

GET-Evidence - Feature #490

Classify web-hits as relevant/not-relevant

05/10/2010 09:25 AM - Madeleine Ball

Status: In Progress	Start date:
Priority: Normal	Due date:
Assigned To: Tom Clegg	% Done: 80%
Category:	Estimated time: 0.00 hour
Target version:	Totalhours:
Billable:	Resolution:
Estimatedhours:	
Hours:	

Description

Each time a variant is seen in a new genome, it should be queued for web-search. The search should take place on chr position, rsID, and when available gene/amino acid change (both one letter and three letter abbreviations).

Logged in users should be able to classify web-search results as relevant/not-relevant. (Incrementing a counter for "relevant" or "not relevant"; user can change their mind but only one "vote" per user!)

Implementation notes:

- Add columns to flat_summary table for autoscore, web hits, genome hits, webscore... and refresh it
- Relax web search criteria: include all single-genome-hit variants
- Add long form AA (and rsid where applicable) to search terms in web search
- Requeue old web searches (to pick up rsid and long form AA results)
- Web hit vote history: {variant, url, oid, timestamp, score}
- Web hit current vote: {variant, url, score}
- UI: "vote yes" and "vote no" buttons (immediate ajax call)
- During vote event, update flat summary if webscore has changed as a result
- Tie = relevant, otherwise majority

Questions:

- Is webscore=0 (not relevant) suitable for variants with no web search results?

History

#1 - 05/28/2010 03:17 PM - Madeleine Ball

Sub-tasks:

- This should be accompanied by a report which shows the number of variants seen in only one genome and sorted by autoscore (a column in the report).
- The dump should have a column for "relevance" which can have the following values:

Relevant (at least 1 user identified a relevant web-hit)

Not-Relevant (at least 1 user identified every web-hit as not relevant)

Not-Reviewed

#2 - 05/28/2010 03:34 PM - Madeleine Ball

The report should be a list of variants only in one genome, sorted according to autoscore and randomized within each autoscore.

IE, in order:

[randomized list of variants in only one genome with autoscore = 6]

followed by

[randomized list of variants in only one genome with autoscore = 5]

...

etc.

It would be nice if we also had a column summarizing the evaluated web hits so far (ie, someone has evaluated this as having N valid hits and M nonvalid hits). It should **not** report the raw number of web hits. (I don't want this to bias our behavior in going through these evaluations for the purpose of the paper, someday we may wish to add it of course.)

#3 - 05/30/2010 05:06 AM - Tom Clegg

The voting system is in place, so you can vote relevant / not relevant for each web hit.

The current decision (if any) for each web hit is shown on the left side of the link.

If you're logged in, you can vote by clicking the "thumbs up" or "thumbs down" icons on the right side of the link.

UI fixes todo:

- Indicate what **your** current vote is for each link (perhaps by highlighting the icon that represents your existing vote)
- Improve icons (suggestions other than fixing the anti-alias/resize ugliness?)
- Tool-tips to indicate what the voting icons do

#4 - 05/30/2010 05:44 AM - Tom Clegg

First stab at a report: <http://evidence.personalgenomes.org/report?type=need-web-review>

Also, the latest-flat dump has a "webscore" column:

- "Y" if at least one web hit has been voted "relevant" by a majority or a tie; or
- "-" if some web hits have not been voted on yet; or
- "N" otherwise (i.e. all web hits are voted "not relevant", or there are no web hits)

#5 - 05/30/2010 02:59 PM - Madeleine Ball

This is great! Can you increase the number of pages displayed at:

- <http://evidence.personalgenomes.org/report?type=need-web-review>
We need to page through all such variants with "autoscore >= 2"

#6 - 06/04/2010 02:25 PM - Tom Clegg

Replying to [comment:10 <https://www.google.com/accounts/o8/id?id=AltOawk3Kp1W2TZ9rVrm-L11IDIFY9UvU5ReVXw>]:

This is great! Can you increase the number of pages displayed at:

- <http://evidence.personalgenomes.org/report?type=need-web-review>
We need to page through all such variants with "autoscore >= 2"

(Fixed the other day but forgot to update ticket, oops.)

This report now shows 1200 hits. The first 846 hits have autoscore >= 2.

#7 - 02/15/2011 03:17 PM - Ward Vandewege

- Project changed from External to GET-Evidence
- Category deleted (GET-Evidence)

#8 - 09/01/2011 07:16 PM - Tom Clegg

- Status changed from New to In Progress
- Priority changed from Urgent to Normal
- % Done changed from 0 to 80