

Dolphin18K for Math Word Problem Solving

1. Introduction

Dolphin18K is a dataset created for math word problem solving, containing 18,460 problems posted by users on the community question answering site, Yahoo! Answers.

2. Data Description

(1) File Format

Field	Description	Example
id	Unique id of a problem.	yahoo.answers. 20061213041448AAxoy5z
original text	Original problem text from yahoo answer.	what is 30 divided by half plus 10? 1st correct answer gets 10 points!!!!?\n
text	Cleaned problem text by removing problem unrelated text, used in our paper.	what is 30 divided by half plus 10? 1st correct answer gets 10 points!!!!?\n
equations	The equations used to solve the problem.	unkn: x equ: $x=30/(1/2)+10$
ans	Problem answer.	70
flag	Annotation status of the problem	0 – annotation done

(2) Equation Annotation

- “unkn”: the variables whose values will be problem answers

- “equ”: <the first equation>
- “equ”: <the second equation>
- ...
- “equ”: <the last equation>

(3) Answer Annotation

- “;”: separate the values of different variables.
- “{}”: any order of variables is allowed.
- “|”: different answer formats.
- “or”: different solutions to the problem.

See the examples below:

Problem	Equations	Answer
3 times the sum of twice a number and 5 is 4 times 2 less than the number. Find the number.	unkn: x equ: $3*(2x+6)=4(x-2)$	-13
Find two consecutive integers whose sum is 27.	para: x unkn: x,x+1 equ: $x+(x+1)=27$	13; 14
Find 2 numbers whose sum AND product are 11.	unkn: x,y equ: $x + y = 11$ equ: $x*y = 11$	{1.113; 9.887}
60 is 140% of what number?	unkn: x equ: $60 = 140/100 * x$	300/7 42.857

3. Dataset Statistic

(1) Manual/Auto

Manual: equations and answers are annotated by human

Auto: equations and answers are annotated automatically
(method of automatic labeling is introduced in [1])

(2) Template Size

Template: Unique form of equation system.

For example, the following equation $x=3+5$ corresponds to template $x=n_1+n_2$.

Template size = number of problems corresponding to the template.

		Equations + Answer	Answer only	Sum
Dev	Manual	909	67	976
	Auto	2,245	507	2,752
	All	3,154	574	3,728
Eval	Manual	3,605	321	3,926
	Auto	8,754	2,052	10,806
	All	12,359	2,373	14,732

Template size	Manual	Auto	All
≥ 1	2,675	7,969	10,644
≥ 2	2,036	5,956	8,229
≥ 6	1,578	4,826	6,827

4. Get the dataset

(1) Download original text from Yahoo! Answers

Run OriginalDataExtractor.py

- Arguments:

-i input_url_file -o output_file -t number_of_threads (default 10)

- Example command line:

```
python OriginalDataExtractor.py -i dev_urls.json -o  
dev_original.json -t 10
```

(2) Get the cleaned-text version

Run CleanVersionExtractor.py

- Arguments:

-i input_organial_file -d input_diff_file -o output_cleaned_file

- Example command line:

```
python CleanVersionExtractor.py -i dev_original.json -d  
dev_diff.pkl -o dev_cleaned.json
```

(3) Get different subsets

Run SubsetExtractor.py

- Arguments:

-i input_cleaned_file -s subset_id_file -o output_subset_file

- Example command line:

```
python SubsetExtractor.py -i dev_cleaned.json -s  
dev_ids\\dev_manual.txt -o dev_ids\\dev_manual.json
```

Reference:

[1] Danqing Huang, Shuming Shi, Chin-Yew Lin, Jian Yin and Wei-Ying Ma. How Well Do Computers Solve Math Word Problems? Large-Scale Dataset Construction and Evaluation. ACL 2016.