Name : Shreya Rajesh Dubey **Intern ID**:241

Tool Name

Homoglyph Detector

Description: What is this tool about?

This tool is designed to detect homoglyph characters—Unicode characters that visually resemble Latin letters but are actually different characters. These homoglyphs can be used in phishing, code obfuscation, and spoofing attacks.

Characteristics / Features

- Scans text files for suspicious homoglyph characters.
- Identifies the Unicode character, its Latin lookalike, and its position in the line.
- Provides detailed output for every line containing homoglyphs.
- Re-encodes standard output to UTF-8 to handle non-ASCII characters properly.

Types / Modules Available

- homoglyph_map: Dictionary mapping Unicode homoglyphs to Latin equivalents.
- get_homoglyph_reasons(line): Function to find homoglyphs in a string.
- detect_homoglyphs_verbose(file_path): Main function to read file and report homoglyphs.

How Will This Tool Help?

- Enhances security audits by detecting potential spoofing attempts in text/code.
- Helps identify obfuscated code or malicious input.
- Supports forensic investigations and malware analysis involving Unicode tricks.

Example Usage

Run the script from the command line with a suspicious input file:

Input File:

```
g00gle.com
youtube.com
faceb00k.com
amazon.com
google.com
```

Command:

```
python homoglyph.py input.py
```

Expected Output :

```
Line 2: youtube.com is suspicious
  -> Character 'o' at position 1 mimics 'o'
Line 5: google.com is suspicious
  -> Character 'o' at position 1 mimics 'o'
  -> Character 'o' at position 2 mimics 'o'
```

Best Case Scenarios

- Reviewing source code or logs for Unicode obfuscation.
- Auditing domain names or usernames in phishing investigations.
- Validating input fields in web applications.

People Who Can Use the Tool

- Security analysts
- Malware researchers
- Forensic investigators
- IT Support

Required Skills

- Basic Python knowledge
- Familiarity with Unicode encoding
- Understanding of security concerns around spoofing and obfuscation

---

Flaws

- Limited to hardcoded homoglyphs; may miss newer or more obscure Unicode tricks.
- No automatic file sanitation or correction.
- Only works on plain text files.
- No GUI or interactive options.

Suggestions to Improve

- Expand homoglyph_map with more Unicode spoofing cases.
- Allow input through command-line arguments instead of hardcoded filenames.
- Add output options (e.g., save report to file).
- Integrate with email/domain scanners or CI/CD pipelines.
- Add confidence scores or risk assessment per line.

Good

- Simple and lightweight.
- Easy to understand and extend.
- Useful in real-world security applications.

Summary

The homoglyph.py script is a lightweight, effective Python tool for identifying suspicious Unicode characters in text files. It's especially useful for detecting spoofing and obfuscation in security-sensitive contexts. While basic in its current form, it provides a strong foundation for further development into a robust security utility.