Tool Name:

By: P.Harini ; **Intern ID:** 385

Homoglyph Link Checker

Description:

This is a Python-based tool for digital investigations, that is detecting suspicious hyperlinks in text files by identifying homoglyph domain attacks, where links seem to be something, but turn out to be something completely different

What Is This Tool About?

This tool addresses cybersecurity and threat intelligence needs by scanning text files for URLs and distinguishing legitimate domains from deceptive lookalikes (e.g., google.com vs. google.com) often exploited in phishing and fraud. It automates the detection of such links where attackers use multilingual/Unicode tricks.

Key Characteristics / Features:

1. Parses any .txt or plaintext file for domains and links.
2. Detects visually confusable Unicode characters (homoglyphs) in URLs.
3. Highlights suspicious/fake links based on script mixing (Cyrillic, Greek, etc.).
4. Supports a wide range of homoglyph scripts (English, Russian, Greek, etc.).
5. Outputs both full link lists and flagged suspect entries.
6. Command-line interface for quick use and automation.
7. Lightweight and requires only Python and the homoglyphs library.
8. Portable, works on Windows, macOS, and Linux.
9. Clear reporting to facilitate incident response or awareness training.
10. Easily extensible for integration with other investigation pipelines.

Types / Modules Available:

- CLI tool for batch scan of text files
- Core detection engine (homoglyphs module)
- Output/report generation

How Will This Tool Help?

- Immediate identification of homograph phishing domains in bulk communications and logs
- Protects users and organizations from credential theft via disguised URLs
- Accelerates cyber threat triage in phishing investigations
- Enables routine pre-send checks for suspicious links in email or chat
- Empowers security awareness training via real examples

Proof of Concept (PoC)

– Example Usage (text format):

Input file:

```
Visit https://google.com for updates.
But beware of: http://google.com/fake  (uses Greek omicron)
And www.facebook.com/login (Cyrillic o)
```

Command:

```
python homoglyph_checker.py testlinks.txt
```

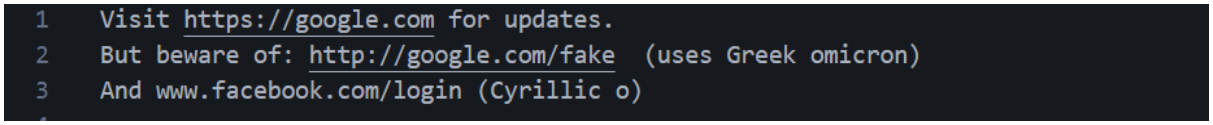
Sample Output:

```
Links found (3):
   https://google.com
   http://google.com/fake
   www.facebook.com/login

Suspicious (potentially fake) links:
   http://google.com/fake
   www.facebook.com/login
```

Screenshots:

Input file (text.txt):



Using Script and output:

```
PS D:\Programming\Programming\Python> python3 .\findhomoglyphs.py .\text.txt
Links found (3):
   https://google.com
   http://google.com/fake
   www.facebook.com/login

Suspicious (potentially fake) links:
   http://google.com/fake
   www.facebook.com/login
PS D:\Programming\Programming\Python>
```

A brief summary:

1. Reads any plain text file
2. Extracts links/domains using regex
3. Scans for confusable Unicode
4. Catches phishing-style disguised domains
5. Supports English, Russian, Greek and Latin (English contains Latin too) lookalikes
6. CLI for fast integration
7. Works on any desktop OS
8. Low resource, minimal dependencies
9. Output: all links + suspicious links
10. Easy to extend for other scripts
11. Useful for defence and awareness
12. Automation-ready
13. No installation required except dependencies
14. Makes phishing audits scalable
15. Maintained, updatable with more scripts if needed

Time to Use / Best Case Scenarios:

- Checking suspicious email dumps
- Auditing readme.txt, chat logs, or bulk content
- After a phishing attempt is reported
- Before whitelisting/allowlisting links
- As part of CI/CD pipeline for documentation/security

When to Use During Investigation:

- During triage of reported phishing incidents
- While auditing user communications for malicious links
- As part of red-teaming/CTI (threat intelligence) exercises
- In anti-phishing user education sessions

Best Person to Use This Tool & Required Skills:

SOC analyst,

Cybersecurity engineer,

IT support, or

Digital forensics examiner.

Required Skills:

- Basic Python and command line usage
- Understanding common phishing patterns
- Familiarity with Unicode character sets (helpful but not required)

Flaws / Suggestions to Improve:

- Limited to text file inputs; add support for HTML/email formats
- May not catch all 0-day or future script tricks; needs regular updates
- No context analysis (e.g., user behaviour after click)
- Add option to auto-generate "confidence score" per link
- Integrate with threat intel feeds for higher precision

Good About the Tool:

- Very lightweight and portable
- Fills a niche in anti-phishing automation
- Clear, actionable reporting
- Simple for any security practitioner to use
- Drastically reduces manual review effort
- Can double as awareness/training tool

In summary:

The homoglyph link checker is essential for detecting phishing and social engineering attempts leveraging the complexity of Unicode. It automates the boring yet critical job of visually inspecting links, making digital investigations faster and safer for everyone.