

Language-Integrated Query

Parameter	Kursinformationen
Veranstaltung:	Vorlesung Softwareentwicklung
Semester	Sommersemester 2022
Hochschule:	Technische Universität Freiberg
Inhalte:	LINQ Konzepte und Anwendung
Link auf den GitHub:	https://github.com/TUBAF-lfi-LiaScript/VL_Softwareentwicklung/blob/master/25_LINQ.md
Autoren	Sebastian Zug, Galina Rudolf & André Dietrich



Fragen aus den Projekten

Frage: Was ist eigentlich eine csv Datei?

Das Dateiformat CSV *comma-separated values* und beschreibt den Aufbau einer Textdatei zur Speicherung oder zum Austausch einfach strukturierter Daten. Die Dateinamenserweiterung lautet `.csv`.

Prüfungsliste.csv

```
# Prüfungen im SoSe
16.07.2022 8:00, Theoretische Physik, mündlich, 30min
18.07.2022 10:00, Technische Mechanik, schriftlich, 2h
...
```

Wie werten wir das Ganze aus?

csvExample1.cs9

```

1 using System;
2 using System.Collections.Generic;
3 using System.Collections;
4 using System.Linq;
5 using System.Text;
6 using Microsoft.VisualBasic.FileIO;
7
8 class Student{
9     public string name;
10    public int id;
11    public string topic;
12
13    public override string ToString(){
14        return $"{this.id} - {this.name}";
15    }
16 }
17
18 class ReadingCSV
19 {
20     public static void Main()
21     {
22         var path = @"Data.csv";
23         List<Student> studentList = new List<Student>();
24         using (TextFieldParser csvReader = new TextFieldParser(path))
25         {
26             csvReader.CommentTokens = new string[] { "#" };
27             csvReader.SetDelimiters(new string[] { "," });
28             csvReader.HasFieldsEnclosedInQuotes = true;
29             csvReader.ReadLine();
30
31             while (!csvReader.EndOfData)
32             {
33                 // Read current line fields, pointer moves to the next line
34
35                 string[] fields = csvReader.ReadFields();
36                 var newRecord = new Student
37                 {
38                     id = Int32.Parse(fields[0]),
39                     name = fields[1],
40                     topic = fields[2]
41                 };
42                 studentList.Add(newRecord);
43             }
44             foreach(var student in studentList){
45                 Console.WriteLine(student);
46             }
47         }
48     }

```

myproject.csproj

```

1 <Project Sdk="Microsoft.NET.Sdk">
2   <PropertyGroup>
3     <OutputType>Exe</OutputType>
4     <TargetFramework>net5.0</TargetFramework>
5   </PropertyGroup>
6 </Project>

```

Data.csv

```

1 StudentID, StudentName, Topic
2 1, Humboldt, Geography
3 2, Hardenberg, Geography
4 3, Rammler, Process engineering
5 4, Winkler, Chemistry
6 5, Reich, Chemistry

```

You must install or update .NET to run this application.

App: /tmp/tmpxocs1zk3/bin/Debug/net5.0/project
Architecture: x64
Framework: 'Microsoft.NETCore.App', version '5.0.0' (x64)
.NET location: /usr/lib/dotnet/dotnet6-6.0.110

The following frameworks were found:

6.0.10 at [/usr/lib/dotnet/dotnet6-6.0.110/shared/Microsoft.NETCore.App]

Learn about framework resolution:

<https://aka.ms/dotnet/app-launch-failed>

To install missing framework, download:

[https://aka.ms/dotnet-core-applaunch?](https://aka.ms/dotnet-core-applaunch?framework=Microsoft.NETCore.App&framework_version=5.0.0&arch=x64&rid=ubuntu.22.10-x64)

[framework=Microsoft.NETCore.App&framework_version=5.0.0&arch=x64&rid=ubuntu.22.10-x64](https://aka.ms/dotnet-core-applaunch?framework=Microsoft.NETCore.App&framework_version=5.0.0&arch=x64&rid=ubuntu.22.10-x64)

Alternative Umsetzung mit `DataFrames` [Link](#). Dafür muss das Paket `ML.NET` installiert werden (vgl. erweiterte Projektkonfigurationsdatei).

csvExample11.cs9

```
1 using System;
2 using System.Collections.Generic;
3 using Microsoft.Data.Analysis;
4
5 class ReadingCSV
6 {
7     public static void Main()
8     {
9         var path = @"Data.csv";
10        var students = DataFrame.LoadCsv(path, separator:',',header:true);
11
12        Console.WriteLine(students.Info());
13        var topicStat = students["Topics"].ValueCounts();
14        Console.WriteLine(topicStat);
15    }
16 }
```

myproject.csproj

```
1 <Project Sdk="Microsoft.NET.Sdk">
2   <PropertyGroup>
3     <OutputType>Exe</OutputType>
4     <TargetFramework>net5.0</TargetFramework>
5   </PropertyGroup>
6   <ItemGroup>
7     <PackageReference Include="Microsoft.Data.Analysis" Version="0.19.1" />
8   </ItemGroup>
9
10 </Project>
```

Data.csv

```
1 StudentID, StudentName, Topic
2 1, Humboldt, Geography
3 2, Hardenberg, Geography
4 3, Rammler, Process engineering
5 4, Winkler, Chemistry
6 5, Reich, Chemistry
```

You must install or update .NET to run this application.

App: /tmp/tmp5q7fgoe5/bin/Debug/net5.0/project
Architecture: x64
Framework: 'Microsoft.NETCore.App', version '5.0.0' (x64)
.NET location: /usr/lib/dotnet/dotnet6-6.0.110

The following frameworks were found:

6.0.10 at [/usr/lib/dotnet/dotnet6-6.0.110/shared/Microsoft.NETCore.App]

Learn about framework resolution:

<https://aka.ms/dotnet/app-launch-failed>

To install missing framework, download:

[https://aka.ms/dotnet-core-applaunch?](https://aka.ms/dotnet-core-applaunch?framework=Microsoft.NETCore.App&framework_version=5.0.0&arch=x64&rid=ubuntu.22.10-x64)

[framework=Microsoft.NETCore.App&framework_version=5.0.0&arch=x64&rid=ubuntu.22.10-x64](https://aka.ms/dotnet-core-applaunch?framework=Microsoft.NETCore.App&framework_version=5.0.0&arch=x64&rid=ubuntu.22.10-x64)

Motivation

Gegeben sei das Datenset eines Comic-Begeisterten in Form einer generischen Liste `List<T>`.

1. Bestimmen Sie die Zahl der Einträge unseres Datensatzes
2. Filtern Sie die Liste der Comic Figuren nach dem Alter und
3. Sortieren Sie die Liste nach dem Anfangsbuchstaben des Namens.

conventionalFiltering

```
1 using System;
2 using System.Collections.Generic;
3
4
5 public class Character{
6     protected string name;
7     public int geburtsjahr;
8     private static int Count;
9     int index;
10 public Character(string name, int geburtsjahr){
11     this.name = name;
12     this.geburtsjahr = geburtsjahr;
13     index = Count;
14     Count = Count + 1;
15 }
16 public override string ToString(){
17     string row = string.Format("{0,6} | {1,-15} | {2,8} |",
18                               index, name, geburtsjahr);
19     return row;
20 }
21 }
22
23 public class Program
24 {
25     public static void Main(string[] args){
26         List<Character> ComicHeros = new List<Character>{
27             new Character("Spiderman", 1962),
28             new Character("Donald Duck", 1931),
29             new Character("Superman", 1938)
30         };
31         Console.WriteLine("Alle Einträge in der Datenbank:");
32         Console.WriteLine("| Index | Name | Ursprung |");
33         foreach (Character c in ComicHeros){
34             Console.WriteLine(c);
35         }
36         // Und nun? Wie filtern wir?
37     }
38 }
```

Cannot open assembly 'main.cs': File does not contain a valid CIL image.

Die intuitive Lösung könnte folgendermaßen daher kommen:

Die Dokumentation von `List<T>` findet sich unter folgendem [Link](#)

1. Wir "erinnern" uns an das `Count` Member der Klasse `List`.
2. Für die Filteroperation implementieren Sie eine Loop. Sie können dazu `foreach` verwenden, weil `List<T>` das Interface `IEnumerable` implementiert.
3. Die Sortieroperation bedingt die Anwendung einer Vergleichsoperation zwischen den Elementen der Liste. Eine Variante ist die Implementierung des Interfaces `IComparable` zu diesem Zweck.

Solution

```
1 using System;
2 using System.Collections.Generic;
3
4
5 public class Character: IComparable{
6     protected string name;
7     public int geburtsjahr;
8     private static int Count;
9     int index;
10 public Character(string name, int geburtsjahr){
11     this.name = name;
12     this.geburtsjahr = geburtsjahr;
13     index = Count;
14     Count = Count + 1;
15 }
16 public override string ToString(){
17     string row = string.Format("{0,6} | {1,-15} | {2,8} |",
18         index, name, geburtsjahr);
19     return row;
20 }
21 public int CompareTo(object obj){
22     if (obj == null) return 1;
23     Character otherCharacter = obj as Character;
24     return string.Compare(this.name, otherCharacter.name);
25 }
26 }
27
28 public class Program
29 {
30     public static void Main(string[] args){
31         List<Character> ComicHeros = new List<Character>{
32             new Character("Spiderman", 1962),
33             new Character("Donald Duck", 1931),
34             new Character("Superman", 1938)
35         };
36         Console.WriteLine($"\\nEinträge in der Datenbank: {ComicHeros.Count}");
37         Console.WriteLine("\\nGefilterte Einträge in der Datenbank:");
38         Console.WriteLine("| Index | Name | Ursprung |");
39         List<Character> ComicHerosFiltered = new List<Character>();
40         foreach (Character c in ComicHeros){
41             if (c.geburtsjahr < 1950) ComicHerosFiltered.Add(c);
42         }
43         foreach (Character c in ComicHerosFiltered){
44             Console.WriteLine(c);
45         }
46         Console.WriteLine("\\nSortierte Einträge in der Datenbank:");
47         Console.WriteLine("| Index | Name | Ursprung |");
48         ComicHeros.Sort();
49         foreach (Character c in ComicHeros){
50             Console.WriteLine(c);
51         }
52     }
53 }
```

Cannot open assembly 'main.cs': File does not contain a valid CIL image.

Eine Menge Aufwand für einen simple Operation! Welche zusätzlichen Probleme werden auftreten, wenn Sie eine solche Kette aus Datenerfassung, Verarbeitung und Ausgabe in realen Anwendungen umsetzen?

Alternativ schauen wir uns weiter im Kanon der `List<T>` Klasse um und realisieren die Methoden `RemoveAll()` oder `Sort()`.

`RemoveAll()` zum Beispiel entfernt alle Elemente, die mit den Bedingungen übereinstimmen, die durch das angegebene Prädikat definiert werden. Interessant ist dabei die Umsetzung. Ein Prädikat ist eine generischer Delegat der einen Instanzen eines Typs `T` auf ein Kriterium hin evaluiert und einen Bool-Wert als Ausgabe generiert.

```
public int RemoveAll (Predicate<T> match);  
public delegate bool Predicate<in T>(T obj);
```

Analog kann `Sort()` mit einem entsprechenden Delegaten `Comparison` verknüpft werden.

```
public void Sort (Comparison<T> comparison);  
public delegate int Comparison<in T>(T x, T y);
```

Solution

```
1 using System;
2 using System.Collections.Generic;
3
4
5 public class Character: IComparable{
6     protected string name;
7     public int year;
8     private static int Count;
9     int index;
10 public Character(string name, int year){
11     this.name = name;
12     this.year = year;
13     index = Count;
14     Count = Count + 1;
15 }
16
17 public override string ToString(){
18     string row = string.Format("{0,6} | {1,-15} | {2,8} |",
19         index, name, year);
20     return row;
21 }
22
23 public int CompareTo(object obj){
24     if (obj == null) return 1;
25     Character otherCharacter = obj as Character;
26     return string.Compare(this.name, otherCharacter.name);
27 }
28 }
29
30 public class Program
31 {
32     private static bool before1950(Character entry)
33     {
34         return entry.year > 1950;
35     }
36
37     private static int sortByYear(Character x, Character y)
38     {
39         int output = 0;
40         if (y.year < x.year) output = 1;
41         if (y.year > x.year) output = -1;
42         return output;
43     }
44
45 public static void Main(string[] args){
46     List<Character> ComicHeros = new List<Character>{
47         new Character("Spiderman", 1962),
48         new Character("Donald Duck", 1931),
49         new Character("Superman", 1938)
50     };
51     ComicHeros.RemoveAll(before1950);
52     //ComicHeros.RemoveAll(x => x.year > 1950);
53     //ComicHeros.Sort(sortByYear);
54     foreach (Character c in ComicHeros){
55         Console.WriteLine(c);
56     }
57 }
58 }
```

Cannot open assembly 'main.cs': File does not contain a valid CIL image.

Allerdings bleibt die Darstellung von komplexeren Abfragen wie filtere die Helden heraus, die vor 1950 geboren sind, extrahiere die Vornamen und sortiere diese in Aufsteigender alphabetischer Folge zu einem unübersichtlichen Darstellungsformat.

Die Methoden für den Datenzugriff und die Manipulation abhängig vom Datentyp (Felder, Objektlisten) und der Herkunft (XML-Dokumente, Datenbanken, Excel-Dateien, usw.).

Welche alternativen Konzepte bestehen für die Verarbeitung von datengetriebenen Anfragen?

Exkurs SQL

Hier folgt ein kurzer Einschub zum Thema *Structured Query Language* (SQL) ... um allen Teilnehmern eine sehr grundlegende Sicht zu vermitteln:

SQL ist eine Datenbanksprache zur Definition von Datenstrukturen in relationalen Datenbanken sowie zum Bearbeiten (Einfügen, Verändern, Löschen) und Abfragen von darauf basierenden Datenbeständen.

Ausgangspunkt sind Datenbanktabellen, die Abfragen dienen dabei der Generierung spezifischer Informationssets:

Buchnummer	Autor	Verlag	Datum	Titel
123456	Hans Vielschreiber	Musterverlag	2007	Wir lernen SQL
123457	J. Gutenberg	Gutenberg und Co.	1452	Drucken leicht gemacht
123458	Galileo Galilei	Inquisition International	1640	Eppur si muove
123459	Charles Darwin	Vatikan Verlag	1860	Adam und Eva

- "Alle Bücher mit Buchnummern von 123400 bis 123500"
- "Alle Buchnummern mit Autoren, die im 19. Jahrhundert erschienen."
- "In welchem Jahrhundert veröffentlichte welcher Verlag die meisten Bücher?"
- ...

SQL basiert auf der relationalen Algebra, ihre Syntax ist relativ einfach aufgebaut und semantisch an die englische Umgangssprache angelehnt. Die Bezeichnung SQL bezieht sich auf das englische Wort "query" (deutsch: „Abfrage“). Mit Abfragen werden die in einer Datenbank gespeicherten Daten abgerufen, also dem Benutzer oder einer Anwendersoftware zur Verfügung gestellt. Durch den Einsatz von SQL strebt man die Unabhängigkeit der Anwendungen vom eingesetzten Datenbankmanagementsystem an.

SQL-Aufrufe sind deklarativ, weil der Entwickler hier nur das WAS und nicht das WIE festlegt. Dabei strukturieren sich die Befehle in 4 Kategorien:

- Befehle zur Abfrage und Aufbereitung der gesuchten Informationen
- Befehle zur Datenmanipulation (Ändern, Einfügen, Löschen)
- Befehle zur Definition des Datenbankschemas
- Befehle für die Rechteverwaltung und Transaktionskontrolle.

Eine Datenbanktabelle stellt eine Datenbank-Relation dar. Die Relation ist Namensgeber und Grundlage der relationalen Datenbanken.

Relationale Datenbank [\[DatenbankSchema\]](#)

Erzeugung der Tabellen

1	CREATE TABLE Student;	
2	INSERT INTO Student SELECT * from ?;	

student.csv		
1	MatrNr,Name	
2	26120,Fichte	
3	25403,Jonas	
4	27103,Fauler	

[
1,		
3		
]		

1	CREATE TABLE hoert;	
2	INSERT INTO hoert SELECT * from ?;	

hoert.csv		
1	MatrNr,VorlNr	
2	26120,5001	
3	25403,5001	
4	27103,5045	


```
[
  1,
  3
]
```

```
1 CREATE TABLE Vorlesung;
2 INSERT INTO Vorlesung SELECT * from ?;
```

vorlesung.csv

```
1 VorlNr,Titel,PersNr
2 5001,ET,15
3 5022,IT,12
4 5045,DB,12
```

```
[
  1,
  3
]
```

```
1 CREATE TABLE Professor;
2 INSERT INTO Professor SELECT * from ?;
```

prof.csv

```
1 PersNr,Name
2 12,Wirth
3 15,Tesla
4 20,Urлаuber
```

```
[
  1,
  3
]
```

Beispiele

Auslesen aller Spalten und aller Zeilen

```
1 SELECT *
2 FROM Student;
```

```
[
  {
    "MatrNr": "26120",
    "Name": "Fichte"
  },
  {
    "MatrNr": "25403",
    "Name": "Jonas"
  },
  {
    "MatrNr": "27103",
    "Name": "Fauler"
  }
]
```

Abfrage mit Spaltenauswahl

```
1 SELECT VorlNr, Titel
2 FROM Vorlesung;
```

```
[
  {
    "VorlNr": "5001",
    "Titel": "ET"
  },
  {
    "VorlNr": "5022",
    "Titel": "IT"
  },
  {
    "VorlNr": "5045",
    "Titel": "DB"
  }
]
```

Abfrage mit eindeutigen Werten

```
1 SELECT DISTINCT MatrNr
2 FROM hoert;
```

```
[
  {
    "MatrNr": "25403"
  },
  {
    "MatrNr": "26120"
  },
  {
    "MatrNr": "27103"
  }
]
```

Abfrage mit Filter und Sortierung

```
1 SELECT VorlNr, Titel
2 FROM Vorlesung
3 WHERE Titel = 'ET';
```

```
[
  {
    "VorlNr": "5001",
    "Titel": "ET"
  }
]
```

LIKE kann mit verschiedenen Platzhaltern verwendet werden: `_` steht für ein einzelnes beliebiges Zeichen, `%` steht für eine beliebige Zeichenfolge. Manche Datenbanksysteme bieten weitere solche Wildcard-Zeichen an, etwa für Zeichenmengen.

ORDER BY öffnet die Möglichkeit die Reihung anzupassen.

Verbund

```
1 SELECT Vorlesung.VorlNr, Vorlesung.Titel, Professor.PersNr, Professor.Name
2 FROM Professor
3 INNER JOIN Vorlesung ON Professor.PersNr = Vorlesung.PersNr;
```

```
[
  {
    "VorlNr": "5022",
    "Titel": "IT",
    "PersNr": "12",
    "Name": "Wirth"
  },
  {
    "VorlNr": "5045",
    "Titel": "DB",
    "PersNr": "12",
    "Name": "Wirth"
  },
  {
    "VorlNr": "5001",
    "Titel": "ET",
    "PersNr": "15",
    "Name": "Tesla"
  }
]
```

JOIN erlaubt es die Relationen zwischen einzelnen Datenbanktabellen aufzulösen. Dabei kann mit **INNER** und **OUTER** bzw. **LEFT** und **RIGHT** die Auswahl über der Schnittmenge beschrieben werden.

Gruppierung mit Aggregat-Funktionen

```
1 SELECT COUNT(Vorlesung.PersNr) AS Anzahl, Professor.PersNr, Professor.Name
2 FROM Professor LEFT JOIN Vorlesung
3 ON Professor.PersNr = Vorlesung.PersNr
4 GROUP BY Professor.Name, Professor.PersNr;
```

```
[
  {
    "Anzahl": 2,
    "PersNr": "12",
    "Name": "Wirth"
  },
  {
    "Anzahl": 1,
    "PersNr": "15",
    "Name": "Tesla"
  },
  {
    "Anzahl": 0,
    "PersNr": "20",
    "Name": "Urlauber"
  }
]
```

[DatenbankSchema] Wikipedia "SQL", Nils Boßung, <https://de.wikipedia.org/wiki/SQL#/media/Datei:SQL-Beispiel.svg>

LINQ Umsetzung

Language Integrated Query (LINQ) zielt auf die direkte Integration von Abfragefunktionen in die Sprache. Dafür definieren C# (wie auch VB.NET und F#) eigene Schlüsselwörter sowie eine Menge an vorbestimmten LINQ-Methoden. Diese können aber durch den Anwender in der jeweiligen Sprache erweitert werden.

LINQexample.cs

```
var query =
    from e in employees
    where e.DepartmentId == 5
    select e;
```

LINQ-Anweisungen sind unmittelbar als Quelltext in .NET-Programme eingebettet. Somit kann der Code durch den Compiler auf Fehler geprüft werden. Andere Verfahren wie *ActiveX Data Objects* ADO und *Open Database Connectivity* ODBC hingegen verwenden Abfragestrings. Diese können erst zur Laufzeit interpretiert werden; dann wirken Fehler gravierender und sind schwieriger zu analysieren.

Innerhalb des Quellprogramms in C# oder VB.NET präsentiert LINQ die Abfrage-Ergebnisse als streng typisierte Aufzählungen. Somit gewährleistet es Typsicherheit bereits zur Übersetzungszeit wobei ein minimaler Codeeinsatz zur Realisierung von Filter-, Sortier- und Gruppiervorgänge in Datenquellen investiert wird.

LINQ Anwendungsfelder ^{[[LinqAnbieter](#)]}

Merkmale von LINQ

- Die Arbeit mit Abfrageausdrücken ist einfach, da sie viele vertraute Konstrukte der Sprache C# verwenden.
- Alle Variablen in einem Abfrageausdruck sind stark typisiert, obwohl dieser in der Regel nicht explizit angegeben wird. Der Compiler übernimmt die Ableitung.
- Eine Abfrage wird erst ausgeführt, wenn Sie über der Abfragevariable iteriert wird. Folglich muss die Quelle in einer iterierbaren Form vorliegen.
- Zur Kompilierzeit werden Abfrageausdrücke gemäß den in der C#-Spezifikation festgelegten Regeln in Methodenaufrufe des Standardabfrageoperators konvertiert. Die Abfragesyntax ist aber einfacher zu lesen.
- LINQ kombiniert Abfrageausdrücke und Methodenaufrufe (`count` oder `max`). Hierin liegt die Flexibilität des Konzeptes.

Diese Veranstaltung konzentriert sich auf die *LINQ to Objects* Realisierung von LINQ. Dabei können Abfragen mit einer beliebigen `IEnumerable`- oder `IEnumerable<T>`-Auflistungen angewandt werden.

[[LinqAnbieter](#)] Wikimedia <https://commons.wikimedia.org/wiki/File:AnbieterLINQ.png>, Author 'Mussklprozz'

Exkurs "Erweiterungsmethoden"

Erweiterungsmethoden ergänzen den Umfang von bestehenden Methoden einer Klasse ohne selbst in diesem Typ deklariert worden zu sein. Man beschreibt eine statische Methode und ordnet diese einer Klasse über den Typ des ersten Parameters zu.

Merke: Erweiterungsmethoden stellen das bisherige Konzept der Deklaration von Klassen (etwas) auf den Kopf. Sie ermöglichen es zusätzliche Funktionalität "anzuhängen".

Das folgende Beispiel unterstreicht den Unterschied zur bereits vorgestellten Methode der partiellen Methoden, die eine verteilte Implementierung einer Klasse erlaubt. Hierfür muss der Quellcode vorliegen, die Erweiterungsmethode `Print()` kann auch auf eine Bibliothek angewandt werden.

Extensions.cs

```

1 using System;
2
3 // Ergänzung mit partiellen Implementierungen - Nur zur Abgrenzung im
4 // enthalten Beispiel
5
6 // .----- Explizite Erlaubnis zur Erweiterung
7 // v
8 public partial class MyPartitalString
9 {
10     public string content;
11     public MyPartitalString(string content)
12     {
13         this.content = content;
14     }
15 }
16
17 public partial class MyPartitalString
18 {
19     public void sayHello() => Console.WriteLine("Say Hello!");
20 }
21
22 ///////////////////////////////////////////////////
23 /
24 class MyString
25 {
26     public string content;
27     public MyString(string content)
28     {
29         this.content = content;
30     }
31 }
32
33 // Erweiterungsmethode in einer separaten Klasse
34 static class Exporter
35 {
36     public static void print(this MyString input, string newString)
37     {
38         Console.WriteLine(input.content + newString);
39     }
40 }
41
42 class Program
43 {
44     public static void Main(string[] args)
45     {
46         MyPartitalString text1 = new MyPartitalString("Bla fasel");
47         text1.sayHello();
48
49         MyString text2 = new MyString("Bla fasel");
50         text2.print("-Hossa");
51     }
52 }

```

Cannot open assembly 'main.cs': File does not contain a valid CIL image.

Erweiterungsmethoden schaffen uns die Möglichkeit weitere Funktionalität zu integrieren und gleichzeitig Datenobjekte durch eine Verarbeitungskette "hindurchzureichen". Erweitern Sie die statische Klasse doch mal um eine Methode, die dem Inhalt der Membervariable `content` zusätzlichen Information einfügt. Das Ganze ist natürlich noch recht behäbig, weil wir zwingend von einem bestimmten Typen ausgehen. Dies lässt sich über eine generische Implementierung lösen.

GenericExtensions.cs

```
1 using System;
2
3 abstract class myAbstractString{
4     public string content;
5     public myAbstractString(string content)
6     {
7         this.content = content;
8     }
9     public void sayClassName() => Console.WriteLine(this.GetType().Name);
10 }
11
12 class myString: myAbstractString
13 {
14     public myString(string content): base(content) {}
15 }
16
17 class yourString: myAbstractString
18 {
19     public yourString(string content): base(content) {}
20 }
21
22 static class Exporter
23 {
24     public static void Print<T>(this T input) where T: myAbstractString
25     {
26         Console.WriteLine(input.content);
27         input.sayClassName();
28     }
29 }
30
31 class Program
32 {
33     public static void Main(string[] args)
34     {
35         myString A = new myString("Bla fasel");
36         A.Print();
37         yourString B = new yourString("Bla blub");
38         B.Print();
39     }
40 }
```

Cannot open assembly 'main.cs': File does not contain a valid CIL image.

Sie können Erweiterungsmethoden verwenden, um eine Klasse oder eine Schnittstelle zu erweitern, jedoch nicht, um sie zu überschreiben. Entsprechen wird eine Erweiterungsmethode mit dem gleichen Namen und der gleichen Signatur wie eine Schnittstellen- oder Klassenmethode nie aufgerufen.

Exkurs "Anonyme Typen"

Anonyme Typen erlauben die Spezifikation eines Satzes von schreibgeschützten Eigenschaften, ohne zunächst explizit einen Typ definieren zu müssen. Der Typname wird dabei automatisch generiert.

Anonyme Typen enthalten mindestens eine schreibgeschützte Eigenschaft, alle anderen Arten von Klassenmitgliedern sind ausgeschlossen.

Extensions.cs

```
1 using System;
2
3 //class Irgendwas{
4 //    public string text;
5 //    public int zahl;
6 //}
7
8 class Program
9 {
10     public static void Main(string[] args)
11     {
12         //          .---- Hier steckt der Unterschied - keine Typangabe
13         //          v
14         var v = new {text = "Das ist ein Text", zahl = 1};
15         Console.WriteLine($"text = {v.text}, zahl = {v.zahl}");
16         Console.WriteLine(v);
17         //v.text = "asfsa";
18         Console.WriteLine($"type = {v.GetType().Name}");
19         var myPropertyInfo = v.GetType().GetProperties();
20         Console.WriteLine("\nProperties:");
21         for (int i = 0; i < myPropertyInfo.Length; i++)
22         {
23             Console.WriteLine(myPropertyInfo[i].ToString());
24         }
25     }
26 }
```

Cannot open assembly 'main.cs': File does not contain a valid CIL image.

Der Vorteil anonymer Typen liegt in ihrer Flexibilität. Die eigentlichen Daten werden entsprechend den Ergebnissen einer Funktion erzeugt.

Exkurs "Enumerables"

```
public interface IEnumerable<out T> : System.Collections.IEnumerable{
    public IEnumerator<T> GetEnumerator();
}

public interface IEnumerator<out T> : IDisposable,
    System.Collections.IEnumerator{
    public object Current { get; }
    public bool MoveNext ();
    public void Reset ();
}
```

Zur Wiederholung soll nochmals ein kurzes Implementierungsbeispiel gezeigt werden. An dieser Stelle wird eine Klasse `myStrings` umgesetzt, die als Enumerationstyp realisiert werden soll. Entsprechend implementiert die Klasse `IEnumerable` das Interface `IEnumerable<string>` und referenziert einen Enumeratorentyp `StringEnumerator`, der wiederum das Interface generische Interface `IEnumerator<string>` umsetzt.

Transformieren Sie folgendes Codefragment in eine UML Darstellung.

GenericEnumerable

```

1 using System;
2 using System.Collections;
3 using System.Collections.Generic;
4
5 class myStrings : IEnumerable<string>{
6     // eigentliche Daten
7     public string [] str_arr = new string[] { "one" , "two" ,"three", "four"
8         , "five"};
9
10    // "Verwaltungsoverhead"
11    public IEnumerator<string> GetEnumerator()
12    {
13        IEnumerator<string> r = new StringEnumerator(this);
14        return r ;
15    }
16    IEnumerator IEnumerable.GetEnumerator()
17    {
18        return GetEnumerator() ;
19    }
20 }
21 class StringEnumerator : IEnumerator<string>{
22     int index;
23     myStrings sp;
24
25     public StringEnumerator (myStrings str_obj){
26         index = -1 ;
27         sp = str_obj ;
28     }
29
30     object IEnumerator.Current{
31         get
32         { return sp.str_arr[ index ] ; }
33     }
34
35     public string Current{
36         get
37         { return sp.str_arr[ index ] ; }
38     }
39
40     public bool MoveNext( ){
41         if ( index < sp.str_arr.Length - 1 ){
42             index++ ;
43             return true ;
44         }
45         return false ;
46     }
47
48     public void Reset( ){
49         index = -1 ;
50     }
51
52     public void Dispose(){
53         // pass
54     }
55 }
56
57 class Program {
58     public static void Main(string[] args){
59         myStrings spp = new myStrings();
60         foreach( string i in spp)
61             System.Console.WriteLine(i);
62     }
63 }

```

Cannot open assembly 'main.cs': File does not contain a valid CIL image.

Welchen Vorteil habe ich verglichen mit einer nicht-enumerate Datenstruktur, zum Beispiel einem array? Im Hinblick auf eine konkrete Implementierung ist zwischen dem Komfort der erweiterten API und den Performance-Eigenschaften abzuwägen.

Einen Überblick dazu bietet unter anderem die Diskussion unter <https://stackoverflow.com/questions/169973/when-should-i-use-a-list-vs-a-linkedlist/29263914#29263914>

LINQ - Grundlagen

Sie können LINQ zur Abfrage beliebiger aufzählbarer Auflistungen wie `List<T>`, `Array` oder `Dictionary<TKey, TValue>` verwenden. Die Auflistung kann entweder benutzerdefiniert sein oder von einer .NET Framework-API zurückgegeben werden.

Alle LINQ-Abfrageoperationen bestehen aus drei unterschiedlichen Aktionen:

- Abrufen der Datenquelle
- Erstellen der Abfrage
- Ausführen der Abfrage

Für ein einfaches Beispiel, das Filtern einer Liste von Zahlenwerten realisiert sich dies wie folgt:

LINQBasicExample

```
1 using System;
2 using System.Threading;
3 using System.Collections.Generic;
4 using System.Linq;
5
6 class Program {
7     public static void Main(string[] args){
8         // Spezifikation der Datenquelle
9         int[] scores = new int[] { 55, 97, 92, 81, 60 };
10        // Definition der Abfrage
11        IEnumerable<int> scoreQuery =
12            from score in scores // Bezug zur Datenquelle
13            where score > 80     // Filterkriterium
14            select score;       // "Projektion" des Rückgabewertes
15        // Execute the query.
16        foreach (int i in scoreQuery)
17        {
18            Console.Write(i + " ");
19        }
20    }
21 }
```

Cannot open assembly 'main.cs': File does not contain a valid CIL image.

Datenquellen

Zugriff	Bedeutung
LINQ to Objects	Zugriff auf Objektlisten und -Hierarchien im Arbeitsspeicher
LINQ to SQL	Abfrage und Bearbeitung von Daten in MS-SQL-Datenbanken
LINQ to Entities	Abfrage und Bearbeitung von Daten im relationalen Modell von ADO.NET
LINQ to XML	Zugriff auf XML-Inhalte
LINQ to DataSet	Zugriff auf ADO.NET-Datensammlungen und -Tabellen
LINQ to SharePoint	Zugriff auf SharePoint-Daten

Im Rahmen dieser Veranstaltung konzentrieren wir uns auf die LINQ to Objects Variante.

Query Ausdrücke

Insgesamt sind 7 Query-Klauseln vorimplementiert, können aber durch Erweiterungsmethoden ergänzt werden.

Ausdruck	Bedeutung
from	definieren der Laufvariable und einer Datenquelle
where	filtert die Daten nach bestimmten Kriterien
orderby	sortiert die Elemente
select	projiziert die Laufvariable auf die Ergebnisfolge
group	bildet Gruppen innerhalb der Ergebnismenge
join	vereinigt Elemente mehrere Datenquellen
let	definiert eine Hilfsvariable

```
class Student{
    public string Name;
    public int Id;
    public string Subject{get; set;}
    public Student(){}
}

// Collection Initialization
List<Student> students = new List<Student>{
    new Student("Max Müller"){Subject = "Technische Informatik", id = 1},
    new Student("Maria Maier"){Subject = "Softwareentwicklung", id = 2},
    new Student("Martin Moraweschek"){Subject = "Höhere Mathematik I", id = 3}
}

// Implizite Typdefinition
var result = from s in students           // Spezifikation der Datenquelle
              where s.Subject == "Softwareentwicklung"
              orderby s.Name
              select new (s.Name, s.Id)   // Projektion der Ausgabe

// explizite Typdefinition
IEnumerable<Student> result = from s in students
                              ...
```

Im vorangehenden Beispiel ist `students` die Datenquelle, über der die Abfrage bearbeitet wird. Der List-Datentyp implementiert das Interface `IEnumerable<T>`. Die letzte Zeile bildet das Ergebnis auf die Rückgabe ab, dem Interface entsprechen auf ein `IEnumerable<Student>` mit den Feldern Name und Id.

Die Berechnung der Folge wird nicht als Ganzes realisiert sondern bei einer Iteration durch den Datentyp `List<Student>`.

Für nicht-generische Typen (die also `IEnumerable` anstatt `IEnumerable<T>` unmittelbar) implementieren, muss zusätzlich der Typ der Laufvariable angegeben werden, da diese nicht aus der Datenquelle ermittelt werden kann.

StudentListExample

```

1 using System;
2 using System.Collections.Generic;
3 using System.Linq;
4
5 public class Student
6 {
7     public string FirstName { get; set; }
8     public string LastName { get; set; }
9     public int[] Scores { get; set; }
10 }
11
12 class Program {
13     public static void Main(string[] args){
14         //ArrayList StudentList = new ArrayList(); <-- Nicht mehr benutzen
15         List<Student> StudentList = new List<Student>();
16         StudentList.Add(
17             new Student{
18                 FirstName = "Svetlana", LastName = "Müller", Scores = new int[]
19                 { 98, 92, 81, 60 }
20             });
21         StudentList.Add(
22             new Student {
23                 FirstName = "Claire", LastName = "O'Donnell", Scores = new
24                 int[] { 75, 84, 91, 39 }
25             });
26         var query = from student in StudentList
27                     where student.Scores[0] > 95
28                     select student;
29         foreach (Student s in query)
30             Console.WriteLine(s.LastName + ": " + s.Scores[0]);
31     }
32 }

```

Cannot open assembly 'main.cs': File does not contain a valid CIL image.

Welche Struktur ergibt sich dabei generell für eine LINQ-Abfrage? Ein Query beginnt immer mit einer `from`-Klausel und endet mit einer `select` oder `group`-Klausel.

Allgemeingültig lässt sich, entsprechend den Ausführungen in [Mössenböck](#) folgende Syntax ableiten:

```

QueryExpr =
    "from" [Type] variable "in" SrcExpr
    QueryBody
QueryBody =
    { "from" [Type] variable "in" SrcExpr
    | "where" BoolExpr
    | "orderby" Expr ["ascending" | "descending"] {"," Expr ["ascending" |
    "descending"]}
    | "join" [Type] variable "in" SrcExpr "on" Expr "equals" Expr ["into"
    variable]
    | "let" variable "=" Expr
    }
    ( "select" ProjectionExpr ["into" variable QueryBody]
    | "group" ProjectionExpr "by" Expr ["into" variable QueryBody]
    ).

```

Mit der isolierten Definition der Abfragen können diese mehrfach auf die Daten angewandt werden. Man spricht dabei von einer "verzögerten Ausführung" - jeder Aufruf der Ausgabe generiert eine neue Abfrage.

DelayedEvaluation

```
1 using System;
2 using System.Threading;
3 using System.Collections.Generic;
4 using System.Linq;
5
6 class Program
7 {
8     public static void Main(string[] args){
9         var numbers = new List<int>() {1,2,3,4};
10        // Spezifikation der Abfrage
11        var query = from x in numbers
12                    select x;
13        Console.WriteLine(query.GetType());
14        // Manipulation der Daten
15        numbers.Add(5);
16        Console.WriteLine(query.Count());
17        // Manipulation und erneute Anwendung der Abfrage
18        numbers.Add(6);
19        Console.WriteLine(query.Count()); // 6
20    }
21 }
```

Cannot open assembly 'main.cs': File does not contain a valid CIL image.

Hinter den Kulissen

Der Compiler transformiert LINQ-Anfragen in der Abfragesyntax in Lambda-Ausdrücke, Erweiterungsmethoden, Objektinitializer und anonyme Typen. Dabei sprechen wir von der Methodensyntax. Abfragesyntax und Methodensyntax sind semantisch identisch, aber viele Benutzer finden die Abfragesyntax einfacher und leichter zu lesen. Da aber einige Abfragen nur in der Methodensyntax möglich sind, müssen sie diese bisweilen nutzen. Beispiele dafür sind `Max()`, `Min()`, oder `Take()`.

Nehmen wir also nochmals eine Anzahl von Studenten an, die in einer generischen Liste erfasst wurden:

```
List<Student> students = new List<Student>({
    new Student{
        Id = "123sdf234"
        FirstName = "Svetlana",
        LastName = "Omelchenko",
        Field = "Computer Science",
        Scores = new int[] { 98, 92, 81, 60 }
    };
    //...
});

var result = from s in students
              where s.Field == "Computer Science"
              orderby s.LastName
              select new {s.LastName, s.Id};
```

Der Compiler generiert daraus folgenden Code:

```
IEnumerable<Student> result = students
    .Where(s => s.Field == "Computer Science" )
    .OrderBy(s => s.LastName)
    .Select(s => new {s.LastName, s.Id});
```

Wieso hat meine Klasse `Student` plötzlich eine Methode `where`? Hier nutzen wir eine automatisch generierte Erweiterungsmethoden.

Dabei wird die eigentliche Filterfunktion als Delegat übergeben, dies wiederum kann durch eine Lambdafunktion ausgedrückt werden. <https://docs.microsoft.com/de-de/dotnet/api/system.linq.enumerable.where?view=netframework-4.8>

Dabei beschreiben die Lambdafunktionen sogenannten Prädikate, Funktionen, die eine bestimmte Bedingung prüfen und einen boolschen Wert zurückgeben.

WhereExample

```
1 using System;
2 using System.Threading;
3 using System.Collections.Generic;
4 using System.Linq;
5
6 class Program {
7     public static bool filterme(int num){
8         bool result = false;
9         if (num > 10) result = true;
10        return result;
11    }
12
13    public static void Main(string[] args){
14        int[] numbers = { 0, 30, 20, 15, 90, 85, 40, 75 };
15        //Func<int, bool> filter = delegate(int num) { return num > 10; };
16        Func<int, bool> filter = filterme;
17        IEnumerable<int> query =
18            numbers.Where(filter);
19        //IEnumerable<int> query =
20        //    numbers.Where(s => s > 10);
21        foreach (int number in query)
22        {
23            Console.WriteLine(number);
24        }
25    }
26 }
```

Cannot open assembly 'main.cs': File does not contain a valid CIL image.

Basisfunktionen von LINQ

Mit LINQ lassen sich Elementaroperationen definieren, die dann im Ganzen die Mächtigkeit des Konzeptes ausmachen.

Filtern

Das Beispiel zur Filterung einer Customer-Tabelle wurde der C# Dokumentation unter <https://docs.microsoft.com/de-de/dotnet/csharp/programming-guide/concepts/linq/basic-linq-query-operations> entnommen.

Die üblichste Abfrageoperation ist das Anwenden eines Filters in Form eines booleschen Ausdrucks. Das Filtern bewirkt, dass im Ergebnis nur die Elemente enthalten sind, für die der Ausdruck eine wahre Aussage liefert.

Das Ergebnis wird durch Verwendung der `where`-Klausel erzeugt. Faktisch gibt der Filter an, welche Elemente nicht in die Quellsequenz eingeschlossen werden sollen. In folgendem Beispiel werden nur die customers zurückgegeben, die eine Londoner Adresse haben.

```
var queryLondonCustomers = from customer in customers
                             where customer.City == "London"
                             select customer;
```

Sie können die logischen Operatoren `&&` und `||` verwenden, um so viele Filterausdrücke wie benötigt in der `where`-Klausel anzuwenden.

FilterExample

```
1 using System;
2 using System.Threading;
3 using System.Collections.Generic;
4 using System.Linq;
5
6 class Program {
7     public static bool even(int value)
8     {
9         return value % 2 == 0;
10    }
11
12    public static void Main(string[] args){
13        var numbers = new List<int>() {-1, 7,11,21,32,42};
14        var query = from i in numbers
15                    where i < 40 && i > 0
16                    select i;
17        foreach (var x in query)
18            Console.WriteLine(x);
19    }
20 }
```

Cannot open assembly 'main.cs': File does not contain a valid CIL image.

Die entsprechenden Operatoren können aber auch um eigenständige Methoden ergänzt werden. Versuchen Sie zum Beispiel die Bereichsabfrage um eine Prüfung zu erweitern, ob der Zahlenwert gerade ist.

Gruppieren

Die group-Klausel ermöglicht es, die Ergebnisse auf der Basis eines Merkmals zusammenzufassen. Die group-Klausel gibt entsprechend eine Sequenz von `IGrouping<TKey,TElement>`-Objekten zurück, die null oder mehr Elemente enthalten, die mit dem Schlüsselwert `TKey` für die Gruppe übereinstimmen. Der Compiler leitet den Typ des Schlüssels anhand der Parameter von `group` her. `IGrouping` selbst implementiert das Interface `IEnumerable` und kann damit iteriert werden.

```
var queryCustomersByCity =
    from customer in customers
    group customer by customer.City;

// customerGroup is an IGrouping<string, Customer> now!
foreach (var customerGroup in queryCustomersByCity) // Iteration 1
{
    Console.WriteLine(customerGroup.Key);
    foreach (Customer customer in customerGroup) // Iteration 2
    {
        Console.WriteLine("    {0}", customer.Name);
    }
}
```

Dabei können die Ergebnisse einer Gruppierung wiederum Ausgangsbasis für eine weitere Abfrage sein, wenn das Resultat mit `into` in einem Zwischenergebnis gespeichert wird.

GroupByExample

```
1 using System;
2 using System.Threading;
3 using System.Collections.Generic;
4 using System.Linq;
5
6
7 class Student{
8     public string Name;
9     public int id;
10    public string Subject{get; set;}
11    public Student(){ }
12    public Student(string name){
13        this.Name = name;
14    }
15 }
16
17 class Program {
18     public static void Main(string[] args){
19         List<Student> students = new List<Student>{
20             new Student("Max Müller"){Subject = "Technische Informatik", id =
21                 1},
22             new Student("Maria Maier"){Subject = "Softwareentwicklung", id = 2},
23             new Student("Martin Morawschek"){Subject = "Höhere Mathematik I",
24                 id = 3},
25             new Student("Katja Schulz"){Subject = "Technische Informatik", id =
26                 4},
27             new Student("Karl Tischer"){Subject = "Softwareentwicklung", id =
28                 5},
29         };
30         var query = from s in students
31                     group s by s.Subject;
32         foreach (var studentGroup in query)
33         {
34             Console.WriteLine(studentGroup.Key);
35             foreach (Student student in studentGroup)
36             {
37                 Console.WriteLine("    {0}", student.Name);
38             }
39         }
40         var query2 = from s in students
41                     group s by s.Subject into sg
42                     select new {Subject = sg.Key, Count = sg.Count()};
43         Console.WriteLine();
44         foreach (var group in query2){
45             Console.WriteLine(group.Count + " students attend in " + group
46                             .Subject);
47         }
48     }
49 }
```

Cannot open assembly 'main.cs': File does not contain a valid CIL image.

Sortieren

Bei einem Sortiervorgang werden die Elemente einer Sequenz auf Grundlage eines oder mehrerer Attribute sortiert. Mit dem ersten Sortierkriterium wird eine primäre Sortierung der Elemente ausgeführt. Sie können die Elemente innerhalb jeder primären Sortiergruppe sortieren, indem Sie ein zweites Sortierkriterium angeben.

```
var queryLondonCustomers = from customer in customers
                             orderby customer.City, customer.Street descending
                             select customer;
```

SortExample

```
1 using System;
2 using System.Threading;
3 using System.Collections.Generic;
4 using System.Linq;
5
6
7 class Student{
8     public string Name;
9     public int id;
10    public string Subject{get; set;}
11    public Student(){ }
12    public Student(string name){
13        this.Name = name;
14    }
15 }
16
17 class Program {
18     public static void Main(string[] args){
19         List<Student> students = new List<Student>{
20             new Student("Max Müller"){Subject = "Technische Informatik", id =
                1},
21             new Student("Maria Maier"){Subject = "Softwareentwicklung", id = 2}
                ,
22             new Student("Martin Morawschek"){Subject = "Höhere Mathematik I",
                id = 3},
23             new Student("Katja Schulz"){Subject = "Technische Informatik", id =
                4},
24             new Student("Karl Tischer"){Subject = "Softwareentwicklung", id =
                5},
25         };
26         var query = from s in students
27                     orderby s.Subject descending
28                     select s;
29         foreach (var student in query){
30             Console.WriteLine("{0,-22} - {1}", student.Subject, student.Name);
31         }
32     }
33 }
```

Cannot open assembly 'main.cs': File does not contain a valid CIL image.

Ausgaben

Die select-Klausel generiert aus den Ergebnissen der Abfrage das Resultat und definiert damit das Format jedes zurückgegebenen Elements. Dies kann

- den vollständigen Datensatz umfassen,
- lediglich eine Teilmenge der Member oder
- einen völlig neuen Datentypen.

Wenn die select-Klausel etwas anderes als eine Kopie des Quellelements erzeugt, wird dieser Vorgang als Projektion bezeichnet.

SelectExample

```
1 using System;
2 using System.Threading;
3 using System.Collections.Generic;
4 using System.Linq;
5
6
7 class Student{
8     public string Name;
9     public int id;
10    public string Subject{get; set;}
11    public Student(){ }
12    public Student(string name){
13        this.Name = name;
14    }
15 }
16
17 class Program {
18     public static void Main(string[] args){
19         List<Student> students = new List<Student>{
20             new Student("Max Müller"){Subject = "Technische Informatik", id =
                1},
21             new Student("Maria Maier"){Subject = "Softwareentwicklung", id = 2},
22             new Student("Martin Morawschek"){Subject = "Höhere Mathematik I",
                id = 3},
23             new Student("Katja Schulz"){Subject = "Technische Informatik", id =
                4},
24             new Student("Karl Tischer"){Subject = "Softwareentwicklung", id =
                5},
25         };
26         var query = from s in students
27                     select new {Surname = s.Name.Split(' ')[0]};
28         Console.WriteLine(query.GetType());
29         foreach (var student in query){
30             Console.WriteLine(student.Surname);
31         }
32     }
33 }
```

Cannot open assembly 'main.cs': File does not contain a valid CIL image.

Einen guten Überblick zu den Konsequenzen einer Projektion gibt die Webseite <https://docs.microsoft.com/de-de/dotnet/csharp/programming-guide/concepts/linq/type-relationships-in-linq-query-operations>

Aufgabe der Woche

Für die Vereinigten Staaten liegen umfangreiche Datensätze zur Namensgebung von Neugeborenen seit 1880 vor. Eine entsprechende csv-Datei (comma separated file) findet sich im Projektordner und /data, sie umfasst 258.000 Einträge. Diese sind wie folgt gegliedert

1880,"John",0.081541,"boy"	
1880,"William",0.080511,"boy"	
1880,"James",0.050057,"boy"	

Die erste Spalte gibt das Geburtsjahr, die zweite den Vornamen, die Dritte den Anteil der mit diesem Vornamen benannten Kinder und die vierte das Geschlecht an.

Der Datensatz steht zum Download unter <https://osf.io/d2vyg/> bereit.

Lesen Sie aus den Daten die jeweils am häufigsten vergebenen Vornamen aus und bestimmen Sie deren Anteil innerhalb des Jahrganges.