Parsing Data Obtained from dbSNP's FTP Site

Created: July 7, 2005; Updated: February 18, 2014.

I'm using the perl script: getXML.pl to query genotypes. Are there perl-based XML parsers available for genotype XML format?

There are many perl modules for parsing XML on CPAN: You could also try XML::Twig or XML::Parser. (**6/11/07**)

I've tried to use a SAX parser to parse the files in Entrez SNP, but the parser is looking for a DTD file called "idxsnp.dtd". Where is it?

The idxsnp.dtd is located in the ftp specs directory.

Where can I find an XML parser for an XML file I have downloaded?

There are XML parser modules on CPAN, but we are unable to provide any help and/or support for these modules. If you have questions about them, you'll have to check on the web or contact the author of the module. (**4/20/06**)

I am parsing files to determine allele discrepancies for RS numbers across mouse strains, and found that each strain with a discrepancy has the flag "gtyFlag1" on the GTypeByInd tag line. Can I use this flag to parse for allele discrepancies?

Yes, you can use the "gtyFlag1" flag to parse for allele discrepancies since this flag is set when two submissions for the same SNP/strain are inconsistent. (**4/8/05**)

Parsing Problems

I've completed a dbSNP batch query, and when I opened it with my parser, I got an error message. When I opened the file with Microsoft Explorer, I got both the data and the file structure. All I need is to see the data.

Try using Microsoft Word or WordPad to view the XML file. Microsoft Explorer doesn't automatically expand the nodes.