University of Missouri

Development of Survey Visualization and Advanced Integrated Data Analysis in TigerAware

By: Rui Huang

Advisor: Dr. Yi Shang

Committee: Dr. Yunxin Zhao, Dr. Tim Trull

Outline

- Introduction
- Related Work
- Design & Implementation
- Demos
- Contribution & Future Work

Outline

- Introduction
  - Existing TigerAware Platform
  - Two Improvements
- Related Work
- Design & Implementation
- Results
- Conclusion & Future Work

Introduction

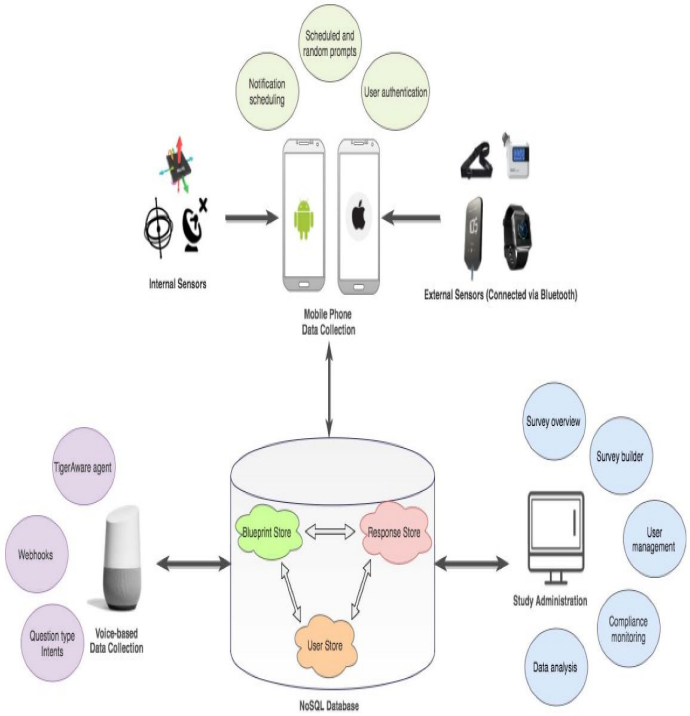
- Researchers across any discipline can follow these six steps to conduct effective survey
  - Define the problem
  - Design the research
  - Design survey questions
  - Deploy Survey
  - Analyze User Responses
  - Write the research report and present its findings

TigerAware Platform →

TigerAware Platform

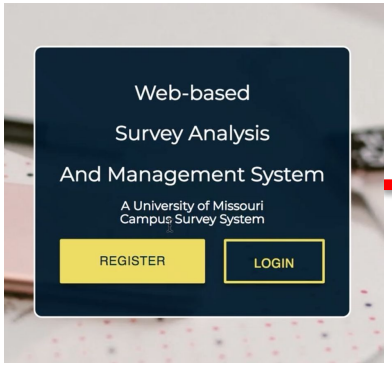
data collection & analysis system

- survey data
  - question responses
- device sensor data
  - GPS
- external sensor data
  - bluetooth breathalyzer



TigerAware System Architecture

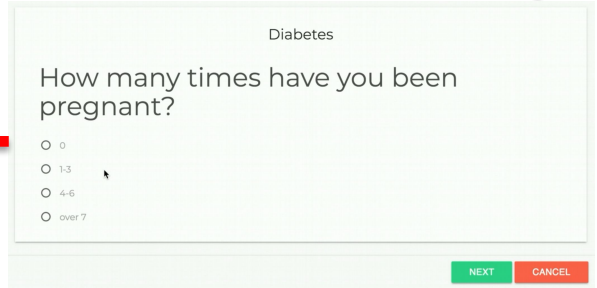
Researchers' point of view



Survey Name

Diabetes

Label	Title	Type	Delete	Edit	Condition
pregnancies	How many times have you been pregnant?	MultipleChoice			
Clucose	Please fill in your plasma glucose concentration in an 2 hour oral glucose tolerance test.	textField			
BloodPressure	What is your blood pressure(mm Hg)?	textField			



Two Improvements in This Project

- Survey Visualization Component
  - interactive
  - exportable
  - intuitive
- Data Analysis Component
  - integrated
  - basic statistics
  - advanced analysis, e.g. Computer Vision & NLP

Outline

- Introduction
- Related Work
- Design & Implementation
- Results
- Conclusion & Future Work

Related Work

- Rogers et al, “Deep Learning at Your Fingertips”, CCNC, 2019
  - all-in-one survey creation, data collection, and data analysis system
  - support both typical statistics(e.g. mean, mode) and advanced deep-learning based analysis(e.g. emotion recognition)

Related Work

- Morrison et al, “An Innovative Mobile Survey and Sensor Data Collection and Analytics System”,IEEE, 2018
  - Design architecture and implement TigerAware system
  - Demonstrate usability of TigerAware system by a number of real world study(e.g. google Assistant Based Diabetes Self Management Study, Driving After Drinking Alcohol Study)

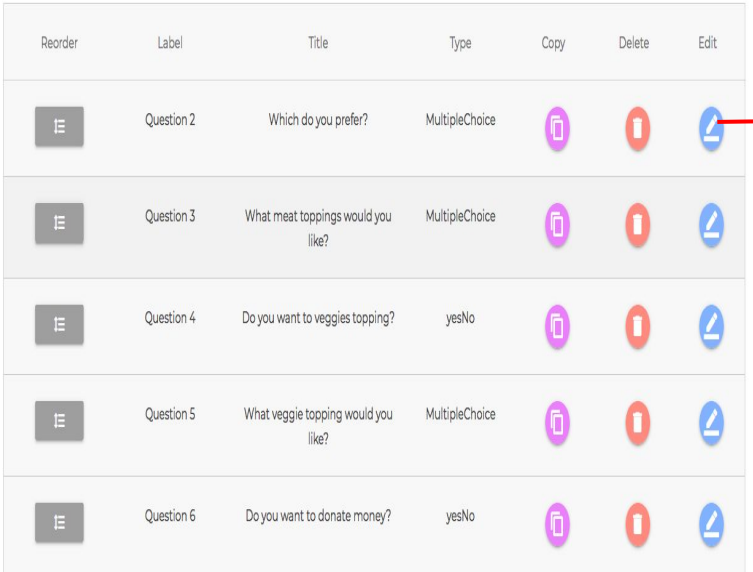
Related Work

- Tutte, “How to Draw a Graph”, *Proceedings* of the London Mathematical Society, 1962
  - propose an algorithm to find planar embedding for planar graph
  - nodes’ position can be determined uniquely as the solution to a system of linear equations

OUTLINE

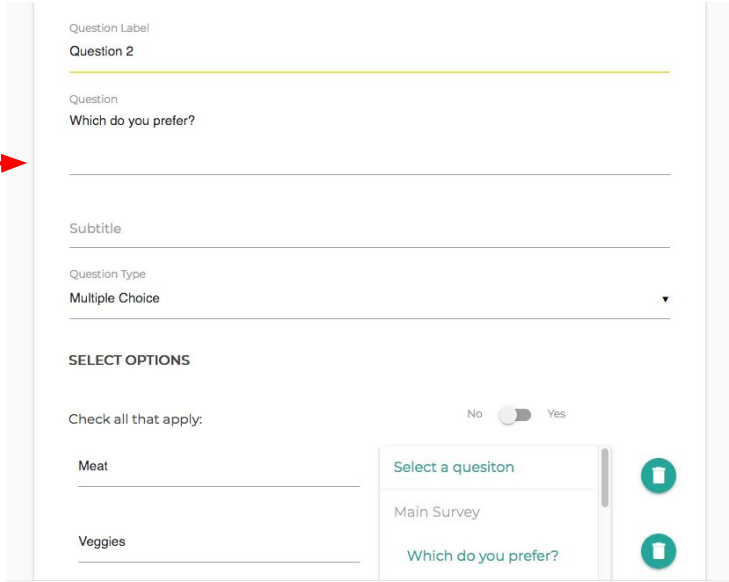
- Introduction
- Related Work
- Design & Implementation
  - Survey Visualization
  - Data Analysis
- Results
- Conclusion

Survey Example



Reorder	Label	Title	Type	Copy	Delete	Edit
	Question 2	Which do you prefer?	MultipleChoice			
	Question 3	What meat toppings would you like?	MultipleChoice			
	Question 4	Do you want to veggies topping?	yesNo			
	Question 5	What veggie topping would you like?	MultipleChoice			
	Question 6	Do you want to donate money?	yesNo			

survey overview



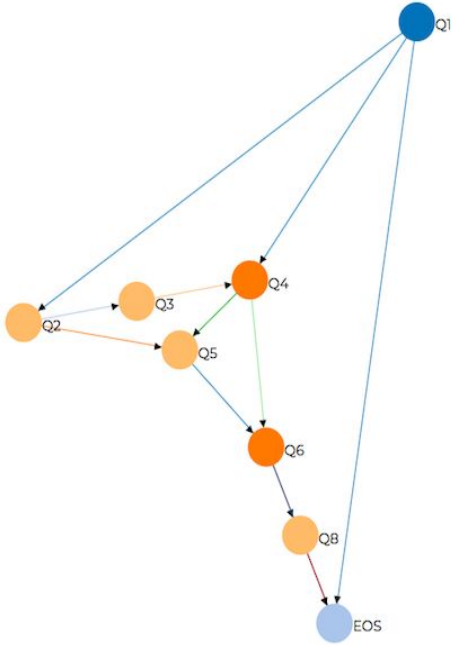
question detail

Survey Visualization Motivation

- issues of creating survey
  - error-prone
  - hard to locate error
  - nonintuitive


```
"survey" : [ {
  "conditionDefault" : "",
  "conditionID" : "",
  "groupID" : "",
  "id" : "Question 1",
  "on" : "",
  "subtitle" : "",
  "title" : "Do you like pizza? ",
  "type" : "yesNo"
}, {
  "choices" : [ "Meat", "Veggies" ],
  "conditionDefault" : "Question 3",
  "conditionID" : "",
  "conditions" : [ {
    "toID" : "Question 3",
    "trigger" : "0"
  }, {
    "toID" : "Question 5",
    "trigger" : "1"
  } ],
  "groupID" : "",
  "id" : "Question 2",
  "multiselect" : false,
```


survey without visualization

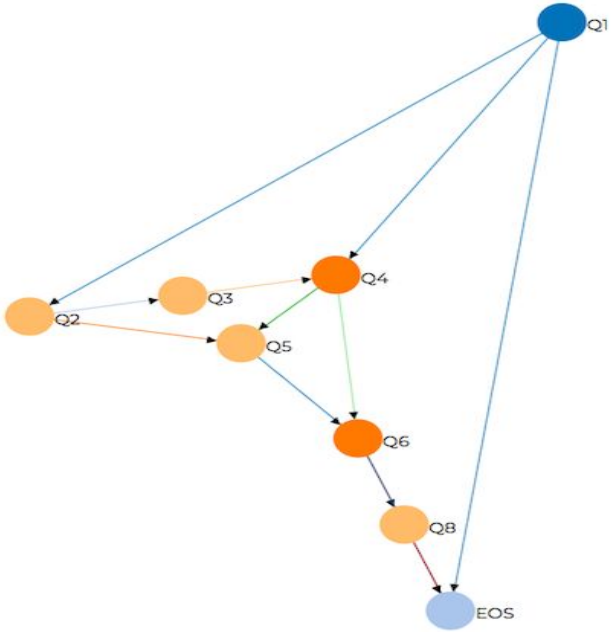


survey with visualization

Visualization Design

Single web page in the TigerAware Dashboard

- survey structure represented as directed graph
- question represented as node
- branches represented as directed edge



Survey Visualization

Visualization Design

- Visualize
  - D3 visualization framework
- Format Converter
  - data format is non-compatible
- Planar Algorithm
  - D3 don’t provide planar embedding, need to be implemented in this project
- Export Survey
  - export graphs as PDF files

Visualization Framework(D3)

D3.js is a JavaScript library for manipulating documents based on data. D3 helps you bring data to life using HTML, SVG, and CSS.

D3 supports force-directed layout, which highly meets the requirements of displaying surveys as directed graphs

Format Converter

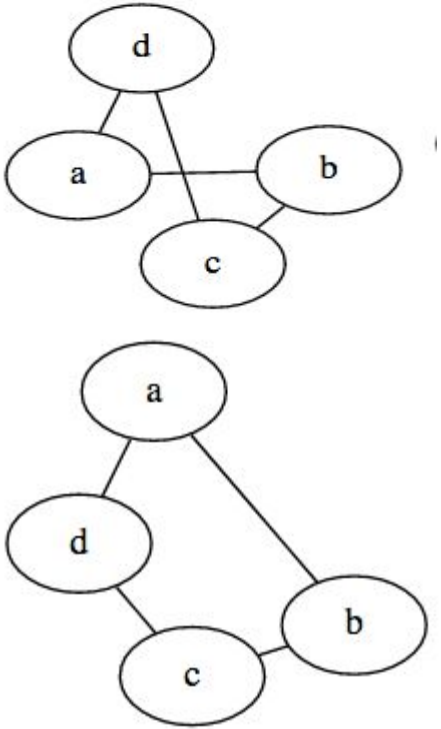
Data Formats not compatible

- TigerAware Data Format
  - questions are stored as JSON object
  - questions connected to each other through pointer
- D3 Data Format
  - questions set
  - edge set
- Adaptor is implemented to convert format

Planar Graph & Embedding

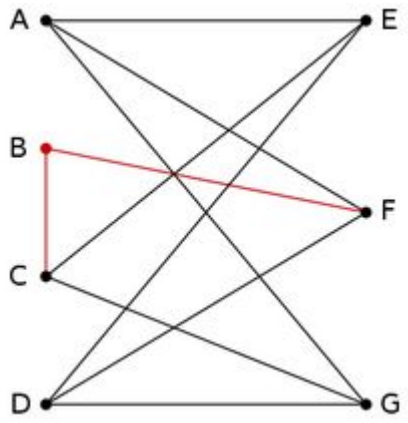
Planar Graph: graph theory, a planar graph is a graph that can be drawn on the plane in such a way that its edges intersect only at their endpoints.

Planar Embedding: such a drawing that no edges cross each other



Kuratowski's and Wagner's theorems

A finite graph is planar if and only if it does not contain a subgraph that is a subdivision of the complete graph K_5 or the complete bipartite graph $K_{3,3}$ (utility graph)
A subdivision of a graph results from inserting vertices into edges (for example, changing an edge •—— • to •—•—•) zero or more times



An example of a graph with no K_5 or $K_{3,3}$ subgraph. However, it contains a subdivision of $K_{3,3}$ and is therefore non-planar.

Tutte's Planar Algorithm

Step 1: fix at least three nodes randomly

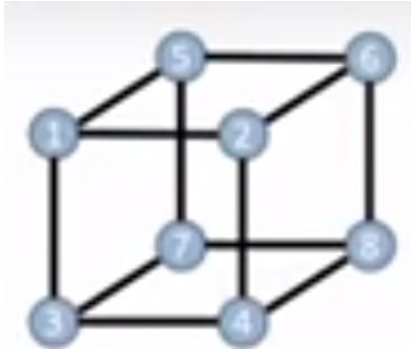
Step 2: create an adjacency matrix L with element Lij = 1/deg(i) for an edge between node i and j

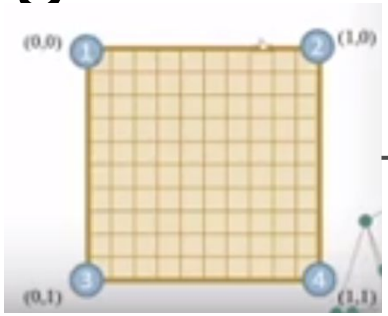
Step 3: generate matrix L' by zero out the rows that already positioned, then create matrix A by subtracting L' from Identity
matrix

Step 4: solve the linear system Ax = bx for x coordinates, where bx is a column vector containing x coordinates for fixed nodes, and 0 for non-fixed nodes.

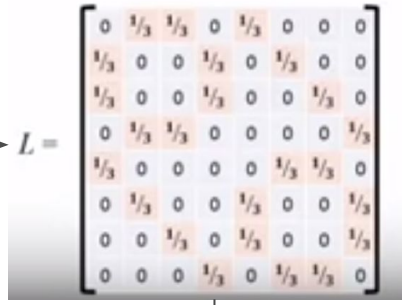
Step 5 : solve the linear system Ay = by for y coordinates, where by is a column vector containing y coordinates for fixed nodes, and 0 for non-fixed nodes

Tutte's Planar Algorithm Example

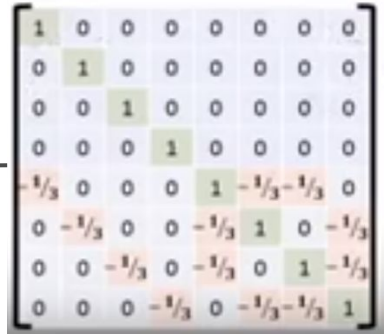




step 1: fix four nodes

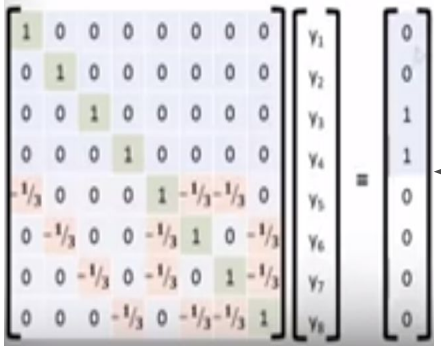


step 2: create L matrix



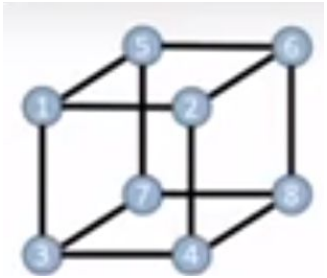
step 3: calculate A matrix

step 4: solution for X



step 5: solution for Y

Tutte's Planar Algorithm Example



$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ -\frac{1}{3} & 0 & 0 & 0 & 1 & -\frac{1}{3} & -\frac{1}{3} & 0 \\ 0 & -\frac{1}{3} & 0 & 0 & -\frac{1}{3} & 1 & 0 & -\frac{1}{3} \\ 0 & 0 & -\frac{1}{3} & 0 & -\frac{1}{3} & 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 0 & -\frac{1}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ -\frac{1}{3} & 0 & 0 & 0 & 1 & -\frac{1}{3} & -\frac{1}{3} & 0 \\ 0 & -\frac{1}{3} & 0 & 0 & -\frac{1}{3} & 1 & 0 & -\frac{1}{3} \\ 0 & 0 & -\frac{1}{3} & 0 & -\frac{1}{3} & 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 0 & -\frac{1}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

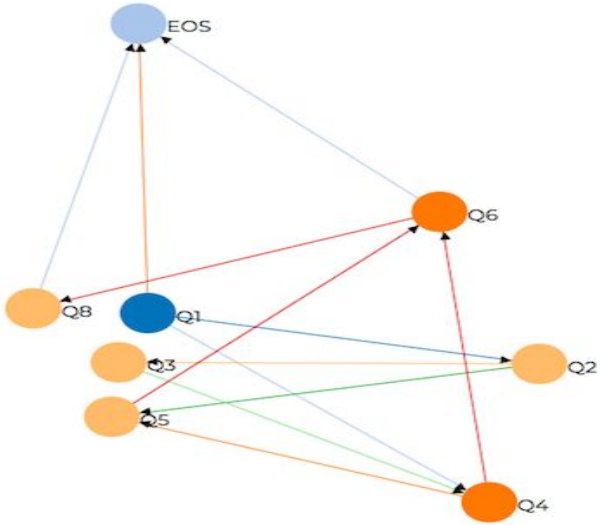
$$x_1y_1 = (0,0), \quad x_2y_2 = (1,0)$$
$$x_3y_3 = (0,1), \quad x_4y_4 = (1,1)$$
$$x_5y_5 = ((x_1y_1) + (x_6y_6) + (x_7y_7))/3$$
$$x_6y_6 = ((x_2y_2) + (x_5y_5) + (x_8y_8))/3$$
$$x_7y_7 = ((x_3y_3) + (x_5y_5) + (x_8y_8))/3$$
$$x_8y_8 = ((x_4y_4) + (x_6y_6) + (x_7y_7))/3$$

Tutte's Planar Algorithm Complexity

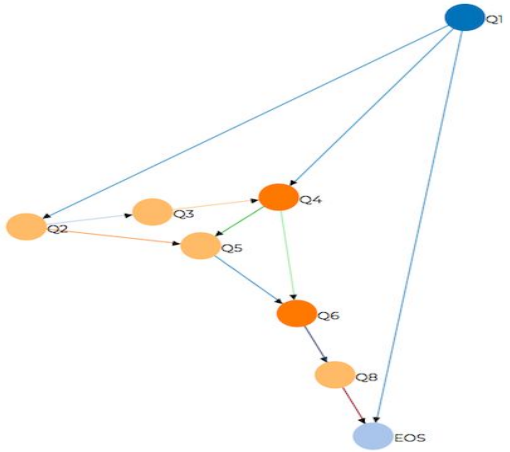
- Time Complexity
  - $O(V^3)$ solving linear system using LU Decomposion

- space complexity
  - $O(V^2)$ saving matrix

Visualization Result



Non-Planar Embedding by D3



Tutte's Planar Embedding

Export Graph as PDF

- Front End
  - user interface
  - send request
  - prompt download notification
- Back End
  - Node.js(Express) server
  - Librsvg convert graph to PDF
  - Return PDF to front end

OUTLINE

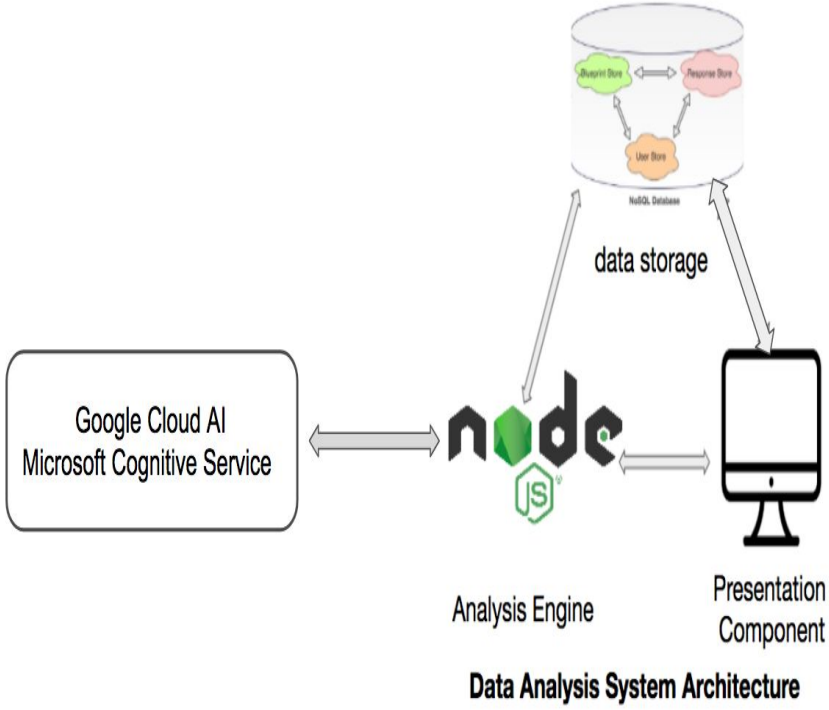
- Introduction
- Related Work
- Design & Implementation
  - Survey Visualization
  - Data Analysis
- Results
- Conclusion

Data Analysis Motivation

- TigerAware lacks ability to provide analysis
  - statistics function not supported(e.g. distribution)
  - advanced analysis function not supported(e.g. NLP)
  - difficult for researcher to draw conclusion
- third-party analytics software is expensive
  - Tableau, Zoho Analytics

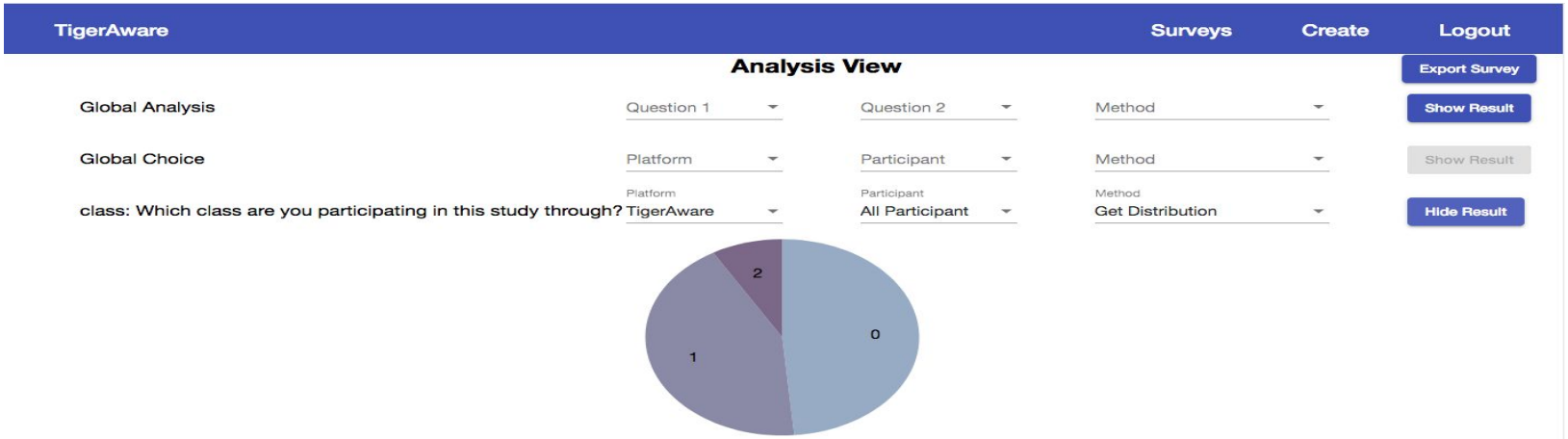
Data Analysis System

- Presentation Component
  - analysis page in TigerAware
- Analysis Engine
  - typical statistics
  - natural language process
  - computer vision
- Data Storage
  - hold survey data



Presentation Component Design

- A single web page in TigerAware Dashboard
- Designed to configure analysis parameters
  - platform, participant, and method.
- Communicate with business layer through HTTP

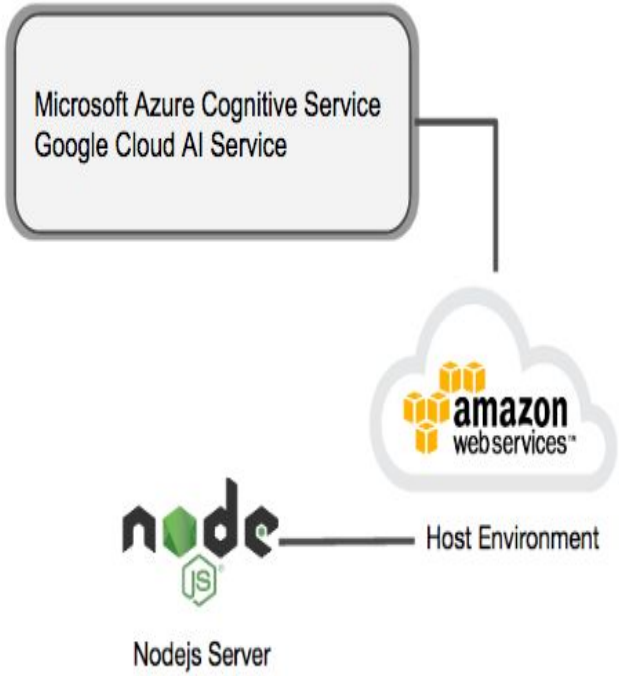


Presentation Component Implementation

- Data Access function
  - fetch survey data
  - interact with firebase through AngularFireDataBase
- Visualization
  - display questions, results
  - support pie chart, clock, word cloud, image,text
- Parameters Setting Module
  - initialize analysis parameters

Analysis Engine

- Provide analysis services
  - TigerAware service
  - Microsoft Azure Cognitive service
  - Google Cloud AI service
- Expose service through API
  - individual analysis API
  - group analysis API
  - export survey & response API



Business Layer Architecture

Analysis Engine - (1) TigerAware Service

Provide in-house analysis service

- word cloud
  - generate word cloud for free-text response
  - filter stop words, stemming
- response distribution
  - distribution for multiple choice question
- export survey & response
  - export survey and response as CSV file

Analysis Engine - (2) Microsoft Service

functions supported by Microsoft

- computer vision
  - emotion detection, image classification, landmark & celebrity detection, etc.
- natural language process
  - sentiment analysis, key phrase extraction, etc.

functions implemented in this project

- computer vision
  - emotion detection
- natural language process
  - sentiment analysis

Analysis Engine - (3) Google Service

functions supported by Google

- computer vision
  - emotion detection,label detection, landmark detection, text extraction, logo detection, etc.
- natural language process
  - sentiment analysis, content classification, entity analysis, syntax analysis, etc.

functions implemented in this project

- computer vision
  - emotion detection,label detection, landmark detection, text extraction, logo detection, etc.
- natural language process
  - sentiment analysis

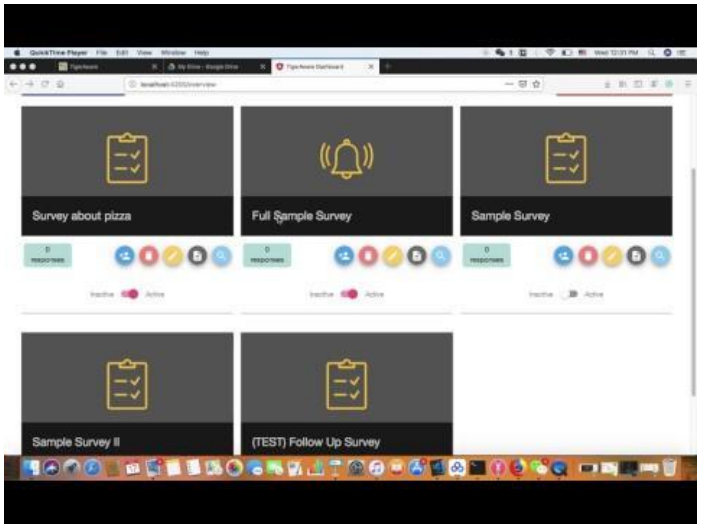
Data Storage

- Firebase realtime database is used in this project
  - Data is synchronized in realtime to every connected client
- survey is organized as a Json object
  - blueprints
  - data
  - users
  - etc

Demos

Survey Visualization Demo

TigerAware Service Demo



Demos

Google Cloud AI Service Demo

Microsoft Azure Cognitive Service Demo

Outline

- Introduction
- Related Work
- Data Overview
- Design & Implementation
- Demos
- Contribution & Future Work

Contribution

- Visualization component implemented in this project has better performance than state-of-the-art library D3

- Data analysis component provides both typical statistics function(e.g. distribution) and advanced analysis(e.g. sentiment analysis, emotion detection) for TigerAware System

Future Work

- For data analysis component, currently only pie chart is supported, more charts can be supported in future

- For in-house tigeraware service, more functions(e.g. sentiment analysis, emotion detection) need be implemented

Thank You!

Questions?

University of Missouri