Real-estate data XML parser and database Import script.

Date: 04-Nov-2015

[8:31:21 AM] Simon Grant: Hi Tahir.

[8:31:25 AM] Simon Grant: Goodthankyou.

[8:32:03 AM] Simon Grant: Next we are going to move onto the next project. I would like you to write a script that parses an xml file.

[8:32:41 AM] Simon Grant: The script will update the database based on information in the xml file. Will you be able to do this?

[8:33:37 AM] Simon Grant: The project is a straight forward Yii project but you don't need to use Yii for now. Just write the script to access the database directly.

[8:34:34 AM] Simon Grant: You can use what ever libraries you need to do xml parse. I am not really PHP programmer but I imagine there are some tools to help you with this.

[9:31:13 AM] Tahir Hussain: Hi Simon.

[9:31:44 AM] Simon Grant: hi

[9:32:48 AM] Tahir Hussain: I will create an XML file parse script. But can you give me the table structure where I will import xml parsed data?

[9:33:03 AM] Simon Grant: yes ofcourse

[9:33:19 AM] Tahir Hussain: And do we have any sample xml file?

[9:34:04 AM] Simon Grant: yes I will give you the specifications soon. First I will upload database to dev server where you can download ok?

[9:34:22 AM] Tahir Hussain: ok great.

[9:34:34 AM] Simon Grant: just give me a few mins

[9:34:39 AM] Tahir Hussain: sure

[9:35:44 AM] Simon Grant: i will leave data in it because it will make it easier

[9:36:11 AM] Tahir Hussain: that's fine

[9:41:34 AM] Simon Grant: ok, file realestate.sql.gz is in your home directory on staging server

[9:41:49 AM] Simon Grant: is a mysql file. Please install on your local system.

[9:42:55 AM] Tahir Hussain: ok let me check

[9:57:41 AM] Tahir Hussain: Ok done. Please send XML sample file

[10:07:33 AM] Simon Grant: This site is a real-estate / property listing site, similar to realestate.com.au. real-estate.com.au is australia's largest property site. They have an xml standard for property listings called REAXML which is what most people use, including us.

[10:08:08 AM] Tahir Hussain: ok

[10:08:26 AM] Simon Grant: you can find an example here http://reaxml.realestate.com.au/docs/residential\_sample.xml

[10:08:41 AM] Simon Grant: the specification for the xml you can find here.

[10:08:44 AM] Simon Grant: http://reaxml.realestate.com.au/

[10:09:26 AM] Simon Grant: don't worry, I don't think you will have to know or use all the elements.

[10:10:01 AM] Simon Grant: just use as a reference. The elements that you will use will be determined by what's in the database, which I think is very small.

[10:11:11 AM] Simon Grant: Our website, feeds off to a number of other client websites. So there is common code in Yii, but different css, image files depending on the clients. So each client will have their own folder.

[10:12:39 AM] Simon Grant: https://ask.realestate.com.au/reaxml/functional-specification/

[10:15:42 AM] Tahir Hussain: ok. So my job is to create a php script that just read that (live) xml file and import data into the given Database. Script can be in simple php.

[10:19:34 AM] Tahir Hussain: Ok I will first analyze the XML file and map it with the database tables.

[10:20:04 AM] Simon Grant: yes you can use php, but would be better if you use some xml library to make it easier.

[10:20:47 AM] Simon Grant: The script will be run every hour through a cron job. You just need to write the script, I will set up the cron job on our server.

[10:21:42 AM] Tahir Hussain: ok I got it.

[10:22:03 AM | Edited 10:22:17 AM] Tahir Hussain: Ok one important thing

[10:22:17 AM] Simon Grant: Because it will run every hour, the information may be duplicated. Or the status of the property might change, for example: from 'on market' to 'sold' or something like that. The script will need to check for that.

[10:22:19 AM] Simon Grant: yes?

[10:23:27 AM] Tahir Hussain: Yes. So the script will check the database for any existed record first and if it finds then the script should update the existing record.

[10:23:36 AM] Simon Grant: yes.

[10:23:46 AM] Tahir Hussain: got it

[10:24:09 AM] Simon Grant: This script is only done for one client at the moment, called Abode property.

[10:24:27 AM] Simon Grant: You can see an example of a listing here http://www.abodeproperty.com.au/real-estate/buy/nsw/burraneer/ap00m9/

[10:26:04 AM] Simon Grant: Try to match up the what you see to the database tables. So each property listing is in the 'listing' table.

[10:27:30 AM] Simon Grant: clients are in 'agency' table. Abode property has id of 4, so you will only be changing their records, understand?

[10:29:06 AM] Tahir Hussain: okay. This script will only import data for the Adobe Property client.

[10:29:08 AM] Tahir Hussain: Got it

[10:29:30 AM] Simon Grant: yes

[10:29:38 AM] Tahir Hussain: great

[10:29:47 AM] Simon Grant: one other thing,

[10:30:39 AM] Simon Grant: I am not very familiar with this application, but when the market status of the property changes, I think you will need to reflect the change in the listing\_market\_history table.

[10:31:22 AM] Simon Grant: The relevant codes are under listing\_market\_history\_code table etc.

[10:31:47 AM] Simon Grant: a lot of tables it seems are used, but you will need to go through them and work out which are the relevant ones.

[10:31:58 AM] Tahir Hussain: ok. It will be updated whenever cron jobs run.

[10:32:19 AM] Tahir Hussain: Yes I will list down all important tables for this script

[10:32:24 AM] Simon Grant: yes , which will be every hour, so make sure we don't duplicate property listings.

[10:32:59 AM] Simon Grant: another important table is locality\_address

[10:33:07 AM] Tahir Hussain: Yes. I will make it sure.

[10:33:21 AM] Simon Grant: and all tables beginning with locality\_

[10:33:51 AM] Simon Grant: if you have any questions, I will try and help.

[10:35:46 AM] Tahir Hussain: sure. thanks

[10:35:56 AM] Simon Grant: Good luck Tahir!

[10:50:26 AM] Tahir Hussain: Realestate.com.au contains different items listing files

[10:50:28 AM] Tahir Hussain:

Business listings

Commercial listings

Commercial Land listings

Rental listings

Residential listings

Residential Land listings

Rural listings

HolidayRental listings

[10:51:03 AM] Simon Grant: yes

[10:51:13 AM | Edited 10:51:20 AM] Tahir Hussain: So the script should be able to import data from each of the file?

[10:51:34 AM] Simon Grant: we should only use what is relevant to what is already in the database

[10:52:00 AM] Simon Grant: is there a field or table that has a listing type?

[10:52:35 AM] Tahir Hussain: ok let me check from the database.

[10:54:46 AM] Simon Grant: try to find relevant table. It may be under one of the listing tables or in one of the locality tables.

[10:55:50 AM] Simon Grant: If we don't make use of types of listing in our database, then we don't use it! We only need to pull in enough data to make our site work. But be careful not to miss any important fields from the xml we can use ;-)

[9:43:46 AM] Tahir Hussain: I have found a few things.

[9:44:36 AM] Tahir Hussain: The database contains property of 5 types (Business, commercial, Rural, Residential, Holiday).

[9:45:12 AM] Tahir Hussain: So we need to import 5 types of property list from the xml

[9:45:40 AM] Simon Grant: which table has these entries?

[9:46:03 AM] Tahir Hussain: listing\_trade\_type\_code

[9:47:40 AM] Simon Grant: ok

[9:49:32 AM] Tahir Hussain: So I will use 5 sample files to import data for agency "Adobe Property"

[9:51:06 AM] Simon Grant: No, I image that when you will recieve a file, it may contain many different types of property types, depending on the xml property

[9:51:56 AM] Simon Grant: you know which xml tag is used for property type?

[9:52:03 AM] Tahir Hussain: yes

[9:53:25 AM] Simon Grant: ok, so you understand? depending on the value of the xml tag, you update the value in the database. I think that for property type, you only need to modify when you add a new listing. After that, property type won't change. Understand?

[9:53:53 AM] Tahir Hussain: Yes I understand.

[9:54:06 AM] Simon Grant: cool, you have another question?

[9:54:46 AM] Tahir Hussain: No more question. I am just about to start development.

[9:54:51 AM] Tahir Hussain: Thank you

[9:54:57 AM] Simon Grant: are you using a library for xml parse?

[9:56:15 AM] Tahir Hussain: Yes I will this one REA XML Parser

[9:56:16 AM] Tahir Hussain: https://github.com/benjy/REA-XML-Parser

[9:56:32 AM] Simon Grant: oh, great!

[9:56:57 AM] Tahir Hussain: But it will need a few modifications.

[9:57:03 AM] Simon Grant: ok

[9:57:38 AM] Simon Grant: at the end of the script, rename the file when it is complete, so that we know not to reprocess it again.

[9:58:06 AM] Simon Grant: i mean rename xml file to \*.xml.complete or something

[9:59:20 AM] Simon Grant: you will probably need to name it to a timestamp .xml.20151231 or something so we don't keep overwriting. That way we keep all xml files that come in every hour.

[10:01:01 AM] Tahir Hussain: ok. once I import data from the xml I will rename it .xml.timestampValue

[10:01:06 AM] Tahir Hussain: That's fine.

[10:01:28 AM] Simon Grant: thanks Tahir

Database tables and fields

1. Table: Listing

Fields and Mapping

1. Reference\_number (Don’t know what is this )
2. Agency\_id: referenced Agency table, value = 4
3. Locality\_address\_id: reference table locality\_address, it contains property address.
4. locality\_address\_visibility\_code: reference table visibility\_code (1=visible)
5. listing\_type\_code: reference table listing\_type\_code(aution, for sale etc)
6. listing\_trade\_type\_code: reference table “listing\_trade\_type\_code” (commercial, rental, business, residential)
7. listing\_market\_status\_code: reference listing\_market\_status\_code(withdraw, sold, on Hold, settled)
8. listing\_marketed: datetime field
9. listing\_publish\_status\_code: reference table: listing\_publish\_status\_code(deleted, withdraw, published)
10. listing\_published: datatime
11. map\_publisher\_code: reference table: map\_publisher\_code
12. map\_publisher\_code: reference table: locality\_address\_visibility\_code
13. headline: Title of the property
14. summery and description: Description of the property
15. hot\_property: Boolean (1,0)
16. external\_reference\_id: 5-6 digits number but don’t know exactly
17. external\_reference\_address:
18. plot\_id
19. status\_code: (Inactive, active, deleted, deactivated)
20. created = datetime
21. modified: datetime
22. external\_reference\_id = will be uniqueId of the property in xml file