**Name: Roshni Bethi**

**Intern ID: 425**

**Homoglyph Link Checker (Tool1)**

**What is a Homograph Link Checker?**

A Homograph Link Checker is a security tool designed to detect URLs that visually imitate legitimate websites by using look-alike characters from different alphabets or number substitutions. These deceptive URLs are often used in phishing attacks to trick users into clicking malicious links.

**How Does a Homograph Checker Work?**

A homograph link checker follows these steps:

**1. Input Handling**

* Accept a URL input from the user (like https://www.go0gle.com)
* Optional: normalize the URL (add https:// if missing)

**2. Domain Extraction**

* Extract just the domain part from the full URL:
  + Input: https://www.g00gle.com
  + Extracted: g00gle.com

**3. Character Analysis**

* Scan for:
  + Numeric substitutions: 0, 1, 3, 5, 8 replacing letters
  + Look-alike characters: Unicode Cyrillic, Greek, or other language characters that resemble Latin
* Use Unicode libraries or string matching for character detection

**4. Reachability Check (Optional)**

* Use an HTTP client (e.g., requests, fetch, curl, etc.)
* Attempt to make a request to the URL
* Report:
  + If the site is reachable (status 200 OK)
  + If it fails or times out (suspicious or broken)

**5. Alert the User**

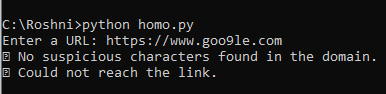
* If suspicious characters or unreachable domains are found, show warning:
  + " Suspicious URL! May be a phishing site."
  + " URL looks clean and reachable."

**Purpose**

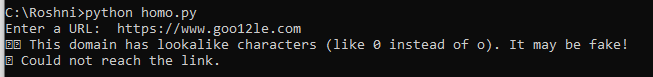
* Identify suspicious domain names
* Prevent phishing or scam attacks
* Alert users to fake websites that look real
* Help improve online safety

**How to Improve It (Advanced Ideas)**

1. Use Punycode Decoding to detect internationalized domains (xn-- format)
2. Compare domain with known safe list (google.com, amazon.in, etc.)
3. Integrate with VirusTotal or threat intelligence APIs
4. Add GUI with buttons using Tkinter (Python) or Electron (JS)
5. Export reports of suspicious URLs scanned



For a fake URL like https://www.goo9le.com, it doesn’t detect suspicious characters (probably because “9” isn’t in the script’s suspicious-character list) but fails to reach the link.



For another fake URL like https://www.goo12le.com, it does detect suspicious characters (likely because of “0” instead of “o”) and warns that the domain may be fake.

**Summary Description (You Can Use Anywhere):**

A Homograph Link Checker is a program that analyzes URLs to detect deceptive domain names designed to look like trusted websites. By identifying numeric and Unicode-based character substitutions (like g00gle.com instead of google.com), it helps protect users from phishing attacks and fraud. The checker works in any programming language by extracting domain names, analyzing characters, and optionally checking if the site is real and reachable.