

AUG Email 2011-Jan

Last Updated: 01/31/2011

NEXT EMAIL

From: Dan Connolly Sent: Monday, January 31, 2011 3:47 PM To: i2b2 AUG Members

Subject: Re: CRC NullPointerException near CacheUtil: anyone else seen this?

Wild... Turned out to be a bug in a JDK that I accidentally installed, which caused JBoss to fail to create the cache service.

For reference: <http://informatics.kumc.edu/work/ticket/308>

On Mon, 2011-01-31 at 18:07 +0000, Dan Connolly wrote:

The CRC cell in one of our i2b2 1.4 installations seems to have lost its marbles; It gives us two "Press F12..." messages during startup and fails to run queries. It seems that all calls to the CRC cell are failing.

Has anyone else seen symptoms like this?

A complete message exchange is attached. The top of the stack trace from the CRC_QRY_getQueryMasterList_fromUserId response is:

```
<response_header> <info>Log information</info> <result_status>
```

```
<status type="ERROR">java.lang.NullPointerException at edu.harvard.i2b2.crc.delegate.setfinder.QueryRequestDelegate.handleRequest
```

```
(QueryRequestDelegate.java:126) at edu.harvard.i2b2.crc.axis2.QueryService.handleRequest(QueryService.java:129) at edu.harvard.i2b2.crc.axis2.
```

```
QueryService.request(QueryService.java:67) at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
```

```
Line 126 of QueryRequestDelegate.java reads:
```

```
Node rootNode = CacheUtil.getCache().getRoot();
```

Dan

NEXT EMAIL

From: Wilson, Brian Sent: Friday, January 28, 2011 2:28 PM Cc: i2b2 AUG Members

Subject: Genomic Data & i2b2 - Meeting Follow up

Hi All,

The following is a link to a i2b2 community wiki page containing the presentation from this morning. <https://community.i2b2.org/wiki/display/GARLIC/GARLIC++AUG+FeedBack>

May I suggest using this page to continue the discussion and help capture and disseminate other ideas, comments and questions the community may have with respect to this topic.

Thank you

Brian

NEXT EMAIL

From: Phillips, Lori C. Sent: Friday, January 28, 2011 8:39 AM To: Wei,Xintao; i2b2 AUG Members Subject: RE: I2B2 installation problem

Xintao,

Try

```
edu.harvard.i2b2.crc.applicationdir=C:\jboss-4.2.2.GA\server\default
```

```
conf
```

```
c rcapp
```

A singular '\' is interpreted as escape char.

Lori Phillips

From: Wei,Xintao Sent: Friday, January 28, 2011 8:32 AM To: i2b2 AUG Members

Subject: I2B2 installation problem

Hi, I am installing I2B2 server 1.5.2 to Windows Server 2008 with SQL server 2005.

I have problems to install I2B2 cells.

For example, my crc_application_directory.properties was edited as:

```
edu.harvard.i2b2.crc.applicationdir=C:\jboss-4.2.2.GA\server\default\conf\crcapp
```

but the error said that:

```
ERROR [STDERR] edu.harvard.i2b2.common.exception.I2B2Exception: Application property file(C:\jboss-4.2.2.GAserverdefaultconfcrcapp\crc.properties) missing entries or not loaded properly
```

The directory is wrong! Folder C:\jboss-4.2.2.GAserverdefaultconfcrcapp\ was created by I2B2 scripts and crc.properties was in this folder.

Based on what I understood, folder C:\jboss-4.2.2.GA\server\default\conf\crcapp should be created, but not C:\jboss-4.2.2.GAserverdefaultconfcrcapp.

Somehow the "\" had been removed by I2B2 scripts.

It happened to other cells too. All "\" had been removed from *.properties files and the following folders were created:

```
C:\jboss-4.2.2.GAserverdefaultconfcrcapp C:\jboss-4.2.2.GAserverdefaultconfrcloaderapp C:\jboss-4.2.2.GAserverdefaultconfontologyapp C:\jboss-4.2.2.
```

```
GAserverdefaultconfworkplaceapp
```

How to solve this problem? Any suggestions?

Thank you!

Xintao

NEXT EMAIL

From: Murphy, Shawn N. Sent: Wednesday, January 26, 2011 1:06 PM To: Peter Beninato; Dan Connolly; i2b2 AUG Members

Subject: RE: what is INDEX FACT_NOLOB ON OBSERVATION_FACT for?

The WHERE clause can look very simple, like a range of patient_num's, but the SELECT statement is forcing the retrieval. SQL Server 2008 has a special way to handle this by allowing the inclusion of "non-query" columns in the index, but other databases do not so you need to use this trick to get fast retrieval times.

Thanks, Shawn.

From: Peter Beninato Sent: Wednesday, January 26, 2011 12:44 PM To: Murphy, Shawn N.; Dan Connolly; i2b2 AUG Members

Subject: RE: what is INDEX FACT_NOLOB ON OBSERVATION_FACT for?

Hi,

What does the WHERE clause look like for this type of PDO retrieval?

Are there criteria on all these columns?

Peter

From: Murphy, Shawn N. Sent: Wednesday, January 26, 2011 9:41 AM To: Dan Connolly; i2b2 AUG Members

Subject: RE: what is INDEX FACT_NOLOB ON OBSERVATION_FACT for?

Hi Dan,

We use this for PDO retrievals. If one does not have an index that includes all the items one uses in a select statement, the index that is used will force the retrieval mechanism to go back to the original database to retrieve the items for the query (since it can not retrieve them from the index). This makes the data retrieval very slow.

Thanks,

Shawn.

From: Dan Connolly Sent: Wednesday, January 26, 2011 11:59 AM To: i2b2 AUG Members

Subject: what is INDEX FACT_NOLOB ON OBSERVATION_FACT for?

I'm bulk loading on the order of 10^7 records, so I disabled the indexes first, then used sqldr, and now I'm rebuilding the indexes. Rebuilding FACT_NOLOB is taking a lot of space and time. I'm curious what it's for. What would happen if I didn't rebuild it?

For reference... it seems to come from this bit of i2b2/edu.harvard.i2b2.data/Release_1-4/NewInstall/Demodata/scripts/crc_create_datamart_oracle.sql

```
CREATE INDEX FACT_NOLOB ON OBSERVATION_FACT (PATIENT_NUM, START_DATE, CONCEPT_CD, ENCOUNTER_NUM, NVAL_NUM,
TVAL_CHAR, VALTYPE_CD, MODIFIER_CD, VALUEFLAG_CD, PROVIDER_ID, QUANTITY_NUM, UNITS_CD, END_DATE, LOCATION_CD,
CONFIDENCE_NUM, UPDATE_DATE, DOWNLOAD_DATE, IMPORT_DATE, SOURCESYSTEM_CD, UPLOAD_ID);
```

p.s. Is that crc_create_datamart_oracle.sql file in the SVN repository somewhere? I don't see it under <http://svn.i2b2.org/svn/repos/i2b2/trunk/>

NEXT EMAIL

From: Phillips, Lori C. Sent: Tuesday, January 25, 2011 10:32 AM To: Zapletal Eric; i2b2 AUG MembersCc: Nicolas Rodon

Subject: RE: strange behaviour in Workbench with characters

Eric, When you drag from Workplace (action red 1) are you dropping it to the query name box or into the group 1 panel? Lori Phillips

From: Zapletal Eric Sent: Tuesday, January 25, 2011 10:21 AM To: i2b2 AUG MembersCc: Nicolas Rodon

Subject: strange behaviour in Workbench with characters

Dear I2B2 community, we are currently facing a strange behaviour in the I2B2 Workbench (1.5.2 version) :

when we drag/drop a request from Workplace to QueryTool (red 1 action), accented characters (E ACCUTE) are not correctly rendered in the QueryTool only. Accented characters are however correctly displayed in the Workplace window.

when we drag/drop a concept from Ontology Tools (green 2 action), concepts are correctly displayed in QueryTool (cf enclosed image : in this example the same concept appears twice in the Query Tool) Furthermore, requests can not be executed when such incorrect characters appear in the QueryTool. Is this behaviour already mark as a bug or the result of a bad installation ? We are currently running version 1.3 of the server... Many thanks in advance for your help ! Sincerely yours Eric Zapletal

NEXT EMAIL

From: Jeff Cowall Sent: Thursday, January 13, 2011 8:38 AM To: i2b2 AUG Members; James Law; Murphy, Shawn N.

Subject: RE: representation of Cancer Registry data as i2b2 observations

Is the modifier UI anticipated for 1.6 release, though? Or later? I'd be very interested in a rough idea of approach and timing - Extend the concept tree or what? Might be worth waiting and planning for.

No commitments, I understand, just thinking....

Thanks, Jeff

On 1/12/2011 at 7:44 PM, "Murphy, Shawn N." wrote:

Hi Jeff and James,

I guess that is true, that the grouping is similar to grouping these kinds of observations in an encounter, and this method will work. What worries me about it is that the false positive still are possible if the user does not specify the same encounter restriction, but this could be overcome by some creative UI changes.

Unfortunately the UI for the modifiers is not ready in 1.6RC's yet, just the data model. We have been focusing on query performance, and that will be the emphasis of RC3 that should be out in the beginning of February.

Thanks, Shawn.

From: Jeff Cowall Sent: Wednesday, January 12, 2011 10:43 AM To: i2b2 AUG Members; James Law; Murphy, Shawn N.

Subject: RE: representation of Cancer Registry data as i2b2 observations

I'm working with James on this, and the data we're interested in is ICD-O oncology diagnoses prepared by Data Abstractors in the Cancer Registry. They refer to the act of determining the proper code as a "diagnostic event", and it is not associated with a particular, specific physical visit by a patient. Each tumor or fluid-born cancer identified receives a unique identifier, too, and that is what gets the specific diagnosis, there can be more than one at a time.

It seems to me that this is more than somewhat analogous to an encounter: The abstractor ("provider") observed this component of the complex ICD-O ("observation"), of a unique tumor/liquid ("patient" component). It's tumor(s) that change over time, and that association needs to be maintained.

If there were a way of associating observations to a specific tumor on a specific patient, that'd work, I think, as well as being conceptually more accurate. Thoughts, anyone?

Thanks, Jeff

Jeff Cowall Sr Data Architect University of Michigan Health System MCIT Clinical Research

On 1/11/2011 at 8:54 PM, "James Law" wrote:

Shawn- Thanks for the quick follow up.

Yes, this is the false positive. If we used the visit dimension as a "fact grouper", the 1.6 UI would allow the end user to remove the false positives by using the "same visit" in each column- am I correct? Sounds like you're not advising we go this path, however, as we would lose the financial visit. (It might be okay for us to not associate these with a financial encounter/visit)

One question on the modifiers. I've read the wiki page on the design of it. While I'm not completely clear on how it works yet, I did see the .png examples attached to the page, which show the ability of users to modify dose amount, and route, and other attributes of a medication. These additional attributes are stored as additional rows in ob_fact. Is the case of meds, is there no change to the ontology tree?

A related question. We've got the 1.62 rc candidate VM running. Is there an example of a med where the modifiers are working?
Thanks again!
James

"Murphy, Shawn N." 1/11/2011 8:34 PM

Hi James,
Just to clarify, I think what you are saying is that if a person had malignant brain cancer and benign skin cancer, the query brain AND benign would return a false positive for this patient.
In 1.6 the use of modifiers should be able to take care of this use case. You could use the visit dimension as a "fact grouper" as well, but you would have to modify the UI, and the table would not be available for visits of course.
Thanks, Shawn.

From: James Law Sent: Tuesday, January 11, 2011 4:48 PM To: i2b2 AUG Members

Subject: representation of Cancer Registry data as i2b2 observations

Hi, We are loading some of our organizations' Cancer Registry data into i2b2 for cohort identification. One element of interest from the registry is the the icd-o code, which is maintained by our registrars. As you may know, the icd-o is a code, but is a composite for the cancer's site, morphology, and behavior. We created an ontology such that a PI could query using site, morphology, or behavior. Thus our ontology tree looks something like:
Registry - Site -- ~~BRAIN - LUNG - ETC~~ Morphology -- ~~8000-8100 - neoplasm, benign~~ -- 8100-8200 -etc
For each ICD-O code, we record multiple facts in the OB_FACT table. One for the site, one for the morphology, one for behavior.
This works well, but creates the potential for a false positive. For example, if a patient had multiple cancers: LUNG /SMALL cell and SKIN /Melanoma
Then a query that contains both LUNG + Melanoma would actually return this patient, even though that is probably not what a PI would expect.
What is the best way to resolve this problem? One thought we've had is to have these observations for the cancer be under a specified visit in the visit dimension. The 1.6 query UI will allow a user to specify they are on the same "visit". Would this be considered a valid use of the visit_dimension table?
Using the new concept code modifier seems like a good solution, but they don't seem appropriate if its something that needs to be in the Ontology tree.
Thanks, James
James Law University of Michigan Health System | MCIT

NEXT EMAIL

From: Nigam Shah Sent: Sunday, January 09, 2011 4:29 PM To: nlp-sig at amia; kr-sig at amia; i2b2 AUG Members

Subject: JBI special issue on "Mining the Pharmacogenomics Literature" – Call for papers

Dear Colleagues,

WHAT: JBI special issue on mining the pharmacogenomics literature Guest editors: K. Bretonnel Cohen, Russ Altman, Yael Garten, Udo Hahn, Nigam Shah

DEADLINE: March 15th 2011

We are inviting submissions for a special issue of the Journal of Biomedical Informatics on the automatic or semi-automatic extraction of relationships between biomedical entities relevant to pharmacogenomics from the research literature. Accepted papers will focus particularly on methods for the extraction of genotype-phenotype, genotype-drug, and phenotype-drug relationships and the novel use of these relationships for advancing pharmacogenomic research. Efforts aimed at creating benchmark corpora as well as comparative evaluation of existing relationship extraction methods are of special interest.

The planned special issue aims to address the gap in coverage of text mining for pharmacogenomics, as an important initial application area of genomics in clinical medicine, and thus an important translational medicine activity. The technical area of the issue is intended to focus particularly on genotype-phenotype-drug relationships. It will include broad categories of work that have been well-studied in the past, specifically text mining and reasoning, but will restrict submissions to applications of that work to the constrained area of pharmacogenomics, and particularly genotype-phenotype-drug relationships.

For example, topics that are solicited include:

Relation extraction between genotypes, phenotypes, and drugs, and other semantic classes relevant to pharmacogenomics

Corpus development for pharmacogenomics text mining

Associating gene variants (mutations, alleles, rs/ss numbers) to the associated gene name

Using text mining to extract information about the association of drugs with clinical phenotypes

The use of biological networks in combination with text mining to facilitate discovery

Work on the corpus of documents linked to by PharmGKB

Reasoning systems applied over the PharmGKB knowledge base

The creation of ontologies to help relate molecular action of drugs to their clinical effects.

Work on named entity recognition (e.g., gene taggers) alone will not be considered responsive to this call. The key feature we seek in submissions is the use of language technologies to understand the molecular basis of drug response, its variability, and its impact on phenotypes at the molecular, cellular, organ and whole organisms level. Approaches that combine text-mining and knowledge-based systems are of special interest.

See full details in the attached CFP.

Regards, Nigam Shah

NEXT EMAIL

From: Russ Waitman Sent: Wednesday, January 05, 2011 6:13 PM To: Keith Marsolo Cc: i2b2 AUG Members; Arvinder Choudhary; Dan Connolly; Dongsheng Zhu

Subject: Re: Representing multiple sources of a diagnosis observation

Keith, That's great to hear. It seems we're the opposite of you:

KUH/UKP hasn't yet migrated to Epic for billing so our billing records reside in Seimens for the hospital and IDX for the clinics. - our Epic rollout has occurred in the hospital first so that's where all our data is largely coming from. The Epic clinic rollout started in Summer of 2010.

Russ

Keith Marsolo 1/5/2011 3:38 PM

Russ,

We pull encounters from pat_enc_dx and problems from problem_list. Our discharge records are located in HSP_ACC_LIST_DX. We don't have many records in HSP_DISCH_DX. Most of our billing records come off an interface, so that might be why they are located in a different table.

We also have plans to load the diagnoses in the MEDICAL_HX table. Those would be loaded with a type of "Medical History." We plan to do the same thing with the data in SURGICAL_HX, except those would go in the procedure portion of the ontology.

Going back to billing diagnoses, I would caution against trying to figure out which ones are listed on the discharge summary, at least if your institution's Epic build is anything like ours. The discharge summary diagnoses are a subset of the billing diagnoses, but we couldn't find any way to distinguish which ones were which.

We haven't tried to tackle inpatient yet.

Keith

On Jan 4, 2011, at 7:07 PM, Russ Waitman wrote:

Thanks for the reply Keith. That's interesting.

So from your example below, somewhere you'd be storing the ICD9 in the observation value slot but this xml would tell us about the source of the problem and is linked off the modifier. That's the underlying approach when you mention xml filtering in this talk:

http://www.i2b2.org/events/slides/i2b2_AUG_201010_Marsolo.pptx on slide 9.

So from that slide, am I to understand you are pulling the encounter dx from pat_enc_dx? discharge dx from hsp_disch_diag? and the problem list from problem_list?

In our environment they aren't coding discharge diagnoses but are the hsp_admit_diag table. The medical_hx table also seems to have a bunch of diagnosis info here,

Russ

Keith Marsolo 1/4/2011 9:42 AM

Russ,

We took an approach based on modifiers, though ours uses XML instead of the more "flat" ontology used in version 1.6. In fact, trying to nest diagnoses (i.e. multiple types/sources) under one concept isn't possible with modifiers as implemented in version 1.6. If you look at any of my AUG presentations, this is what I call "XML-based filtering."

I'll paste a snippet of one of our early versions of this. I'm not saying we have the perfect solution. In fact, we're still trying to figure out the best format to use. We've looked at XML, XML schemas and now we're exploring RDF, which should potentially allow us to query based on attribute.

Without going into too much detail, this is useful for some of our outcomes work where the "drilldown" attributes of interest have different names, a potentially different set of values, but are still related and need to be linked somehow.

In any case, here's a snippet of the XML, as it would appear in the ontology metadata (my apologies for the long string of code). There's a corresponding blob of XML that we store in the observation_blob column (not shown). You can set the values in our query tool ("constrain by XML" context menu) and then the appropriate facts are retrieved.

If you were interested in this method, our code is available, though we're not in a position to offer any support (such as life in the world of academic software development).

Keith

```
<ontology_xml> <context>
<val>diagnoses</val> </context> <xml_type>
<val>2</val> </xml_type> <type Name="ProblemList">
<chronic> <val>N</val> <val>Y</val>
</chronic>
<class_of_problem> <val>Stage 2</val> <val>Acute</val> <val>Chronic</val> <val>Temporary</val> <val>End Stage</val> <val>Stage 1</val>
<val>Minor</val>
</class_of_problem>
<hospital_pl> <val>N</val> <val>Y</val> </hospital_pl>
<priority> <val>Low</val> <val>Medium</val> <val>High</val>
</priority>
<status> <val>Active</val> <val>Deleted</val> <val>Resolved</val>
</status> </type> <type Name="EncounterDx">
<primary> <val>N</val> <val>Y</val>
</primary>
<class_of_problem> <val>Active</val> <val>Inactive</val>
<val>Acute</val> <val>Chronic</val> <val>Temporary</val>
</class_of_problem>
<chronic> <val>N</val> <val>Y</val>
</chronic> </type> <type Name="DischargeDx">
<admit_dx> <val>N</val> <val>Y</val>
</admit_dx> <final_admit_dx> <val>Yes</val> <val>No</val> <val>Unknown</val> <val>Clinically Undetermined</val> <val>Exempt from POA reporting<
/val> </final_admit_dx> <final_dx_exclude>
<val>yes</val> <val>no</val> </final_dx_exclude> </type> </ontology_xml>
```

On Jan 4, 2011, at 10:13 AM, Russ Waitman wrote:

Hi, This may be a fine point but we're wondering how other places are handling the loading of "diagnosis" observations that come from several clinical or billing activities. We are first migrating data in from several places in our Epic's EMR but will then be adding diagnoses from our outpatient IDX billing system.

For example, within Epic's Clarity database there are several tables of interest:

PAT_ENC_DX has diagnoses associated with an encounter

PROBLEM_LIST has diagnoses assigned to a patient's problem list

MEDICAL_HX diagnoses from medical history contacts

HSP_ADMIT_DIAG inpatient admission diagnoses

more details here <http://informatics.kumc.edu/work/wiki/DiagnosisMapping>

Coming from our billing system, we have multiple ICD9 codes assigned to a visit.

On one hand, a researcher might want to see all these diagnoses loaded into a single "diagnosis" ontology tree so they can find any occurrence of a diagnosis.

But, other people may have questions which are strictly focused on inpatient diagnoses entered by a provider.

Are people tending towards one diagnosis bucket or splitting by source? Or, is this a place in 1.6 where a modifier would be utilized?

Russ Waitman Associate Professor Director of Medical Informatics Department of Biostatistics University of Kansas Medical Center

NEXT EMAIL

From: Peter Beninato Sent: Tuesday, January 04, 2011 12:57 PM To: Wei,Xintao; i2b2 AUG Members

Subject: RE: I2B2 installation on windows server 2008

Hi,

Have you installed Apache yet?

The httpd directory is where the apache webserver resides.

From: Wei,Xintao Sent: Monday, January 03, 2011 8:41 AM To: i2b2 AUG Members

Subject: I2B2 installation on windows server 2008

Hi, I am trying to install I2B2 server 1.5.2 to windows server 2008. I have some troubles to the I2B2 server because the I2B2 installation guide is for Linux. Where should I copy the admin directory to the "httpd directory" in windows?

Thank you!

Xintao

NEXT EMAIL

From: Dan Connolly Sent: Monday, January 03, 2011 10:02 AM To: i2b2 AUG Members

Subject: basic stats to aid concept hierarchy navigation

This is the first part of an article in our project blog; see the full article for technical details...

Concept stats: how many needles are in which parts of our i2b2 haystack?

For the HERON research data repository we're building, we're using I2B2, which involves navigating a huge vocabulary of medical terminology. We found ourselves wishing that the terms were tagged with some clues about how often they occur in our database. Now that the enhancement ticket is done, we find it really does help.

I2B2 lets researchers identify patient cohorts by querying a database of "observation facts" about patients. All the facts are collected in one table, indexed by concept. Concepts may be demographics (age, gender, etc.), diagnoses, labs, medications, or procedures, and the tool supports hierarchical navigation and text search:

screenshot: concept browser without stats

About 14,000 concepts are included with the I2B2 software. As we extract concepts from health records at KU Medical center, some obviously match the i2b2 concepts but many do not. For example, we have not yet determined the relationship between local medication codes and the NDC system. This complicates the already daunting task of navigating the hierarchy of medical terminology. It's quite frustrating to poke around in the dark, running query after query, not knowing which concepts have real data attached to them.

We find that pre-computing the number of facts and the number of associated patients and including these statistics in the hierarchy makes navigation much more efficient:

screenshot: concept browser with stats

We can see at a glance that we have no records of dispensing ANTI-OBESITY drugs, at least by that exact categorization.

We can also see that we only have rich data (diagnoses, labs, meds, ...) about roughly 10% of the 2 million patients in our database; this is due to rather recent deployment of an electronic medical record system compared to the long-standing use of computerized billing systems.