

Library Kata „Dateidubletten aufspüren“

Entwickle eine Bibliothek, mit der man Dateidubletten in einem Verzeichnisbaum finden kann.

Der Kontrakt der Bibliothek sollte so aussehen:

```
interface IDublettenprüfung {
    IEnumerable<IDublette> Sammle_Kandidaten(string pfad);
    IEnumerable<IDublette> Sammle_Kandidaten(string pfad, Vergleichsmodi);

    IEnumerable<IDublette> Prüfe_Kandidaten(IEnumerable<IDublette>);
}

interface IDublette {
    IEnumerable<string> Dateipfade {get;}
}

enum Vergleichsmodi {
    Größe_und_Name,
    Größe
}
```

Zuerst wird die Methode *Sammle_Kandidaten()* aufgerufen. Sie durchläuft alle Dateien in einem Verzeichnisbaum und vergleicht Dateien nur sehr grob. Standardmäßig werden Dateiname und -größe verglichen, auf Wunsch aber auch nur die Dateigröße. Dateien, die danach für gleich gehalten werden, liefert die Funktion als *IDublette* zurück.

In einem zweiten Pass werden dann potenzielle Dubletten auf ihre tatsächliche Gleichheit geprüft. Jetzt wird der komplette Inhalt repräsentiert durch einen **[MD5 Hash]** verglichen. Dateien, die nun gleich sind, werden wieder als *IDublette* zurückgeliefert [1].

Variation #1

Ergänze *IDublettenprüfung* um einen Mechanismus, um den Fortschritt der Methoden zu melden.

Ressourcen

[MD5 Hash] http://de.wikipedia.org/wiki/Message-Digest_Algorithm_5

Endnoten

[1] Bedenke, dass es in einer Liste von Kandidaten Gruppen von tatsächlichen Dubletten geben kann.