Understanding Task Design Trade-offs in Crowdsourced Paraphrase Collection

Youxuan Lucy Jiang, Jonathan Kummerfeld and Walter Lasecki
Paraphrases are useful across a wide range of NLP tasks
Paraphrases are useful across a wide range of NLP tasks

Sentence
Paraphrase 1
Paraphrase 2
...

Motivation
Experiment
Results
Paraphrases are useful across a wide range of NLP tasks

Sentence
Paraphrase 1
Paraphrase 2
...

Motivation
Experiment
Results
Paraphrases are useful across a wide range of NLP tasks.
Paraphrases are useful across a wide range of NLP tasks

The boy desires the girl to believe him.

The boy desires to be believed by the girl.

The boy has a desire to be believed by the girl.

The boy’s desire is for the girl to believe him.

…
Paraphrases are useful across a wide range of NLP tasks

What is the smallest state by area?

Which state has the least area?

Name the state with the smallest area...

\[ \lambda x_1.\text{smallest}(x_2, \lambda x_1.\text{state}(x_1), \lambda x_1.\lambda x_2.\text{area}(x_1,x_2)) \]
Paraphrases are useful across a wide range of NLP tasks

What is the smallest state by area?
Which state has the least area?
Name the state with the smallest area

SELECT state.state_name FROM state
WHERE state.area = (SELECT MIN(state.area) FROM state);

...
Paraphrases are useful across a wide range of NLP tasks

What is the smallest state by area?
Which state has the least area?
Name the state with the smallest area

---

```
SELECT state.state_name
FROM state
WHERE state.area = (SELECT MIN(state.area) FROM state);
```
Many factors affect the quality and diversity of crowdsourced paraphrases
Many factors affect the quality and diversity of crowdsourced paraphrases
Many factors affect the quality and diversity of crowdsourced paraphrases

Title
Many factors affect the quality and diversity of crowdsourced paraphrases.
Many factors affect the quality and diversity of crowdsourced paraphrases.
Many factors affect the quality and diversity of crowdsourced paraphrases.
Many factors affect the quality and diversity of crowdsourced paraphrases.
Many factors affect the quality and diversity of crowdsourced paraphrases.
Many factors affect the quality and diversity of crowdsourced paraphrases
Many factors affect the quality and diversity of crowdsourced paraphrases.
Many factors affect the quality and diversity of crowdsourced paraphrases.

Title
Instructions …

Examples

Prompt Sentence
Many factors affect the quality and diversity of crowdsourced paraphrases.

$0.05 - participation

Instructions …

Examples

Prompt Sentence
Many factors affect the quality and diversity of crowdsourced paraphrases

$0.05 - participation

$0.05 - bonus
Many factors affect the quality and diversity of crowdsourced paraphrases.
Many factors affect the quality and diversity of crowdsourced paraphrases
Many factors affect the quality and diversity of crowdsourced paraphrases.
Many factors affect the quality and diversity of crowdsourced paraphrases
Many factors affect the quality and diversity of crowdsourced paraphrases
Many factors affect the quality and diversity of crowdsourced paraphrases
Many factors affect the quality and diversity of crowdsourced paraphrases.
Many factors affect the quality and diversity of crowdsourced paraphrases.
Many factors affect the quality and diversity of crowdsourced paraphrases
Many factors affect the quality and diversity of crowdsourced paraphrases.
Many factors affect the quality and diversity of crowdsourced paraphrases.
Many factors affect the quality and diversity of crowdsourced paraphrases
Priming has a significant impact

- Motivation
- Experiment
- Results

Correctness
- Syntactic
- Lexical Variation
- Lexical & Syntactic

Diversity
Priming has a significant impact

Motivation

Experiment

Results

Correctness

Diversity

Syntactic
Lexical Variation
Lexical & Syntactic
No Examples
Priming has a significant impact
Priming has a significant impact
Priming has a significant impact
Jargon is a challenge

Motivation

Experiment

Results

Correctness

Diversity

Course Questions

Questions

Advising

Dialogue

GeoQuery

WSJ

Ubuntu

IRC
Jargon is a challenge

Correctness and Diversity for different domains:

- **Course Questions**: Correctness 70, Diversity 100

Other domains include Advising Dialogue, GeoQuery, WSJ, Ubuntu, IRC.
Jargon is a challenge

Correctness

Diversity

Course Questions
Advising Dialogue
GeoQuery
WSJ
Ubuntu IRC

Motivation
Experiment
Results
The first systematic study of trade-offs when crowdsourcing paraphrases

Thank you!

Youxuan Lucy Jiang,
Jonathan Kummerfeld and
Walter Lasecki

Supported by IBM