

Survey Search

...

by Pawsitive

Beth Bailey, Leoson Hoay, Ruixue Li

<https://github.com/bethbailey/survey-search>

Outline

- Project Overview
- Site Structure and User Flow
- Awesome stuff we did
- Challenges we faced
- Future work
- Demo (Yay!)
- Questions/Comments

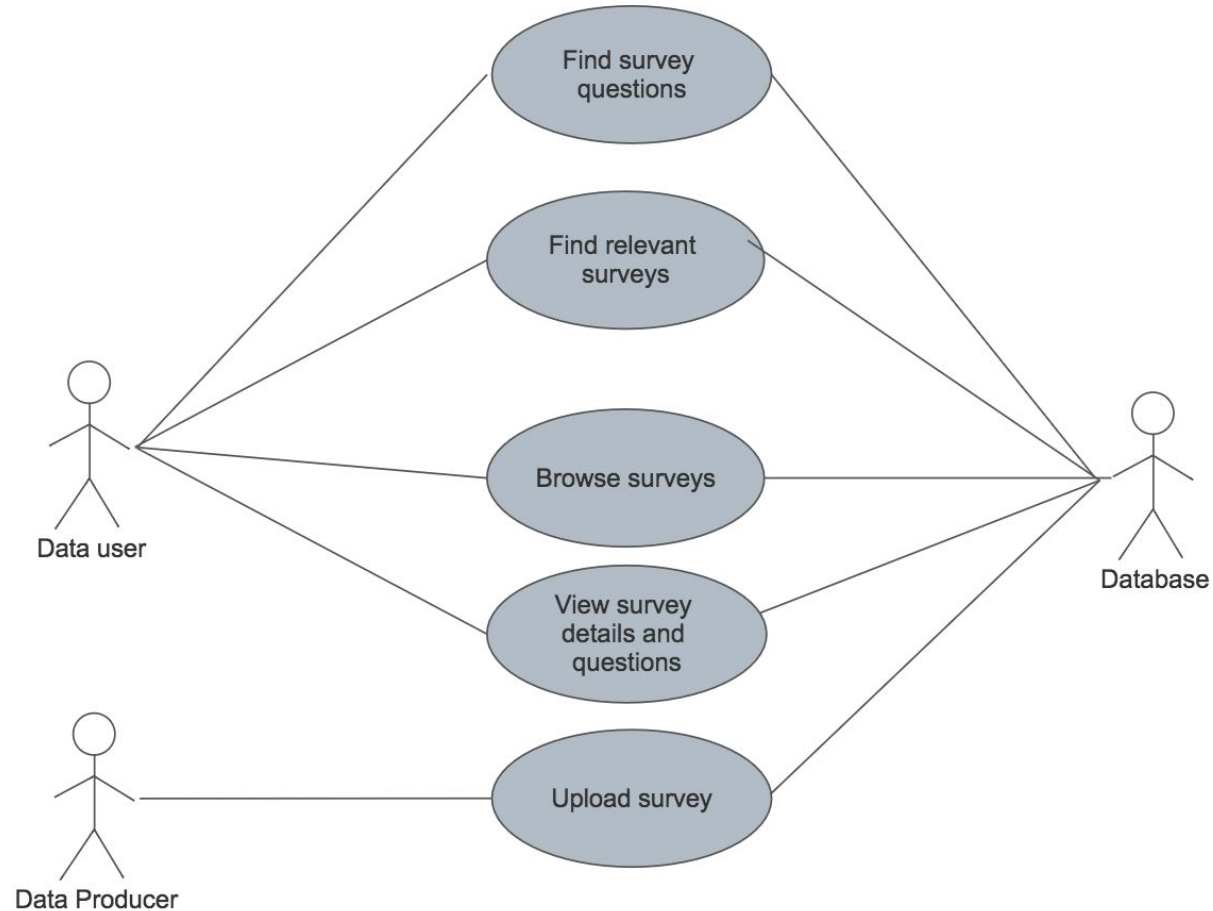
Project Overview

GOAL: Create an easy-to-use survey data repository.

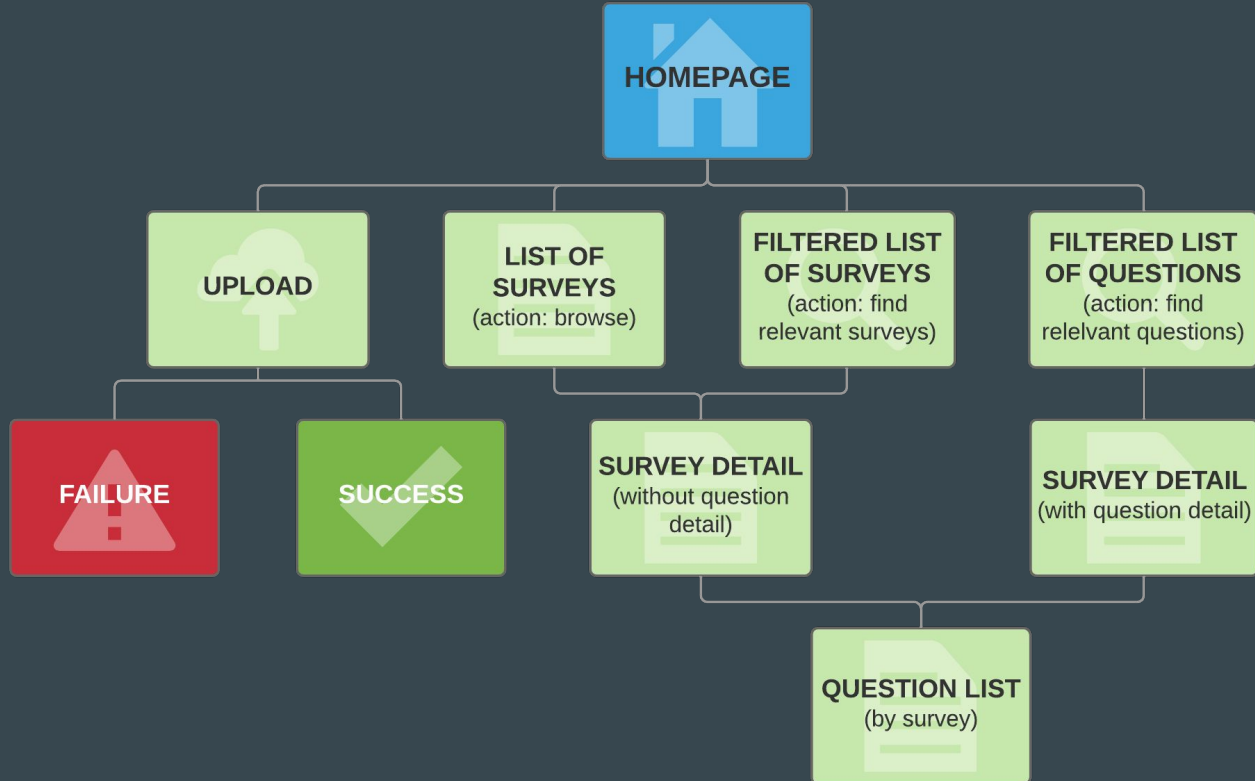
What makes this **different** than other data repositories? Users can:

1. Search *multiple surveys* from different sources.
2. Search surveys on a *granular level* (for individual questions) as well as for entire surveys
3. *Upload and share* their own survey data

Use Case Diagram: Survey Search

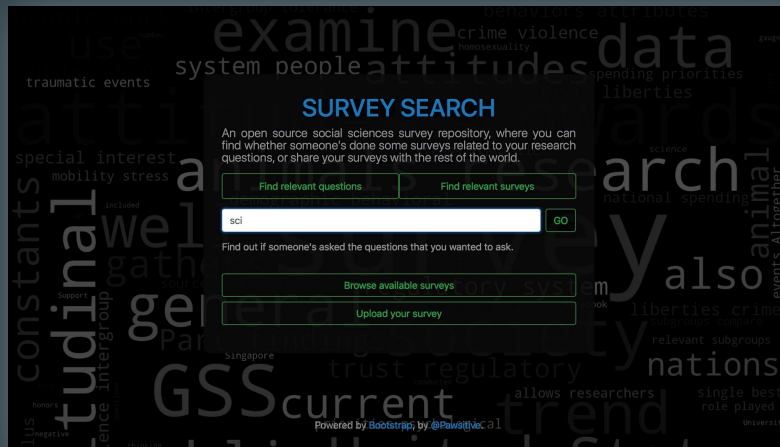


PAWSITIVE SITE MAP



Cool Stuff We Learned!

- Data processing/handling files in Python (pypdf2, docx2txt)
 - Data from surveys, such as the GSS, Pew surveys, and IPSOS surveys
- Lucidchart
- Building a website with Django
 - TF-IDF (term frequency-inverse document frequency) based custom search algorithm (sklearn)
 - Bootstrap frontend
- Word cloud generation



Challenges

- Survey data cleaning is time-consuming and hard
 - Data is in many formats
 - Difficult to reuse code
- A lot of programming involves using resources and reading documentation
 - Don't be afraid to try things
- Django's modularity makes it versatile, but learning curve is steep
 - Integrating multiple modules: search, form, files, word cloud, etc.
- Learning the best practices of software development
 - Version control
 - Teamwork is hard, communicate with and love your teammates <3

Features to come?

- Shared search history/autocomplete
 - Visualization of search patterns
- Advanced query engine
 - A smarter search that accounts for typos and word similarity
 - Show similar surveys (possibly using doc2vec)
- Visualizations of survey data
- User file upload directly to site (alternative to links)

Let's do a demo!

Questions/Comments?

