Google apps @UDel.edu

Wei Zhong <zhongwei@udel.edu>

[ntcir11ann-math:00026] NTCIR 11 Wikipedia Math Search Subtask

2 messages

Moritz Schubotz <schubotz@tu-berlin.de> Wed, Aug 27, 2014 at 5:32 PM
Reply-To: ntcadm-math@nii.ac.jp
To: ntcir11ann-math@nii.ac.jp
Cc: ntcir11adm-math@nii.ac.jp

(Please find attached a PDF version of the announcement.)
Dear participants,

The first round of the NTCIR-11 Math Search Wikipedia Subtask [1] has completed.

Seven runs have been submitted by four different teams with 100 queries.

The 100 queries were generated in the following way $(1\le i\le 100)$:
(1) first, a random English Wikipedia article A_i containing math
formulae is chosen;
(2) a formula f_i is chosen randomly (with n_i variables) from the
article A_i;
(3) for every variable v_j with $1\le j\le n_i,$ occurring in the
formula f_i,
a uniform random number $0<y_j<1$ was calculated;
(4) given a constant threshold z, and the nesting level of the variable l_j,
if $y_j>z*l_j$, the variable was replaced using the qvar concept.
In the formula f_i, K is the set of indices of selected variables,
and $k\in K$ with $1\le k \le #(K)$ gives the element number of
variables are selected.
Using the qvar concept, the kth replaced variable
$u_k:=v_{K(k)}$ is expressed as "$?\{xk\}$".

The runs are evaluated as follows.

Each run r ($1\le r\le 7$) with query i selects $P(r,i)$ articles.
The function $S(r,i,p(r,i))$ gives the pth ranked article ($1\le
p(r,i) \le P(r,i)).
If A_i is an element of $B(r,i):=\{S(r,i,x):x\in\{1,\ldots,P(r,i)\}\}$,
then $X(r,i)=1$ otherwise $X(r,i)=0$.
The number of correct results for each run is given by
$R(r)=\sum_{i=1}^{100} X(r,i)$
which are given in the following table:

r	R(r)
7	91
1	88
5	87
6	87
2	75
3	48
4	1

In a second evaluation, we only considered the top ranked t articles.
If A_i is an element of
$C(r,i,t):=\{S(r,i,x):x\in\{1,\ldots,\min(t,P(r,i))\}\}$,

then $Y(r,i,t)=1$ otherwise $Y(r,i,t)=0$. Then the number of correct results for each run is given by $Q(r,t)=\sum_{i=1}^{100} Y(r,i,t)$. In the following table we list all results for $t\in\{1,2,3,4,5,10,100,1K,10K,10M\}$:

r	t	Q(r,t)
7	10M	91
7	10K	89
1	10K	88
1	10M	88
7	1K	87
6	10M	87
5	10M	87
1	1K	86
5	10K	86
6	10K	85
1	100	84
5	1K	84
6	1K	83
7	100	79
1	30	78
5	100	76
1	20	75
6	100	75
2	100	75
2	1K	75
2	10K	75
2	10M	75
1	10	71
7	30	69
1	5	69
7	10	69
7	20	69
7	5	68
2	20	68
5	30	68
2	30	68
1	4	67
2	10	67
6	10	67
6	20	67
6	30	67
1	2	66
7	2	66
1	3	66
7	3	66
7	4	66
6	5	66
5	10	66
5	20	66
5	5	65
2	5	65
6	2	64
6	3	64
6	4	64
7	1	63
2	2	63
5	2	63
5	3	63
2	3	63
5	4	63

2	4	63
1	1	62
6	1	61
5	1	60
2	1	58
3	1K	48
3	10K	48
3	10M	48
3	100	12
3	30	5
3	20	4
3	10	3
4	100	1
4	1K	1
4	10K	1
4	10M	1

Note that one team was not able to participate because the input data contained several non-well formatted XML articles. I apologize for this inconvenience. Using our recently committed change to MediaWiki, we are now able to generate fully well-formatted XML and this will not be a problem in the future.

An improved version of the input data is scheduled to be released on September 1, 2014.
The 100 queries will not change, but the result submission process will be automated
using a recently developed web interface. Since the evaluation is already automated,
results can quickly be returned to the participant.

Another modification is that the retrieval unit in the Wikipedia Subtask are formulas and not Wikipedia articles. This means for the next round, beginning on September 1, 2014, participants must include references to the formulae.

Best,
Moritz Schubotz

[1] http://ntcir11-wmc.nii.ac.jp/index.php/NTCIR-11-Math-Wikipedia-Task

--
Moritz Schubotz
TU Berlin, Fakultät IV
DIMA - Sekr. EN7
Raum EN742
Einsteinufer 17
D-10587 Berlin
Germany

Tel.: +49 30 314 22784
Mobil.: +49 1578 047 1397
Fax: +49 30 314 21601
E-Mail: schubotz@tu-berlin.de
Skype: Schubi87
ICQ: 200302764
Msn: Moritz@Schubotz.de

ntcir11wmc2ann.pdf
47K

Moritz Schubotz <schubotz@tu-berlin.de> Mon, Sep 1, 2014 at 9:33 PM
Reply-To: ntcadm-math@nii.ac.jp
To: Moritz Schubotz <schubotz@tu-berlin.de>
Cc: ntcir11ann-math@nii.ac.jp, ntcir11adm-math@nii.ac.jp

Dear all,

as announced, the improved data set is available for download now

http://demo.formulasearchengine.com/images/ntcir-11-wmc-1b.zip

There is one minor update to the format. The page IDs have been replaced with revision IDs. The advantage is that the link to the page revision will not change in the future.

The results can be upload via a web interface. The web interface supports just a simple CSV result format.
If participants prefer to use the XML format we will provide an other interface.

To upload results, an account with special rights is required.
Please contact me to retrieve an account.

Best
Moritz
[Quoted text hidden]