.

The final steps of this query are also confusing:

insert into BlueHerondata.QUERY_GLOBAL_TEMP ( patient_num , panel_count)
select patient_num ,1 as panel_count from (select j1.patient_num from BlueHerondata.observation_fact j1, BlueHerondata. MASTER_QUERY_GLOBAL_TEMP j2 where j1.patient_num = j2.patient_num group by j1.patient_num ) t
;
update BlueHerondata.QUERY_GLOBAL_TEMP set panel_count =2 where exists ( select 1 as panel_count from (select j1.patient_num from BlueHerondata.observation_fact j1, BlueHerondata.MASTER_QUERY_GLOBAL_TEMP j2 where j1.patient_num = j2.patient_num group by j1. patient_num ) t where BlueHerondata.QUERY_GLOBAL_TEMP.panel_count = 1 and BlueHerondata.QUERY_GLOBAL_TEMP.patient_num = t. patient_num )
;
insert into BlueHerondata.DX ( patient_num ) select * from ( select distinct patient_num from BlueHerondata.QUERY_GLOBAL_TEMP where panel_count = 2 ) q

At this point in the query, the MASTER_ table contains patient_num and a masterid:6704 or masterid:7577 identifier in each row. The query then uses a join to the observation_fact table in order to produce a distinct list of patient_nums. It does this twice actually, but without any kind of predicate so it's effectively the same thing as a select distinct patient_num from the MASTER_ table would be. For instances with very large observation_fact tables this may be pretty slow just to perform an aggregation of patient_nums. It's not needed to ensure data integrity either, you know where the patient_nums came from, no need to check them like a constraint at this point. Also, this kind of breaks the logic, and makes this an OR between the two master sets that represent previous queries. If an AND is the target here, you really want a distinct list of patients having a count of distinct masterid sets = 2 like:

select patient_num
from MASTER_QUERY_GLOBAL_TEMP
group by patient_num
having count(distinct master_id) = 2
or something similar.

Darren W. Henderson
University of Kentucky
Division for Biomedical Informatics
Bio-Pharmacy Complex Rm182
789 S Limestone
Lexington, KY 40536

On Mon, Jul 16, 2012 at 1:17 PM, Dan Connolly <dconnolly.kumc> wrote:
The new Query-in-Query feature of i2b2 1.6 looks quite interesting, but we haven't managed to get it to work.If anyone else has related experience to share, I'd appreciate it. In particular, our clinical informatics coordinator, Tamara, reported this problem (#1257) on behalf of one of the investigators she is working with:
I searched the following. Group 1: Kluding DPN 40-70 Front.@20:24:10[7-12-2012][jdenton] - which returns 337 patients Group 2: Robbins-DM II - Couples-exclu.@13:15:33 [5-17-2012] [jdenton] - which returns 505 patients Result should be 54 patients included in both searches, but HERON returns 824. i2b2 should and the 2 groups but not sure what it does.
I thought perhaps 824 was a typo for 842, the sum of 337 and 505, but I double-checked qt_query_result_instance.real_set_size, and it's really 824. I'm attaching the I2B2_REQUEST_XML and the GENERATED_SQL from the qt_query_master record from when that query was run.I took a look, but got overwhelmed quickly. If anyone can help diagnose the problem, I'd sure appreciate it!
Dan Connolly

## Plugin dev – i2b2.sdx.Master [07-16-2012]

From: Brian Ostasiewski [bostasie.wakehealth] Sent: Monday, July 16, 2012 1:40 PMTo: Dan ConnollyCc: i2b2 AUG MembersSubject: RE: Plugin dev - i2b2.sdx.Master

This actually got resolved, I believe it was an access to the prsRecord when the prsRecord was null that caused the javascript to stop, it just happened to be at the end of that bit of code so it was not a visibly interrupting error and there was no indication in the console, but it kept the ajax call from going through.

An example of what I ran into is line 59 of the Dem1Set example: prsRecord.sdxInfo.sdxKeyValue is accessed and prsRecord may in fact be null if you were to make it optional and no element was dropped in.

So it's nothing inherent in the system, just be careful how/where you reuse code 😄


From: Dan Connolly [dconnolly.kumc]
Sent: Monday, July 16, 2012 11:57 AM
To: Brian Ostasiewski
Cc: i2b2 AUG Members
Subject: Re: Plugin dev - i2b2.sdx.Master

Dan Connolly, KUMC Medical Informatics<http://informatics.kumc.edu/>
913-945-6741


On Mon, 2012-06-04 at 18:50 +0000, Brian Ostasiewski wrote:
In order to support drag-n-drop concept fields, if I use the typical sequence of registration:
i2b2.sdx.Master.AttachType(...);
i2b2.sdx.Master.setHandlerCustom(..., "DropHandler", ...);
It seems to make it REQUIRED that the drop fields be populated (I wish to keep one optional) before I can execute an ajax call with a function I've defined by:
i2b2.<myplugin>.ajax._addFunctionCall(...);

Interesting... when we developed a plug-in, drop fields seemed to be optional; our use case actually required data there, and our initial attempt was buggy until we added an explicit check.http://informatics.kumc.edu/work/ticket/847

I don't see the mechanism that governs this in the js-i2b2 code, can anyone confirm/explain?

I believe the mechanism is constraints on the model.
For example, in the timeline widget, the callback for changing tabs (and hence consuming data) includes a check to see if any concepts have been dropped:
if ((i2b2.Timeline.model.concepts.length>0) && i2b2.Timeline.model.prsRecord) {
This is satisfied in i2b2.Timeline.conceptDropped by:
i2b2.Timeline.model.concepts.push(sdxData);

Is there a way to override this requirement?
Also, I could get around this if there was a way to revert the AttachType and setHandlerCustom calls - can I repeat those calls with some null value to reverse my original specification?

Thanks,
Brian


## July Meeting [07-13-2012]

**From:** Russ Waitman [rwaitman.kumc] **Sent:** Friday, July 13, 2012 11:42 AM **To:** i2b2 AUG Members; Churchill, Susanne E. **Cc:** Nathan Graham; Tamara McMahon **Subject:** Re: July Meeting
Susanne,
I think we forgot to give you our poster title.  It will be:
The University of Kansas' HERON takes flight: FY2012 enhancements include support for NCI cancer center designation (tumor registry, biospecimens, death index, and R integration), UHC hospital quality measures technical diagnoses, enhanced Medications, and REDCap Registry Integration.
Tamara M. McMahon, Nathan J. Graham, Daniel W. Connolly, Brennan W. Connolly , John Keighley, Mani Nair, Bhargav Adagarla, Arvinder Choudary, Lemuel R. Waitman
Division of Medical Informatics, Department of Biostatistics, University of Kansas Medical Center, Kansas City, KS
Russ WaitmanDirector of Medical InformaticsAssociate Professor, Department of BiostatisticsAssistant Vice Chancellor for Enterprise AnalyticsUniversity of Kansas Medical Center913-945-7087 rwaitman.kumc "Churchill, Susanne E." <SCHURCHILL.PARTNERS> 6/22/2012 12:30 PM
Colleagues,
For those of you bringing posters (we strongly encourage you to do so):
1)  Please limit your size to 4' x 4' so we can get two on each mounting board.2)  Please send you poster title to me.
We're adding new items to the agenda so keep an eye on it at www.i2b2.org/work/aug.html.
Thanks!
Susanne


## Open source standard medication ontology being used by SHRINE [07-05-2012]

**From:** Key, Dustin [key.d.ghc] **Sent:** Thursday, July 05, 2012 7:41 PM **To:** 'Russ Waitman'; Andrew J. McMurry **Cc:** i2b2 AUG Members; Dan Connolly;

Nathan Graham **Subject:** RE: Is there an open source "standard" medication ontology that is being used for SHRINE and is it the i2b2 demo ontology?
Russ,

We built the Rx ontology with UMLS, mapping NDC to Rxcui and it works for our purposes.  I think our implementation is probably not as good as what is discussed in the article, which I would like to emulate some day.  There is strong evidence of elided drugs in our case.  I think this is due to both imperfect method, and perhaps the dropping of obsolete drugs from the source vocabs (I've seen this happen in the case of CPT.)    I was learning about UMLS and drugs as I went and haven't had a chance to go back and clean up the logic.  I'd love to hear / learn more about the right build method and the potential of dropped codes in UMLS.    Perhaps a team effort of setting up and documenting the right build approach / script would be of use here?

We use NDF-RT, VANDF, and MRREL, and populate concept_dimension and observation_fact with rxcuis.  (VANDF seems to be a new and necessary UMLS component of the build, that has shown up since the article was published.)  We link the concept dimension to the i2b2 table via a path name which is a string of umls MRHIER atoms and rxnorm concept codes.  The c_names presented in the ontology are the SCDFs and [brand names] for the non generics.  I'd be happy to share more of  our experience, code, ontology, etc. if it is helpful.

Dustin


**Dustin Key** | PROGRAMMER ANALYST
Group Health Research Institute
PHONE 206-287-2916| CDS 220-2916E-MAIL key.d.ghc
www.grouphealthresearch.org

**From:** Russ Waitman [rwaitman.kumc] **Sent:** Thursday, June 28, 2012 10:38 AM **To:** Andrew J. McMurry **Cc:** i2b2 AUG Members; Dan Connolly; Nathan Graham **Subject:** Re: Is there an open source "standard" medication ontology that is being used for SHRINE and is it the i2b2 demo ontology?
Hi Andrew,
Thanks for the prompt reply.  We are aware of that article and contacted Matvey who informed us that the resulting ontologies were not freely available.
We will have it covered from getting Epic meds mapped to GCNSEQNO or NDC and then getting to RxNorm CUIs.  We're tracking our work on this topic in our ticket here *http://informatics.kumc.edu/work/ticket/1048* and your free to use our stuff when its done.
What we want is once we're there, which ontology should we use for Meds that uses those CUI codes?
Is it possible to get a copy of that ontology from you, NCBO, someone else in i2b2 compatible format (aka insert into CONCEPT_DIMENSION)?
Our sense is we would want to have our folders line up to be really cross institutional compatible.  Otherwise, you have to build your queries down at the concept code level instead of leveraging higher level paths.  We can recreate things ourself now that NDF-RT is in the UMLS with RxNorm but even then, I doubt we'd get things to line up exactly right with what you have done.
Is the i2b2 demo data "Medications" based on that standard ontology or something else?
That demo data seems to be mapped down at the NDC level and not using RxNorm CUIs.
I would think the NDC would be too low a level and that most researchers would want to operate a level up in the RxNorm hierarchy.  Ideally pill size should be handled in a dose modifier so you don't have to join all the pill sizes together.
Secondarily, does the keeper of the ontology have update scripts somewhere that run against RxNorm and NDF-RT that handle the weekly or monthly updates as new NDCs and drugs come on market?  Or, a way to just pickup the new concept_dimensions in a zip file or something?
Russ

"McMurry, Andrew J." <Andrew_McMurry.hms.harvard> 6/28/2012 12:01 PM We have mapped several Medication lists to RxNorm at the Harvard hospitals.
From there, we used the National Drug File medication tree created by the VA (called NDF-RT).
This paper provides a useful description http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041416
I will check if this process has been done for EPIC specifically.
Zak just suggested to me that we try this for EPIC if we have not done so already.
Cheers,
andy

On Jun 28, 2012, at 9:55 AM, Russ Waitman wrote:
Hi,
We are now re-doing our Medications so that in addition to tagging our medication observations with the local terminology (Epic medication_id), we are going to map things to RxNorm.
We're motivated to do this because the local Epic ontology we harvest from Clarity is leaving a fair number of observations unmapped and also because we want to enter this world of interoperability.
Our approach to mapping is if we get hit on GCNSEQNO  we use it, elseif we have NDC we use it.  Our fallback in then going to be using MedEX for an NLP lookup to RxNorm http://knowledgemap.mc.vanderbilt.edu/research/content/medex-tool-finding-medication-information
**Questions:**
1.  The i2b2 demo ontology is at the NDC level.  Is that what we should use and do a look up from our RxNorm CUI to a representative NDC?  Or, has someone developed another RxNorm CUI tree?
2.  Do other places find that the i2b2 demo ontology covers most everything that people use in your medical center?
3.  Do all the UCalifornia and Harvard sites run on this medication ontology?
If that's the case we might favor the NDC mapping over the GCNSEQNO.
Otherwise, we'd then need to create an ontology similar to what Recombinant did and now that NDF-RT is native in the UMLS this should be more straightforward.  Or, we could see if NCBO has a downloadable i2b2 compatible ontology.  Or, we could investigate things like this 3M announcement:  http://www.hddaccess.com/
We want to go the right way because we think if we bake our own ontology that may be problematic if the goal is to have search across institutions and have it work on concept paths as opposed to concept codes.  Otherwise, you can share data but the whole ontology concept path thing won't work unless we harmonize our paths across sites.
For example our local ontology has Medications\Anticoagulants\Heparins  and Medications\Anticoagulants\Low Molecular Weight Heparins which are two concept paths likely unmeaningful to anyone else using i2b2.
Thank you for any guidance you can provide,
Russ Waitman
Associate Professor, Director of Medical Informatics
Department of Biostatistics
Assistant Vice Chancellor for Enterprise Analytics
University of Kansas Medical Center
913-945-7087

rwaitman.kumc http://informatics.kumc.edu