

Hashsuche in Listen

Autor & Copyright: Dipl.-Ing. Harald Nahrstedt

Version: 2016 / 2019 / 2021 / 365

Erstellungsdatum: 01.08.2010

Überarbeitung: 01.12.2023

Quelle: Eigene Anwendung

Beschreibung:

Die Hashsuche ist ein effektives Verfahren zum Durchsuchen von umfangreichen, nicht geordneten Listen. Die Methode entstand 1953-56 bei IBM. Sie ist auch unter dem Namen Gestreute Speicherung bekannt

Anwendungs-Datei: 07-09-01_HashsucheListen.xlsm

1 Das Hash-Prinzip

Die Grundidee ist, dass Elemente einer Liste oder Datei durch ganzzahlige Schlüssel gekennzeichnet und mithilfe einer Funktion (Hashfunktion) einer Adresse (Hashindizes) zugeordnet werden (Bild 1).

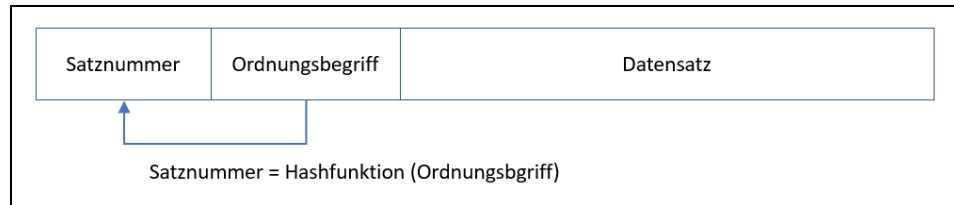


Bild 1. Das Hash-Prinzip

Ordnet die Hashfunktion zwei verschiedenen Schlüsselwerten die gleiche Adresse zu, so spricht man von Kollision. Da der Ordnungsbegriff nur eine Teilmenge des Datensatzes darstellt, sind Kollisionen zwangsläufig. Möglichst wenige Kollisionen erhält man, wenn als Hashfunktion die Divisionsreste bei Teilung durch eine Primzahl p gewählt werden, da die Reste $\text{mod}(p)$ stark streuen.

Methoden für Hashfunktionen:

- Divisionsrest-Verfahren
- Zerlegungsmethoden
- Basistransformationen
- Nichtnumerische Schlüssel
- Arboreszenzen (verallgemeinertes Divisionsrestverfahren)

Methoden bei Kollisionen:

- Lineare Kollisionsstrategie
- Kollisionsstrategie mittels Pseudozufallszahlen
- Quadratische Kollisionsstrategie
- Doppelhashverfahren

2 Suchen in einer technischen Liste

Wir suchen uns eine Liste, wie zum Beispiel die Liste Technische Regel für Dampfkessel (TRD). Zu finden in Wikipedia und ergänzen sie um einen Suchbegriff (Bild 2).

	A	B	C
1	1	Allgemeines, Aufbau und Anwendung	Allgemein
2	100	Allgemeine Grundsätze für Werkstoffe	Werkstoffe
3	101	Bleche	Bleche
4	102	Nahtlose und elektrisch preßgeschweißte Rohre aus Stahl	Rohre
5	103	Stahlguss	Stahlguß
6	106	Schrauben und Muttern aus Stahl	Stahlschrauben
7	107	Kesselteile aus Formstahl und Schmiedestücken	Kesselteile
8	108	Gusseisen mit Lamellengraphit und Gusseisen mit Kugelgraphit	Gußeisen
9	110	Armaturengehäuse	Armaturen
10	201	Schweißen von Bauteilen aus Stahl; Fertigung – Prüfung	Schweißprüfung
11	202	Prüfung von Fertigteilen aus Stahlblech	Fertigteile
12	203	Nahtlose Sammler und ähnliche Hohlkörper – Fertigung und Prüfung	Sammler
13	300	Festigkeitsberechnung von Dampfkesseln	Festigkeitsprüfung
14	301	Zylinderschalen unter innerem Überdruck	Zylinderschalen
15	303	Kugelschalen und gewölbte Böden unter innerem und äußerem Überdruck	Kugelschalen
16	304	Gewölbte Flammrohrböden	Flammrohrböden
17	305	Ebene Wandungen, Verankerungen und Versteifungsträger	Wandungen
18	306	Zylinderförmige Schalen unter äußerem Überdruck	Zylinderschalen

Bild 2. Beispiel einer technischen Liste

Als Hashfunktion wird zunächst jedes Zeichen des Suchbegriffs (in seiner ganzen Länge) in seinen ASCII-Wert umgewandelt. Aus der Quersumme wird dann mittels $\text{Adresse} = (\text{Summe} \bmod 41) + 1$ (1) die Speicheradresse ermittelt.

In einem ersten Schritt wollen wir einmal testen, wie viele Kollisionen die Formel erzielt. Dazu schreiben wir eine Prozedur, die in der Spalte D die Anzahl der Speicherzugriffe summiert. In einem Modul wird dazu die Prozedur nach Code 1 installiert.

Codeliste 1. HashTest im Modul modHash

```
Option Explicit

Sub HashTest()
    Dim objTab As Object
    Dim iMaxRow As Integer
    Dim i As Integer
    Dim j As Integer
    Dim sKey As String
    Dim iKey As Integer
    Dim lSum As Long

    Set objTab = ActiveSheet
    iMaxRow = objTab.UsedRange.Rows.Count
    For i = 1 To iMaxRow
        sKey = objTab.Cells(i, 3)
        lSum = 0
        For j = 1 To Len(sKey)
            lSum = lSum + Asc(Mid(sKey, j, 1))
        Next j
    Next i
End Sub
```

```

        Next j
        iKey = lSum Mod 41 + 1 'Hashfunktion
        objTab.Cells(iKey, 4) = objTab.Cells(iKey, 4) + 1 'Testeinträge
    Next i
    Set objTab = Nothing
End Sub

```

Diese Prozedur ermittelt für das Testbeispiel maximal 6 Kollisionen bei Armaturen und eine relativ gut gestreute Verteilung. Nun können wir damit beginnen, die Liste in Tabelle1 in eine nach dem Hashschlüssel geordnete umzustellen. Dazu benutzen wir die Tabelle2 und den nachfolgenden Code.

Codeliste 2. Hash Datentransfer im Modul modHash

```

Sub Hashttransfer()
    Dim objTab1 As Object
    Dim objTab2 As Object
    Dim iMaxRow As Integer
    Dim i As Integer
    Dim j As Integer
    Dim sKey As String
    Dim iKey As Integer
    Dim lSum As Long

    Set objTab1 = Sheets("Tabelle1")
    Set objTab2 = Sheets("Tabelle2")
    iMaxRow = objTab1.UsedRange.Rows.Count
    For i = 1 To iMaxRow
        sKey = objTab1.Cells(i, 3)
        lSum = 0
        For j = 1 To Len(sKey)
            lSum = lSum + Asc(Mid(sKey, j, 1))
        Next j
        iKey = lSum Mod 41 + 1 'Hashfunktion
        If objTab2.Cells(iKey, 1) = "" Then
            objTab2.Cells(iKey, 1) = objTab1.Cells(i, 1)
            objTab2.Cells(iKey, 2) = objTab1.Cells(i, 2)
            objTab2.Cells(iKey, 3) = objTab1.Cells(i, 3)
        Else
            'Kollision
            j = 42
            Do While Not objTab2.Cells(j, 1) = ""
                j = j + 1
            Loop
            objTab2.Cells(j, 1) = objTab1.Cells(i, 1)
            objTab2.Cells(j, 2) = objTab1.Cells(i, 2)
            objTab2.Cells(j, 3) = objTab1.Cells(i, 3)
        End If
    Next i
    Set objTab1 = Nothing
    Set objTab2 = Nothing
End Sub

```

Das Ergebnis zeigt Bild 3. Natürlich macht eine solche Anwendung erst bei einer großen Datenmenge Sinn.

	A	B	C
1	108	Gusseisen mit Lamellengraphit und Gusseisen mit K	Gußeisen
2	503	Prüfung vor Inbetriebnahme – Bauprüfung und Was	Bauprüfung
3	520	Richtlinie für das Verfahren der Erlaubnis zur Erricht	Prüfrichtlinien
4	411	Ölfeuerungen an Dampfkesselanlagen	Ölfeuerungen
5	101	Bleche	Bleche
6	412	Gasfeuerungen an Dampfkesseln	Gasfeuerungen
7	414	Holzfeuerungen an Dampfkesseln	Holzfeuerung
8	1	Allgemeines, Aufbau und Anwendung	Allgemein
9	509	Richtlinie für das Verfahren der Bauartzulassung vo	Richtlinie
10	452	Anlagen zur Lagerung von druckverflüssigtem Amm	Ammoniakkesselanlagen
11	300	Festigkeitsberechnung von Dampfkesseln	Festigkeitsprüfung
12	431	Rauchgas-Wasservorwärmer für Dampfkessel der C	Rauchgasvorwärmer
13	505	Wiederkehrende Prüfung – Äußere Prüfung	Äußenprüfung
14	460	Anlagen zur Verminderung von luftverunreinigender	Rauchgasreinigung
15	102	Nahtlose und elektrisch preßgeschweißte Rohre au	Rohre
16	309	Schrauben	Schrauben
17	201	Schweißen von Bauteilen aus Stahl; Fertigung – Prü	Schweißprüfung
18	203	Nahtlose Sammler und ähnliche Hohlkörper – Fertic	Sammler

Bild 3. Die Hash-Liste

Mit der dritten Prozedur *HashSuche* wird nun das Suchen mit der Hash-Methode getestet.

Codeliste 3. Hash Suche

```

Sub HashSuche ()
    Dim objTab As Object
    Dim sKey As String
    Dim iKey As Integer
    Dim lSum As Long
    Dim j As Integer

    Set objTab = Sheets("Tabelle2")
    sKey = InputBox("Schlüssel = ", "HASH-SUCHE")
    lSum = 0
    For j = 1 To Len(sKey)
        lSum = lSum + Asc(Mid(sKey, j, 1))
    Next j
    iKey = lSum Mod 41 + 1 'Hashfunktion
    If objTab.Cells(iKey, 3) = sKey Then
        MsgBox Str(iKey) & " / " & objTab.Cells(iKey, 2), _
            vbInformation, "GEFUNDEN"
    Else
        j = 42
        Do While Not objTab.Cells(j, 1) = ""
            If Trim(objTab.Cells(j, 3)) = sKey Then
                MsgBox Str(j) & " / " & objTab.Cells(j, 2), _
                    vbInformation, "GEFUNDEN"
            End If
            j = j + 1
        Loop
    End If
    Set objTab = Nothing
End Sub

```