

TDB DUPLICATE CHECKER

Research Project:

TDB DUPLICATE CHECKER

Author:

Sufyan Qadir

Table of Content

1. Introduction	2
2. Problem Statement	2
3. Methods and Tools	2
3.1. Approach	2
4. Tests and Results	3
5. Reference	4

1. Introduction

The tdb file's function is to store numerous translation units along with various languages. However, there is a problem when some translation units have duplicate English tags in the database. The cost of managing the TDB file increases as a result. Because the tdb file can contain more than 100,000 lines, manually finding these issues takes some time. **Automating this procedure and incorporating it into the system would be beneficial in terms of saving time and money.**

2. Problem Statement

The database contains **duplicate English tags** for some translation units. As a result, management of the TDB file becomes more expensive.

3. Methods and Tools

3.1. Approach

Possible Solution	<p>We write a Python script that will find these duplicates and save them in a text file. This may be incorporated into the system by the DevOps developers:</p> <ul style="list-style-type: none">• Select file path• Keep id and English tags in a dictionary.• To find duplicates, use collections.• Enter the outcome in the text file.
Tools	<ul style="list-style-type: none">• Python (collections, re)• Gitlab

4. Tests and Results

The file was opened and copied to the terminal as expected. The parenRegular expression was used to match the parent tag, its child tag, and the parent tag's id attribute. The IDs in the text file were successfully compared to the IDs in the database, and duplicates were found.

5. Reference

- *Python | Find keys with duplicate values in dictionary. (2019, March 8). GeeksforGeeks.*
<https://www.geeksforgeeks.org/python-find-keys-with-duplicate-values-in-dictionary/>

Internal Use Only

Internal Use Only