

## CS4395 Web Crawler Project

### (1) How I created my knowledge base:

To create my knowledge base, I first decided that my topic was going to be about bowling. I had initially wanted to do a knowledge base on racing games, but I believe that was either too broad a topic to get the kinds of information that I wanted or I picked a starting link that gave me too broad of information. After grabbing my urls, I scraped the text from each site by finding all the p tags and getting the text from there, as this would get me the bulk of the information that I needed while ignoring most of the irrelevant information. I then did a simple cleaning of the data by removing any newlines and tabs from it. For getting the top 40 terms from the dataset, I decided to use tf-idf as there would naturally be words that appear in all the documents like “bowling” that I wouldn’t want to include as I wanted to include words that would be more specific to certain sites. Using tf-idf ensured that I would be getting more relevant words as opposed to using just term frequency. In picking the 10 relevant terms, I scanned through my collected data and tried to pick words that covered topics that I considered to be either important or interesting. I also tried to consider the content of each of the urls and gave some preference to specific topics that had a url completely dedicated to that specific topic (for instance, I included the term “Truman” since in my dataset there was a url dedicated to the bowling alley built in the white house). Finally, for the knowledge base, I simply took my 10 relevant terms and made them keys of a dict. For the values of the dict, I included any sentences that had one of the relevant terms in it and added it to that key’s value set.

Knowledge base screenshots (on the left is the relevant word associated with the sentence):

```
target: Bowling is a target sport and recreational activity in which a player rolls a ball toward pins (in pin bowling) or another target (in target bowling).
target: The term bowling usually refers to pin bowling (most commonly ten-pin bowling), though in the United Kingdom and Commonwealth countries, bowling could
target: Bowling pins are the target of the bowling ball in pin bowling variations.
target: In target bowling, the aim is usually to get the ball as close to a mark as possible.
alley: In 1908, the now-oldest surviving bowling alley for the tenpin sport was opened in Milwaukee, Wisconsin, in the basement of the Holler House tavern, co
alley: In 1909, the first ten-pin bowling alley in Europe was installed in Sweden, but the game failed to catch on in the rest of Europe until after World War
alley: M.Scott Mahaskey/POLITICO The bowling alley in the basement of the EEOB.
alley: By M. SCOTT MAHASKEY 05/12/2014 01:22 PM EDT The Harry S. Truman Bowling Alley is located in the basement of the Eisenhower Executive Office Building.
```

```
ball: When hit by the ball, synthetic pins usually sound different from wooden pins.
ball: In target bowling, the aim is usually to get the ball as close to a mark as possible.}
ball: M.Scott Mahaskey/POLITICO A guest toes the line while delivering a ball.
ball: Visitors also can see a display on pinboys, a behind-the-scenes look at a television bowling production, murals commemorating member
pins: By this time, too, there were many variations of "pin" games, and also of games where a ball was thrown at objects other than pins.
pins: In the 1880s, Brunswick Corporation (founded 1845) of Chicago, Illinois, maker of billiard tables began making bowling balls, pins,
pins: Scoring rules are the same as the 10-pin game, except bowlers receive a third shot to convert pins left after their second-ball spar
pins: (Ex: Bocce Ball, an Italian lawn game) The earliest known forms of bowling date back to ancient Egypt,[5] with wall drawings depict
pins: Once again, the main difference between ten pin and candlepin is the size of the balls and pins.
```

shoe: In performance or competitive bowling shoes, the sliding shoe features a sole that will allow the bowler easily slide during their shoe: The braking bowling shoe has a sole that provides traction and is usually made of a higher friction material, such as rubber.

shoe: While it's probably obvious that you don't want too tight of a fit, it's also important to avoid a bowling shoe that is too loose.

shoe: The shoe for sliding is the one opposite of the hand with which you bowl.

egypt: A British anthropologist, Sir Flinders Petrie, discovered in the 1930's a collection of objects in a child's grave in Egypt that

egypt: [6][7] Remnants of bowling balls were found among artifacts in ancient Egypt going back to the Egyptian protodynastic period in 3

egypt: This set of ball games may be traced back as far as ancient Egypt.

egypt: (Ex: Bocce Ball, an Italian lawn game) The earliest known forms of bowling date back to ancient Egypt,[5] with wall drawings dep

egypt: [8] What is thought to be a child's game involving porphyry (stone) balls, a miniature trilithon, and nine breccia-veined alabas

branham: Branham's family moved from Detroit to the San Fernando Valley in 1978.

## Top 10 terms:

- target
- alley
- ball
- pins
- shoe
- egypt
- branham
- pluckhahn
- truman
- bwaa

(2) sample dialogue that could be created with a chatbot based on the knowledge base

### Initial dialogue

Hello! I am Layne, a bowling chatbot! What is your name?

*I am Laura*

Hi Laura! Do you like bowling?

*Yes, I like going to the bowling alley and playing a few rounds*

Great! What sort of ball do you use?

*I just use one at the bowling alley*

I see, did you know [fact about bowling ball]

I did not, that is interesting

What score do you usually get when playing?

*I'm not too good, so I usually only get an 80 or 90*

That's too bad. Do you play candle pin, 10 pin, or duck pin?

*What is duck pin?*

Duck pin uses pins that are shorter than traditional pins

*Oh I see, I play 10 pin*