

**A NATURAL LANGUAGE APPLICATION FOR PREDICTION OF HIGHEST RISKS IN A GOOGLE MAP**

By Rumela Saraswati and Sunny Anand



December 7, 2015

the University of Texas at dallas

Richardson, Dallas

ARCHITECTURAL DIAGRAM

**STAGE 2**

Collect the data in a corpus XML file

Crawl each of the links

Grab the headlines from HTML/URL links

**Get Safety Prediction**

**Enter Location**

USER

Sentence Tokenize

Find all unique action words

Define the verb tag list

Find all action component (verbs or nominals)

POS Tag using NLTK POS Tagger

Search Pattern ‘Accident Report’

Websites

Data Sources

For each Search Pattern collect the Headlines from the websites

**STAGE 1**

Manually analyze the file to extract the features

Create the list of safety prediction for each action word

Create observation file

Using Demo tool to determine Semantic Role

**STAGE 3**

Extract the predicate, nominal using SRL Demo tool

Create the CSV file with the features

Make the road safety prediction

Determine the most frequently occurring cause for a particular action component

Determine the Type of prediction based on the action component

**Ouput from stage 2 and stage 3**

**STAGE 4**