



# Web Crawler

*Written by:*

*Erik Rodriguez and Arielle Posadas*

## Overview

We chose Elden Ring as our topic for the web crawler assignment. Our base URL was: <https://www.pcgamer.com/elden-ring-guide>. The program scraped the first 20 relevant links from the base URL and created 20 files containing the text. We initially thought of using external links as well, however, we found that the links outside of the URL did not contain relevant information.

Next, we used NLTK's sentence tokenizer to extract each sentence to a list. Afterwards, text from each file was preprocessed by converting everything to lowercase, removing stop words and punctuation, tokenizing the sentence into words, lemmatizing each word, and joining the words back into the sentence.

Next, we calculated the term frequency (TF) across all files and created dictionaries. Each key is added to the vocabulary set. Once the TF dictionary was done, the inverse document frequency (IDF) uses each TF dictionary and the vocabulary set to create the IDF dictionary. Finally, we use each TF dictionary and the IDF dictionary to create the TF-IDF dictionary.

Finally, we built the knowledge base by manually picking ten terms from our dictionary and scrapping a sentence from the Elden Wiki page for each term. We decided to do this to give context about the definition of the terms.

## Program Output

### Scrapped Links

```
~/Pr/NLP-Portfolio/05 - Webcrawler  master *2 +21 !1 ?3  python3 web-crawl.py
https://www.pcgamer.com/lets-speculate-way-too-much-about-the-elden-ring-expansion-image/
https://www.pcgamer.com/elden-ring/
https://www.pcgamer.com/all-elden-ring-quests-side-best/
https://www.pcgamer.com/elden-ring-character-creation/
https://www.pcgamer.com/all-elden-ring-classes-origins/
https://www.pcgamer.com/best-elden-ring-keepsakes-starter-gifts/
https://www.pcgamer.com/elden-ring-spirit-calling-bell-location/
https://www.pcgamer.com/all-elden-ring-quests-side-best/
https://www.pcgamer.com/elden-ring-tips-guide/
https://www.pcgamer.com/elden-ring-crafting-kit-cookbook-recipes/
https://www.pcgamer.com/how-to-level-up-elden-ring-melina-location-refuse-accept/
https://www.pcgamer.com/elden-ring-fun-builds/
https://www.pcgamer.com/best-elden-ring-weapons-list/
https://www.pcgamer.com/best-elden-ring-armor-sets-location/
https://www.pcgamer.com/best-elden-ring-spells-magic-sorceries-incantations/
https://www.pcgamer.com/how-to-respec-elden-ring-larval-tear/
https://www.pcgamer.com/best-elden-ring-ashes-of-war/
https://www.pcgamer.com/how-to-use-elden-ring-great-rune-arc/
https://www.pcgamer.com/best-elden-ring-talisman-pouch-slots-locations/
https://www.pcgamer.com/where-to-get-smithing-stones-elden-ring/

len of vocab: 2855
len of idf_dict: 2855
```

### Top 30 Terms using TF-IDF

```

len of vocab: 2855
len of all_tf_dicts: 20
len of idf_dict: 2855
len of tf_idf: 2855

Top 30 terms using the tf-idf dictionary:
1      ('cookbook', 0.14929366712149064)
2      ('smithing', 0.0612598028604092)
3      ('rune', 0.044841008648671694)
4      ('talisman', 0.04183713934279121)
5      ('melina', 0.040776100244349155)
6      ('armorer', 0.03732341678037266)
7      ('larval', 0.0348261324215716)
8      ('warrior', 0.03314187747342243)
9      ('calling', 0.03128732220006665)
10     ('creation', 0.03126828799419518)
11     ('tba', 0.031102847316977215)
12     ('rich', 0.030052666394676655)
13     ('somber', 0.02874541879172956)
14     ('affinity', 0.025357968459606132)
15     ('tear', 0.02457745612232908)
16     ('upgrade', 0.02071632910482161)
17     ('unique', 0.020271736877793798)
18     ('amber', 0.01940993726432116)
19     ('activate', 0.018710674510147244)
20     ('missionary', 0.01866170839018633)
21     ('dlc', 0.018587946696944486)
22     ('shadow', 0.01815733789315427)
23     ('ancient', 0.01754382060255896)
24     ('bearing', 0.017247251275037734)
25     ('reallocate', 0.015781041994385757)
26     ('incantation', 0.015095436027500745)
27     ('divine', 0.015072918315150497)
28     ('spread', 0.015046613047584789)
29     ('sorcery', 0.01473405948158148)
30     ('lone', 0.014686114332492931)

```

## Knowledge Base

smithing: You can visit the Upgrades page to learn more about the Upgrade Mechanics of Elden Ring.

melina: Melina wears a black robe and approaches players while they are resting at different Sites of Grace.

talisman: They're often recognizable by their bolster text (e.g.

larval: Core of a creature of mimicry known as a silver tear. As much as a substance as it is a living organism.

cookbook: Return to Fextralife Return to wiki home Ask on Discord Contact Us Accept Terms and Save Close Preview Close Editor Contributions to Fextralife Wikis are licensed under a Custom License.

rune: The compass at the top of the screen indicates the direction of the lost runes.

affinity: See Status Effects for all status buildups and other effects applied to your character.

dlc: Elden Ring features free DLC in the form of an update released in December 2022, that introduced the Colosseum battles, where players can engage in new pvp modes and six player multiplayer.

divine: See the page Divine Tower of Caelid for an explanation of how to get in there; the Divine Tower of Liurnia is associated to Ranni's questline.

arcs: Rune Arc boost is lost upon death.

## Possible Chatbot Output

Since the information obtained was sourced from guides and the Elden ring wiki we would use this to help players progress on their journey.



Hello, I am Elden bot! How may I assist you today?



I just started with the game but I have heard of a character called Melina. Who is she?



Melina wears a black robe and approaches players while they are resting at different Sites of Grace.



Oh I see. I was also wondering if Elden Ring features any expansions?



Elden Ring features free DLC in the form of an update released in December 2022, that introduced the Colosseum battles, where players can engage in new PVP modes and six player multiplayer..