

## I. Abstract

A web crawler is a program that takes a starting URL and jumps from link to link extracting data. I created a web crawler that searches links for information related to Nessie: the monster of Loch Ness. The starting point for the crawler was the Wikipedia page for Nessie found at [https://en.wikipedia.org/wiki/Loch\\_Ness\\_Monster](https://en.wikipedia.org/wiki/Loch_Ness_Monster).

## II. Creation of Knowledge Base

### A. Selection of Key Words

The web crawler starts at the Wikipedia page for the Loch Ness monster and jumps to a new link based on a few criteria that I modified as the project went on. I noticed certain pages were no longer in use and some links simply went to images. BeautifulSoup was used to extract paragraphs of data from the web pages meeting my URL criteria. The raw data for each page was stored in its own file.

These raw text files were then scraped and cleaned: removed of their new line characters and excessive spaces and placed into new files. Then, I created a function to get important words from the corpus, or the collection of scraped pages. The important terms were determined using TD-IDF. I determined using domain knowledge that the most important terms out of the TD-IDF terms were 'sightings', 'elusive', 'photographs', 'appearing', 'police', 'books', 'collection', 'nessie', 'day', and 'since'. I created the knowledge base around these terms.

### B. Implementation

In order to create the knowledge base, I utilized SQLite3. SQLite is a tool used to create database files with tables within them. Using this tool, I created a database called "database.db" and inserted a table called "nessie". The schema for this table is simple:

```
CREATE TABLE IF NOT EXISTS nessie (  
    id int NOT NULL AUTO_INCREMENT,  
    term varchar(20) NOT NULL,  
    context varchar(100) NOT NULL,  
);
```

This is shown in this line of code:

```
cur.execute('CREATE TABLE IF NOT EXISTS nessie (id INTEGER PRIMARY KEY AUTOINCREMENT, term TEXT NOT NULL, '  
            'context TEXT)') # put the table schema in the db
```

The program then goes through a hardcoded list of important words and grabs sentences that contain those words. It inserts the sentences into the nessie table shown here:

```
for sentence in sentences:
    if word in sentence:
        cur.execute("INSERT INTO nessie (term, context) VALUES(?, ?)", (word, sentence))
        connection.commit()
```

The result is a database with 1 table containing almost 240 rows. Each row has a unique id and can be queried later based on the associated important word or “term” like this:

```
SELECT context,
FROM nessie,
WHERE term = 'sightings'
```

Here is what the finished knowledge base looks like:

Name	Type	Schema
▼ Tables (2)		
▼ nessie		CREATE TABLE nessie (id INTEGER PRIMARY KEY AUTOINCREMENT, term TEXT NOT NULL, context TEXT)
id	INTEGER	"id" INTEGER
term	TEXT	"term" TEXT NOT NULL
context	TEXT	"context" TEXT

Table: nessie		
Filter in any column		
	id	term
	Filter	Filter
1	1 sightings	the scientific community explains alleged sightings of the loch ness monster as hoaxes, wishful thinking, and the misidentification of mundane objects.
2	2 sightings	public interest skyrocketed, with countless letters being sent in detailing different sightings describing a "monster fish," "sea serpent," or "dragon," with the final name ultimately settling on "loch ness monster."
3	3 sightings	ronald binns considers that this is the most serious of various alleged early sightings of the monster, but all other claimed sightings before 1933 are dubious and do not prove a monster tradition before that date.
4	4 sightings	this sighting triggered a massive amount of public interest and an uptick in alleged sightings, leading to the solidification of the actual name "loch ness monster."
5	5 sightings	it has been claimed that sightings of the monster increased after a road was built along the loch in early 1933, bringing workers and tourists to the formerly isolated area.
6	6 sightings	another photograph seemed to depict a horned "gargoyle head", consistent with that of some sightings of the monster; however, sceptics point out that a tree stump was later filmed during operation deepscan in 1987, which bore a striking ...
7	7 sightings	a number of explanations have been suggested to account for sightings of the creature.
8	8 sightings	binns does not call the sightings a hoax, but "a myth in the true sense of the term" and states that the "monster is a sociological ... phenomenon.
9	9 sightings	[better source needed] although some sightings describe a v-shaped wake similar to a boat's, others report something not conforming to the shape of a boat.
10	10 sightings	eels are found in loch ness, and an unusually large one would explain many sightings.
11	11 sightings	sightings in 1856 of a "sea-serpent" (or kelpie) in a freshwater lake near leurbost in the outer hebrides were explained as those of an oversized eel, also believed common in "highland lakes".
12	12 sightings	many scientists now believe that giant eels account for many, if not most of the sightings.
13	13 sightings	in a 1982 series of articles for new scientist, maurice burton proposed that sightings of nessie and similar creatures may be fermenting scots pine logs rising to the surface of the loch.
14	14 sightings	however, no sightings of the monster were reported in 1755.
15	15 sightings	despite setbacks (including lucy falling to the bottom of the loch), about 600 sightings were reported where she was placed.
16	16 sightings	robert rines explained that the "horns" in some sightings function as breathing tubes (or nostrils), allowing it to breathe without breaking the surface.
17	17 sightings	according to holiday, this explains the land sightings and the variable back shape; he likened it to the medieval description of dragons as "worms".
18	18 elusive	a single frame was published in his 1961 book, the elusive monster.
19	19 photographs	evidence of its existence is anecdotal, with a number of disputed photographs and sonar readings.
20	20 photographs	a survey of the literature about other hoaxes, including photographs, published by the scientific american on 10 july 2013, indicates many others since the 1930s.
21	21 photographs	although 21 photographs were taken, none was considered conclusive.
22	22 photographs	concurrent with the sonar readings, the floodlit camera obtained a pair of underwater photographs.
23	23 photographs	british naturalist peter scott announced in 1975, on the basis of the photographs, that the creature's scientific name would be nessiteras rhombopteryx (greek for "ness inhabitant with diamond-shaped fin").
24	24 photographs	some of the photographs, despite their obviously murky quality and lack of concurrent sonar readings, did indeed seem to show unknown animals in various positions and lightings.
25	25 appearing	in october 1871 (or 1872), d. mackenzie of balnain reportedly saw an object resembling a log or an upturned boat "wriggling and churning up the water," moving slowly at first before disappearing at a faster speed.
26	26 appearing	bartender david munro reported a wake he believed was a creature zigzagging, diving, and reappearing; there were reportedly 26 other witnesses from a nearby car park.
27	27 appearing	wind conditions can give a choppy, matte appearance to the water with calm patches appearing dark from the shore (reflecting the mountains).

### III. Sample Dialog

Bot: Hello! I'm Nessie the Loch Ness monster bot! What is your name?

User: Hi! I'm Katie.

B: Nice to meet you, Katie! Have you ever been to Scotland?

User: No, I haven't.

B: That's too bad! Have you ever seen any photographs of me?

User: Yes, I have seen a lot.

B: So, you know how beautiful I am! I look like a giant eel with legs! Do you think I am pretty?

User: No, I don't think giant eels are pretty.

B: Oh no! What about smart ones? Did you know I have been being spotted since the early 1930's?

User: I did not know that.