

## #108 Textanalyse

Gegeben sind das Textfile *HIF Beschreibung.txt* .

**1. Aufgabe:** Lies das File ein und zerlege jede Zeile in einzelne Wörter.

Achte darauf, dass du unnötige Zeichen, wie Satzzeichen, Anführungszeichen, Ziffern, usw. mit geeigneten String-Methoden entfernst. Ebenso sind Jahreszahlen, Datumsangaben, usw. keine gültigen Wörter und sollen ignoriert werden.

Erstelle zwecks Übersichtlichkeit, eine statische Methode, die aus einer Zeile ein String-Array erzeugt und zurückgibt, dass nur Wörter enthält.

**2. Aufgabe:** Erstelle ein Dictionary, das für jedes Wort die Anzahl im Text mitzählt und füge alle Wörter des Texts in das Dictionary ein.

**3. Aufgabe:** Implementiere Methoden (auch als Extensionmethoden des Dictionaries möglich) für die Ermittlung der folgenden Werte:

- Anzahl der verschiedenen Wörter
- Anzahl der Gesamtanzahl der Wörter
- Ermittlung des häufigsten Worts und seiner Anzahl

**Hinweis:** Die Werte können zurückgegeben werden, entweder als

- ein KeyValuePair,
- zwei Outputparameter oder
- ein Value Tuple, wie es seit .NET Framework 4.7.1 unterstützt wird.

Siehe dazu <https://docs.microsoft.com/en-us/dotnet/csharp/tuples>

oder die Kurzgeschichte <https://www.youtube.com/watch?v=GL563msVGHI>

Gib alle Werte im Main übersichtlich aus.

**4. Aufgabe:** Definiere ein weiteres Dictionary, das für jeden Anfangsbuchstaben, die Liste aller verschiedenen Wörter speichert. Befülle dieses Dictionary ebenfalls während des Einlesens des Files.

**Achtung:** Die Liste darf keine doppelten Einträge enthalten.

Überprüfe vor dem Einfügen eines Worts in die Liste, ob es nicht schon enthalten ist.

**Achtung:** Es soll keine Unterscheidung zwischen Groß- und Kleinbuchstaben geben!

D. h. der Key im Dictionary soll nur 'a' oder 'A' sein und die Liste soll gleichermaßen Wörter mit großem und kleinem Anfangsbuchstaben speichern.

Z. B. Anfangsbuchstabe A -> Ausbildung, auch, Als, ...

Hinweis: Wörter *der* und *Der* (am Satzanfang) können als zwei Wörter gespeichert werden.

**5. Aufgabe:** Der Benutzer gibt im Main einen Buchstaben ein.

- Berechne unter Verwendung beider Dictionaries, wieviele Wörter es mit diesem Anfangsbuchstaben gibt und gib diese Anzahl aus.
- Gib das häufigste Wort mit diesem Anfangsbuchstaben und die Anzahl aus.

### Freiwillige Aufgabe:

Erstelle eine Klasse TextAnalyzer und definiere sinnvolle Methoden, so dass das Main eine sinnvolle Struktur hat.

Beispiel:

```
static void Main(string[] args)
{
    TextAnalyzer ta = new TextAnalyzer();
    ta.ReadFile("HIF Beschreibung.txt");
    // TODO: Ausgabe der Anzahl der verschiedenen Wörter
    // TODO: Ausgabe der Anzahl der gesamten Wortzahl
    // TODO: Ausgabe des häufigsten Worts und seiner Anzahl

    do
    {
        Console.Write("Geben Sie den Anfangsbuchstaben ein: ");
        if (!char.TryParse(Console.ReadLine(), out char letter))
        {
            Console.WriteLine("Die Eingabe war ungültig. Geben Sie bitte nur ein Zeichen ein.\n");
            continue;
        }

        // TODO: Ausgabe des häufigsten Worts mit dem Anfangsbuchstaben letter
        // TODO: Ausgabe der gesamten Anzahl von Wörtern mit dem Anfangsbuchstaben letter
        break;
    }
    while (true);

    Console.ReadKey();
}
```

Hinweis:

`char.TryParse` gibt `true` zurück, wenn ein Zeichen richtig geparsed wurde, also der Benutzer genau ein Zeichen eingegeben hat. Dann wird dieses Zeichen in den Output-Parameter `letter` kopiert.

`char.TryParse` gibt `false` zurück, wenn beim Parsen ein Fehler aufgetreten ist.