

# Web Crawler Report

## (1) describe how you created your knowledge base:

I used a dictionary to create my knowledge base. I computed the TF-IDF of the terms in 15 documents from relevant links on creating homebrew games for the GameBoy. After I computed the TF-IDF of all of the terms, I found 10 relevant terms in the top 40 terms and made a knowledge base out of them.

For the key, value pair of the knowledge base, the key was the relevant term, and the value was a list of every sentence in the corpus that contained the relevant term.

I stored the knowledge base in human readable form as a .txt file, and pickled it to store it as a binary file.

## Screenshots of the knowledge base:

```
1 {'assembly': ['# game boy color bootstrap rom [ ] unused palettes [ ] '
2               'colorization palettes in the bios [ ] rom disassembly [ ] # '
3               'hardware dmg schematics [ ] - hardware schematics .',
4               '# misc introduction to game boy hacking [ ] - workshop '
5               'introducing basic assembly , debugging and reverse engineering '
6               '.',
7               'binjgb on the web [ ] - porting of the binjgb emulator to web '
8               'assembly .',
9               'porting a go game boy emulator to webassembly [ ] about swotgb '
10              '[ ] - notes about the development of a game boy emulator in '
11              'javascript .',
12              'wiz [ ] - a high-level assembly language for writing homebrew '
13              'on retro console platforms ( game boy , nes , atari 2600 , and '
14              'more ) .',
15              'gbasm-rs [ ] - an opinionated rust based compiler for game boy '
16              'z80 assembly code .',
17              'gbasm [ ] - a javascript based compiler for game boy z80 '
18              'assembly code .',
```

# Web Crawler Report

```

2862 'gbdk': ['# compilers gbdk [ ] - maintained and modernized gbdk ( game boy '
2863         'development kit ) powered by an updated version of the sdcc '
2864         'toolchain .',
2865         'gbdk-go [ ] - a compiler translates go programs to c code .',
2866         'the output c code is built into gb rom by gbdk .',
2867         'complete list of open source emulators # tools # engines zgb [ ] - '
2868         'a little engine for creating games for the original game boy ( '
2869         'expands gbdk , more info here [ ] ) .',
2870         'retr0 gb [ ] - an engine for creating games ( expands gbdk ) .',
2871         '# development tools gbextended [ ] - c library extending gbdk .',
2872         'gbdk-lib-extension [ ] - a small set of sources and tools for the '
2873         'game boy development kit by michael hope .',
2874         'works with gbdk-2020 and rgbds .',
2875         'gbtiles [ ] - converts .gbr files created with harry mulder 's '
2876         'tile designer ( gbtd ) and .gbm files created with harry mulder 's '
2877         'map builder ( gbmb ) to different formats for use with the game boy '

```

```
'gameboy': ['creating photo realistic images with neural networks and a '
'gameboy camera [ ] the game boy printer [ ] - an in-depth '
'technical document about the printer hardware , the '
'communication protocol and the usual routine that games used for '
'implementing the print feature .',
'game boy cartridge pcb photos [ ] # custom cartridges emulating '
'a gameboy cartridge [ ] - emulating the functionality of a game '
'boy cartridge with the development board stm32f4 .',
'homebrew-gameboy-cartridge [ ] - eagle library , schematic , '
'and board files for a cartridge pcb using an atmel at49f040 as '
'rom .',
'homebrew gameboy color cartridge [ ] - board layout for an '
'eprom powered cartridge .',
'gameboy-mbc5-mbc1-hybrid [ ] - cpld implementation of a '
'mbc5/mbc1 hybrid cartridge .',
"the difference between the 80 and the newboy [ ]"]
```

Top 10 Relevant terms:

```
["gameboy", "game", "score", "gbdk", "development", "nes", "software", "programming",  
"assembly", "tutorials"]
```

# Web Crawler Report

(2) write up a sample dialog you would like to create with a chatbot based on your knowledge base

Sample dialog:

USER: Are GameBoy games written in **assembly** or some other language?

CHATBOT RESPONSE: Gameboy games are written *'in a c-like language called gbl , with many other features like tile and map extraction , wla-dx assembly , and more .'*

USER: Can you tell me about the **gameboy** architecture?

CHATBOT RESPONSE: Sure! The I have information about the *'gameboy specification , including info about important memory locations , the cpu 's design , and definitions for the complete z80 instruction set .'*

USER: are there any **Programming** guides you can tell me about making games for the GameBoy?

CHATBOT RESPONSE: Sure, I can tell you about *'programming guides , tutorials and tools to develop software for game boy using the development toolchains '*

## REFERENCES:

[How To Solve 403 Forbidden Errors When Web Scraping | ScrapeOps](#)

[NLP — Text Summarization using NLTK: TF-IDF Algorithm | by Akash Panchal | LessenText.com | Towards Data Science](#)

[TF-IDF from scratch in python on a real-world dataset. | by William Scott | Towards Data Science](#)

[TF IDF | TFIDF Python Example](#)

[5 maximum values in a python dictionary - Stack Overflow](#)

[TF\(Term Frequency\)-IDF\(Inverse Document Frequency\) from scratch in python . | by Yassine Hamdaoui | Towards Data Science](#)

# Web Crawler Report

[How can I use pickle to save a dict \(or any other Python object\)? - Stack Overflow](#)

[Write a dictionary to a file in Python - GeeksforGeeks](#)