Article n.47

Managing Web Experiments for Psycholinguistics: An Example from Experimental Semantics/Pragmatics

Article by J. Degen and J. Tonhauser
Presentation by Sungwoo Kwon

Linguistics and Data Science 2023F

October 29, 2023

Outline

Introduction

- Presenting the workflow of a study conducted in the past by the authors, rather than data itself.
- Experimental Data from HITs(Human Intelligence Tasks).
- Research Topic
  - Investigating the extent to which variability in a content's projectivity is predicted by that content's at-issueness.
  1. Does Felipe regret that Mike visited Alcatraz?
  2. Felipe doesn't regret that Mike visited Alcatraz.
  3. Does Felipe think that Mike visited Alcatraz?
  4. Felipe doesn't think that Mike visited Alcatraz.

Data Management

- The Organizational Framework for Collaboration
  - Github, Git, Slack
- Project Repository Structure
  - Generally contains
    - *README.md*
    - *experiments*
    - *results* (or *analysis*)
    - *paper*
  - Additionally contains
    - *data*
    - *models*
    - *talks*

Data Collection

- Amazon Mechanical Turk
  - A crowdsourcing platform for HITs
  - many advantages over in-lab studies
    - ★ Larger and diverse participants
    - ★ Rapid completion times
- Workflow in AMT
  - Web-programmed experiments using JS,HTML,CSS etc.
  - Supersubmiterator for interfacing with AMT
  - Directory management

Supersubmiterator

- A tool for managing external HITS on Amazon Mechanical Turk
- Makes it easy to retrieve the data once the experiment is completed, with participant IDs anonymized and results formatted into a .csv file.
- Up-to-date with current AMT configurations.
- Provides automatic batching feature.

Topics

- approveQualificationRequests
- approveWork
- assignQualification
- blockWorker
- createQualificationType
- deleteHITs
- evaluateQualificationRequests
- extendHITs
- getBalance
- getQualificationRequests
- getResults
- grantBonus
- loadHITs
- makeTemplate
- rejectQualificationRequests
- rejectWork
- resetAccount
- reviewResults
- revokeQualification

Property	Description
title	Title of the HIT
description	Description of the HIT
keywords	Keywords to associate with the HIT
reward	Reward for the HIT formatted as $0.00 USD
assignments	Maximum number of assignments available for this HIT
annotation	Value you use to identify this HIT **Tip** You can merge values from the input file into this property. To do this, use the syntax: `${the field name}`, where the field name is the column name from the input file. Amazon Mechanical Turk uses Apache Velocity to perform the merge. For more information about merge syntax, go to http://velocity.apache.org.

`https://docs.aws.amazon.com/AWSMechTurk/latest/AWSMturkCL/Welcome.html`

Supersubmiterator

```
{
"liveHIT":"yes",
"title":"Language study",
"description":"You will rate short dialogs.",
"experimentURL":"https://web.stanford.edu/~jdegen/exp1a/",
"keywords":"language research fun cognitive science linguistics university",
"USonly?":"yes",
"minPercentPreviousHITsApproved":"95",
"frameheight":"650",
"reward":"1.00",
"numberofassignments":"250",
"assignmentduration":"1800",
"hitlifetime":"2592000",
"autoapprovaldelay":"60000"
}
```

Web-Based Experiment Pipeline

1. Program experiment as external website and copy it to web space.
2. Create *mturk* directory inside the *experiments* directory where all Mechanical Turk–related files are stored and from where the Submiterator command will be called.
3. Create *experiment.config* file (see example in figure 47.1) in *mturk*.
4. Run from command line to post experiment to Mechanical Turk:

   ```
   submiterator posthit experiment
   ```

5. Run from command line to retrieve data from Mechanical Turk:

   ```
   submiterator getresults experiment
   ```

6. Run from command line to reformat data for worker ID anonymization and easy analysis in R:

   ```
   submiterator reformat experiment
   ```

7. Copy generated data file *experiment.csv* to its corresponding *data* directory in *results* directory.

R

- Open Source Statistical Software.
- Ease in managing active analysis scripts and plots.
- Visualisation.
- Syntax highlighting and prediction.
- Integration with R Markdown and knitr.

Data Analysis

- *lme4*, *brms*, *tidyverse*, *ggplot2*, *lsmeans*
- Results Directory
  - data
  - graphs
  - rscripts

Conclusion

- Organizational framework is useful.
  - Reproducibility
  - Generalization
- Preregistration
  - Documenting one's research plan prior to its realization in a public repository, such as OSF.
  - Separating the confirmatory aspects of the research from the exploratory aspects.
  - Allowing for input from colleagues.
  - Allowing journal to accept researchers' papers conditionally.

Quiz

- What are the main reasons Organizational Framework is useful for studies involving human-generated data, especially in Humanities and Social Sciences? (Two answers)