# Debugger Visualizers for SharpDevelop

* Quick introduction to the project (to get the reader really interested)
  + Object graph visualizer, debugger tooltips, collection visualizer
  + Screenshots
  + There is no other IDE that does this
  + Started as GSOC 2009 project
  + maybe a group photo of us, the developers, from the meeting in Austria
* Introduction to SharpDevelop source
  + Overview of SharpDevelop architecture, link to Daniel’s CodeProject article
  + How to build an addin (since debugger is an addin)
* Debugger.Core
  + What are Expressions, Values and how they work, how Evaluate() works
  + what are Permanent references and what they are useful for
  + how expression caching works in Debugger.Core
* Object graph visualizer
  + Motivation
  + The failed attempt to this in VS and why it failed
  + Describe the graph building algorithm, why it is possible in O(n)
  + design – how is model reused when eg. just one node is expanded, what happens if a node is collapsed (just the edge is removed), lazy evaluation on scrolling, example object diagram of node content-tree (ContentProperty nodes etc.). Example object diagram of whole graph (shown also using Object graph visualizer).
  + how tree layout works (classical two-pass, show the “too much wasted space” case, how could it be improved?) – how horizontal / vertical is done.
  + Discussion about layouting software, why none of them is good for incremental layout (dynagraph seems to be dead, link to list of layouting software). Why we picked Graphviz at least for spline routing (and why we abandoned it in the end.)
  + Spline routing
  + graph matching and animation
  + (spline morphing – if we do it, we could try, might be hard / slow)
  + memory leaks and weak events
  + (node control caching- if I get it working) – Creating User controls is *really* slow (baml parsing), I’d just want to reuse the ones from the previous graph
  + How to interrupt graph building if the user continues the debugger (eg. F10 for step)
  + Visualizing a Permanent reference instead of Expression (eg. for recursive methods)
* More interesting details of Debugger.Core
  + interesting parts of Debugger.Core, why we decided to derive DebugType from System.Type (end of 2009).
  + ICorDebug + multithreading – unfortunately impossible, it’s STA (also some general talk about COM)
  + Slow evals, why they are slow and what could be done to do batch evals in the debuggee – inject batch-evaluator in the debuggee, then send a command to evaluate set of expressions.
* How profiling helped
* Collection visualizer
  + Motivation
  + Data virtualization for scrolling in collection visualizer (per column also)
  + How data virtualization was implemented generally and is actually used in both visualizers and in the tooltips as well
  + WPF UI virtualization
* Debugger tooltips
  + Motivation, (VS already has them)
  + Design – how the editor knows if to open a text tooltip (eg. over a method name) or a debugger tooltip. Iterating all addins and accepting the first one – common practice.
  + Debugger tooltips for IEnumerable – using “new List<object>(IEnumerable)” in the debuggee. How to do it for non-generic IEnumerable? Pros: no PermanentReferences needed. Cons: cannot display infinite collections – VS does it the same way.
  + How debugger tooltips are linked to the visualizers (opening visualizers from tooltips and how it is done)
  + How they are done using WPF Popup, (“+”) is a styled ToggleButton, what other options we had, how they were done in SD3 using a lot of code which we now don’t have to write thanks to WPF Popups
  + Closing the Popups (by clicking anywhere outside, F10 for debugger step etc.)
  + (Popups and keyboard input – if I figure this out, now they are not focusable)
* SharpDevelop console (ie. VS’s Immediate window)
  + Opening debugger tooltip instead of just text-dumping the evaluation result to the console
  + Overload resolution (if user types Foo(123, Bar().Len), how to tell which overload of Foo to call)
* (WPF – if the thesis is too short, otherwise see WPF Unleashed)
  + General concepts – markup, styles, templates

# Thanks

I would like to thank MSc. Pavel Ježek for his helpful, constructive advice on this thesis.

I would like say a very honest “thank you” to David Srbecký, the author of the SharpDevelop Debugger, and Daniel Grunwald, the lead developer of SharpDevelop IDE. They both deserve a lot of respect for their work.

Last but not least, thanks to Google for supporting students to work on interesting open source projects.

# Introduction

The idea for this thesis comes from a desire for better debugging features in current IDEs. We have had a vision about innovative ways of debugging – namely exploring object graphs and collections, and we implemented this vision as new functionality for SharpDevelop – the open source IDE for .NET. First, we describe motivation behind these features – why we built them even though developers could live without them so far. Then we describe the problems we had to solve when figuring out whether or not it is possible to implement these features, and how to implement them. We also describe the architecture of SharpDevelop – focusing on the part important to us, which is the Debugger. Finally, an important part of this thesis focuses on detailed design of these features and their integration into SharpDevelop .

To give the reader a picture of our work right in the beginning, Let’s start with a preview of the features we implemented:

## Object graph visualizer

Screenshot

While debugging, the Object graph visualizer lets the user explore object graphs referenced by the variables in user’s code. For example, if the user has an instance of a linked list in the program, the visualizer displays the linked list in a similar fashion people would draw it on a whiteboard. The graph is updated live as the user steps in the debugger.

## Collection visualizer

Screenshot

While debugging, the Collection visualizer provides insight into the contents of collections of objects. The main point is that it displays properties of the individual collection items so that the user doesn’t have to dig into them manually.

## Debugger tooltips

Screenshot

Debugger tooltips are a popular feature of Visual Studio. SharpDevelop 3 also has debugger tooltips but they do not support IEnumerable collections, which is quite an important feature. As a part of this thesis we reimplemented the tooltips in WPF for SharpDevelop 4.0 and we added support for IEnumerable collections.

## Contribution of this thesis

The visualizers are not copied from any existing IDE or debugger and come completely from the ideas of the author of this thesis. As a result of this thesis SharpDevelop is currently the only IDE to have such features and we are expecting user feedback to see if these features will bring value to users in every-day development.

## Background

The first ideas for this thesis come from the beginning of 2009 when I was experimenting with visualizing object graphs using the Visual Studio debugger (<http://coding-time.blogspot.com/2009/03/debugger-visualizer-for-visual-studio.html>, thinking about improving the way people debug data structures. Then I stumbled upon a poster for Google summer of code 2009 at the Faculty of Mathematics and Physics in Prague. Seeing that SharpDevelop (IDE written in C#) was among the mentoring organizations and they even wanted to work on their debugger, applying was a no-brainer.

Google summer of code (http://code.google.com/soc/) is a program run by Google. Open source projects can apply and Google selects the most interesting projects (participating organizations include Firefox, OpenOffice, Mono, gcc, Haskell and many other organizations). Then students apply with detailed proposals of their ideas. Google distributes approximately 1000 slots among the organizations, based on how popular the organization is (that is how many applications the organization received). For example, SharpDevelop had 5 slots in 2009. Then it is up to the people from the mentoring organization to select the students they like the most. These students then work for 3 months fulltime on their projects and receive 4500$ from Google, provided the mentoring organization confirms that the student did a good job. The unofficial goal of Summer of code is that students stay with the project after the summer and become contributors.

Thanks to Google summer of code I had an excellent opportunity to work with brilliant people that SharpDevelop team members are. They are definitely the best people to work with I have met so far. The reason is their passion and skill. They are not only writing code in their free time for free, but also they are excellent developers writing very high quality code from which one can learn a lot. They are very willing to help with any questions and discussions with them are just effective.

Photo Ischl?

SharpDevelop meeting in August 2009 in Austria. Left to right: Tomasz Tretkowski (Gsoc: C++ Backend Binding), Daniel Grunwald (Senior Developer, Architect), Martin Koníček (Gsoc: Debugger visualizers), Siegfried Pammer (Gsoc: Xaml Binding), David Srbecký (Debugger), Peter Forstmeier (SharpDevelop Reports), Christoph Wille (Project Management).

# Motivation and goals

## Object graph visualizer

### Motivation

todo



### Goals

todo

## Collection Visualizer

### Motivation

The second addition to SharpDevelop debugger is Collection visualizer – a hopefully better way to view collections of objects.

Say we are debugging a program which works with a collection of customers. By collection we mean List, ObservableCollection, array, IQueryable, or any other IEnumerable. This is how collections of objects are currently (as of 2010) displayed in most IDEs:

Screenshot

The obvious problem is that we cannot view all the customers at the same time but we have to drill down one-by-one. How do we quickly get an overview of what is in the collection or see properties of the customer we are interested about?

This is how collection visualizer solves this problem:

Screenshot of collection visualizer

As can be seen from the screenshot each row of the table represents one item of the collection. Each column then represents one property. This is very similar to how relational data are typically being displayed. Indeed, collection visualizer can also be used to display collection of objects loaded from a database. In such case the user sees the data in a very natural way – the same way a database management tool would display them.

### Goals

todo

The Collection visualizer, apart from different view on the data, brings two new features to SharpDevelop:

* The debugger can display any IEnumerable collections.
* The debugger can display collections of reasonable sizes.

In SharpDevelop 3 it was possible to debug ILists and arrays using the debugger tooltips. However when displaying a List the implementation of tooltips in Debugger.AddIn would eagerly get all the List items from the debuggee one-by-one through the debugger API. For a List with just thousand items this would hang whole SharpDevelop for several seconds (or minutes) since the expensive calls to get individual items have to be executed on the main thread.

The Collection visualizer solves this large collection problem by lazily getting next items only as the user scrolls down. This way collections of almost any size can be explored (more on this later).

## Debugger tooltips

### Motivation

### Goals

# Analysis

## Introduction to SharpDevelop IDE

SharpDevelop (<http://www.icsharpcode.net/opensource/sd/>) is a free open source IDE for .NET written almost entirely in C#. Its development started from scratch in 2000. Version 4.0 (as of November 2010) supports C# 4, Visual Basic 10, F#, IronPython, Boo and C++.

screenshot

The level of support for the individual languages varies. For example, F# code-completion is currently in development. For C++ Visual Studio is better, but C++ is not the main focus of SharpDevelop. Support for development in C# is comparable to Visual Studio and in some areas it supersedes Visual Studio. All this information is for SharpDevelop 4.0 as of November 2010.

SharpDevelop runs on Windows (for Linux there is MonoDevelop which was forked from SharpDevelop). SharpDevelop uses .NET SDK for the build process (that is MSBuild and the compilers). Version 4.0 support targetting .NET versions 2.0, 3.0, 3.5 and 4.0. Older projects can be automatically updated to target .NET 4.0. SharpDevelop uses the same project and solution file format as Visual Studio – therefore, it can be used side-by-side with Visual Studio for working on the same projects.

SharpDevelop currently does not support:

* Web application development. However, it can open and debug web application created by Visual Studio. ASP.NET MVC support is planned for the next version (5.0).

As SharpDevelop is free, it makes sense to compare it the free Express version of Visual Studio. SharpDevelop offers these features which are not present in Visual Studio Express:

* Integrated profiler
* NUnit integration (with test runner in SharpDevelop)
* Subversion and Git integration
* Code coverage (PartCover integration?)
* .NET Reflector integration, PInvoke integration
* Reports (think CrystalReports)
* Debugger visualizers
* Productivity features (ReSharper-like)
* Extensibility

Extensibility is an important feature: SharpDevelop can be extended or modified in almost any way. Its API is well designed and in case there is an extension point missing, the SharpDevelop team is open to good contributions from anyone and will definitely include them in official releases.

Being an open source IDE written entirely in C# makes SharpDevelop a very interesting project for programmers who would like to learn advanced topics and have their work used by many people.

## SharpDevelop architecture

Let’s have a high-level look at the architecture of SharpDevelop, its system of AddIns and how our Debugger visualizers fit into the picture.

### ICSharpCode.Core

Let’s first look at ICSharpCode.Core (further referred to as the Core) which is the very basis on which SharpDevelop is built.

The Core is a framework for building extensible applications. It provides an AddIns architecture, where AddIns can extend other AddIns. As in many frameworks, the developer of the actual application has to use the framework explicitly to provide extension points in his application. The goal of the Core is to be easy to use so that users can use it naturally to provide extension points in their applications easily.

The Core has 2 versions- ICSharpCode.Core.WinForms and ICSharpCode.Core.Presentation, designed to be used in Windows Forms and WPF applications respectively. They both reference ICSharpCode.Core which contains all the common functionality which is not UI-specific.

Let’s now look at the AddIn infrastructure – the AddIn tree.

#### The AddIn tree

Composed out of .addin xml files to a single tree.

One AddIn does this:

<Path name = "/SharpDevelop/Browser/Toolbar">

<ToolbarItem id = "Back"

icon = "Icons.16x16.BrowserBefore"

tooltip = "${res:AddIns.HtmlHelp2.Back}"

class = " SharpDevelop.BrowserDisplayBinding.GoBack"/>

<ToolbarItem id = "Forward"

icon = "Icons.16x16.BrowserAfter"

tooltip = "${res:AddIns.HtmlHelp2.Forward}"

class = " SharpDevelop.BrowserDisplayBinding.GoForward"/>

<ToolbarItem id = "Separator1" type = "Separator"/>

<ToolbarItem id = "GoHome"

icon = "Icons.16x16.BrowserHome"

tooltip = "${res:AddIns.HtmlHelp2.Homepage}"

class = "SharpDevelop.BrowserDisplayBinding.GoHome"/>

[...]

Then in our application we can call:

toolStrip = ToolbarService.CreateToolStrip(this,

"/SharpDevelop/Browser/Toolbar");

this.Controls.Add(toolStrip);

And we get a ToolBar with three buttons and a separator.

Another AddIn does this:

<Path name = "/SharpDevelop/Browser/Toolbar">

<Condition name = "SolutionOpen" action = "Disable">

<ToolbarItem id = "SyncHelpTopic"

icon = "Icons.16x16.ArrowLeftRight"

tooltip = "${res:AddIns.HtmlHelp2.SyncTOC}"

class = "HtmlHelp2.SyncTocCommand"

insertafter = "Separator1"/>

[...]

Which causes the Toolbar to have a new button. The call to ToolbarService.CreateToolStrip(this, "/SharpDevelop/Browser/Toolbar") combines everything on the path "/SharpDevelop/Browser/Toolbar" together and returns a toolbar with four buttons.

In the second example we can also see a Condition. The Condition tag causes the ToolbarItem to be enabled only if the SolutionOpen condition evaluates to true.

Interesting attributes of ToolbarItem are **class**, which is the name of the class which handles click on the toolbar button and **insertafter** which specifies at which position the item should be inserted. Insertafter refers to an **id** of another item.

So far we have seen two types of tags which can be inserted inside some Path in the AddInTree: Condition and ToolbarItem. These XML tags are called **codons** in SharpDevelop terminology. There are six types of codons supported by default (make it a table):

* Class - Creates object instances by invocating a type's parameterless constructor.
* FileFilter - Creates file filter entries for the OpenFileDialog or SaveFileDialog.
* Include - Includes one or multiple items from another location in the addin tree. You can use the attribute "item" (to include a single item) **or** the attribute "path" (to include all items from the target path).
* Icon - Used to create associations between file types and icons.
* MenuItem - Creates a System.Windows.Forms.ToolStrip\* item for use in a menu.
* ToolbarItem - Creates a System.Windows.Forms.ToolStrip\* item for use in a toolbar.

Further in Implementation mention how we are using Class to register debugger visualizer.

##### Doozers

Say that these codons are actually not hardwired, but they are converted from codons to objels by their corresponding **doozers** – e.g. ClassDoozer, MenuItemDoozer etc. Users can implement their own codons, e.g.:

<Path name="/SharpDevelop/Services/DebuggerService/Debugger">  
        <Debugger id="DefaultDebugger"  
                  supportsStepping = "true"  
                  supportsExecutionControl = "true"  
                  supportsAttaching = "true"  
                  supportsDetaching = "true"  
                  class="ICSharpCode.SharpDevelop.Services.WindowsDebugger"/>

</Path>

This is a good way to provide parameters to the instance from the AddIn definition file. This is possible because the DebuggerDoozer exists and has been registered:

<Runtime>  
        <Import assembly=":ICSharpCode.SharpDevelop">  
 <Doozer name="Debugger" class="ICSharpCode.SharpDevelop.Debugging.DebuggerDoozer"/>  
 [...]  
 </Import>  
</Runtime>

The doozer converts the codon to an instance. Maybe show IDoozer definition.

Also the SolutionOpen condition we have seen is not hard-wired but it is provided by a class called SolutionOpenConditionEvaluator:

<Runtime>  
        <Import assembly=":ICSharpCode.SharpDevelop">  
<ConditionEvaluator name="SolutionOpen" class="ICSharpCode.SharpDevelop.SolutionOpenConditionEvaluator"/>  
 [...]  
 </Import>  
</Runtime>

Maybe show IConditionEvaluator condition.

#### Localization

In XML: label = "${res: AddIns.Profiler.ProfilingView.CpuCyclesText }"

In code: StringParser.**Parse**("${res:AddIns.Profiler.ProfilingView.CpuCyclesText}")

The resource identifier can be any string but in SharpDevelop we are using namespace-like notation to prevent collisions.

#### PropertyService

maybe

#### Remarks

The Core is not dependent on SharpDevelop. Rather, SharpDevelop is an application built using the Core. We will get to this later in the Architecture overview. <-Forward reference – I think OK.

For more information about SharpDevelop.Core see <http://www.codeproject.com/KB/cs/ICSharpCodeCore.aspx> and http://www.codeproject.com/KB/cs/LineCounterSDAddIn.aspx

### Reusability

Several parts of SharpDevelop are written as completely standalone libraries. These include:

* NRefactory (C# and VB parser and AST)
* ICSharpCode.SharpDevelop.Dom (representation of type system)
* AvalonEdit (code editor with syntax highlighting)
* Debugger
* Profiler
* Usage data collection
* Reports

All of these libraries are completely reusable. They are integrated to SharpDevelop by an AddIns which act as “glue” to provide library’s functionality. E.g. Debugger.Core is a managed debugger library and Debugger.AddIn contains UI and SharpDevelop-specific logic. In the same fashion, AvalonEdit is a drop-in code editor with support for syntax highlighting and AvalonEdit.AddIn adds SharpDevelop-specific behavior, like split-view, context actions etc.

Picture of editor, document, parser: parse text to NRefactory AST, update DOM, discard AST

Screenshot of CC with picture – design: obtained from DOM, ILanguageBinding.HandleKeyPress, AvalonEdit.ShowCompletion.

### NRefactory

### ICSharpCode.DOM

#### ICSharpCode.SharpDevelop

As we said earlier, SharpDevelop is built using the ICSharpCode.Core. SharpDevelop is composed out of AddIns. The main AddIn is the SharpDevelop “itself” – contained in the project ICSharpCode.SharpDevelop. We can find the definition of this AddIn in the file ICSharpCode.SharpDevelop.addin.

The SharpDevelop AddIn contains base of SharpDevelop functionality, including the project system and main UI. The SharpDevelop AddIn is not a working IDE though – it rather provides interfaces to be implemented by AddIns. For example, code editing, code completion, and debugger are all implemented as AddIns extending ICSharpCode.SharpDevelop. That said, also all all the other AddIns extend ICSharpCode.SharpDevelop.

### The editor

### Language bindings

### Project system

IProject. MSBuildBasedProject is the base, does loading using Microsoft.Build.

### Debugger.Core

### Debugger.AddIn

## Object graph visualizer

### Existing work

There have been attempts to solve similar problem we are trying to solve with our Object graph visualizer. Probably the most significant one is Data Display Debugger (<http://www.gnu.org/software/ddd/>) which is a graphical frontent to the command line debuggers like GDB or Python debugger pydb. None of these debuggers supports debugging manager .NET programs though while managed .NET programs are our point of interest.

## Collection Visualizer

### Existing work

There are also some existing tools to visualize collections in the debuggee, for example <http://davidhayden.com/blog/dave/archive/2005/12/26/2645.aspx> which is a Visual Studio debugger visualizer. The problem with this visualizer is that it is designed to visualize only one specific type – ShoppingCart, and the debuggee and the visualizer most both share a reference to this type. We haven’t found any visualizer which works the same way our Collection visualizer works.

## Debugger tooltips

### Existing work

There is a standard feature of Visual Studio very similar to our debugger tooltips. SharpDevelop 3 has debugger tooltips as well. Since SharpDevelop 4 was completely rewritten to WPF it needed new debugger tooltips so we implemented them. Also, the tooltips for SharpDevelop 3 didn’t support exploring instances of IEnumerable collections and we successfully added this.

# Implementation

## Common base for visualizing collections

## Object graph visualizer

## Collection Visualizer

## Debugger tooltips

# Conclusion and future work

# User documentation

# Introduction to SharpDevelop

# Object graph visualizer

## First attempt to implement an object graph visualizer

The implementation described in this thesis is actually a second version of object graph visualizer. The first version was implemented for Visual Studio. The attempt was a failure (details described further) but provided important lessons on how some problems should (not) be approached.

## Graph building

Definition: Object graph is an oriented graph. Its vertices are in-memory instances. There is an oriented edge from vA to vB if and only if instance A has a direct reference to B.

The problem: Given a reference to an instance in the debuggee, build an object graph representing all instances reachable from this instance (up to maximum depth in case the graph is too large).

Our solution: generic with GetNode left unexplained

In the first implementation, the GetNode method was implemented like this: code from Fabric.

The important lesson learned from the first implementation was that the calls to the .NET debugging interface are quite expensive. In particular because the graph building needed O(n2) debugger calls (where n is the size of the graph) it was practically unusable for graphs larger than 10 objects, which is still in a very reasonable range of what users want to visualize.

The current implementation looks like this:

build a hashtable: 'hashCode' -> (list of objects with hash code == 'hashCode')  
find if object seen(o) {  
    candidates = hashtable[o.GetHashCode()] // objects with same hashCode  
    if no candidates, the object is new  
    if some candidates, compare their addresses to o.Address  
      if no address equal (the hash code was just a coincidence) -> o is new  
      if some address equal, o already seen  
}

**ObjectGraphBuilder.GetExistingNodeForValue – uses addresses – might result in incorrect graph!**

#### .NET hashcodes

<http://stackoverflow.com/questions/750947/-net-unique-object-identifier>

Our object graph building algorithm accounts for different instances having same hash codes. The reason is that the hash code is a 32-bit integer, therefore the space of all hash codes is limited and it is not possible to guarantee that two different objects will have always different hash codes.

One could think that 32-bit space is big enough and in practice we will almost never encounter different instances with same hash codes. However, tests on CLR (Microsoft implementation of CLI) are quite surprising. This code only needs to generate as low as 5000 objects to encounter two different objects with same hash codes:

Code from SO

On second thought this result is not that surprising at all. If the hash codes were generated completely at random there would be quite a high probability of hitting one value twice after 5000 attempts (this is the same like the birthday paradox). Then, generating hash codes for example sequentially is not a good idea either as we know – hash codes should be well distributed if we want them to serve their purpose – serve keys for hash tables.

Note that the test program always outputs the same number on the same system with given version of .NET. This shows that the hash code generation is deterministic.

#### Hash code implementation in the runtime

In .NET, the runtime representation of every instance has a data member which stores the hash code for the instance. This data member is assigned by the runtime and it is accessible from managed code through Object.GetHashCode() method. User can override this method but the original runtime-assigned hash code is still accessible through RuntimeHelpers.GetHashCode().

We are always working with the original hash codes so that user defined hash codes don’t interfere with the graph building algorithm. If we used GetHashCode() and the user overrode it e.g. to always return 0 the algorithm would still work but would run in O(n2) instead of O(n).

## Incremental stability of graph layout

One of the important lessons learned during the first implementation was that the graph update during stepping was very chaotic: when the object graph changed just a little (e.g. a new leaf node was added), the layout determined by the layout library could change radically. In other words, the graph layout was incrementally unstable.

To resolve this problem we started looking for ways to make the layout incrementally stable (small graph change results in small layout change) and we also implemented animation which moves existing nodes from old positions to new positions.

It is clear that to make the layout incrementally stable, the layout engine must either support incremental layout (almost no engine does) or there must be a possibility for the user to control the layout to some extent.

The layout engines considered were:

List of graph layout engines

Dynagraph – not maintained

Graphviz

GraphSharp – this library was not available at the time of implementing the object graph visualizer. It is open source and written in C#. With some modifications to it could be a good candidate.

The options were not perfect. As no engine was both good enough and supported incrementally stable layout, we decided to go in the direction of using an engine which has no notion of incremental stability and controlling the layout ourselves.

As for the graph engine, we decided for Graphviz.

As for the graph layout, we decided that we would actually calculate all the positions of the nodes ourselves and let Graphviz just route edges.

## Node layout

Any graph layout algorithm needs to solve two problems:

* Determine the positions of nodes (node layout)
* Determine the paths of edges (edge routing)

These do not necessarily have to be two separate steps. However, Graphviz first determines positions of all nodes (of course edge information is considered in this procedure) and after this step is finished, the positions of nodes are fixed and edges are being routed, one by one.

As we said, we implemented the first step (node layout) ourselves and used Graphviz for the second step (edge routing). The reason is that no existing layout engine was satisfying in terms of incremental stability (i.e. small graph change should result in small layout change).

Let’s see how we approach node layout.

The problem (node layout): Given an oriented graph, determine position of each node in the plane.

Such description of the problem is too vague. It should also state what property of the layout the solution should try to optimize. For example, Graphviz’s dot algorithm tries to achieve layout in which as many edges as possible point downwards. Other algorithm might try to achieve minimal edge crossing. Other algorithm (Graphviz neato) tries to achieve “natural” layout by doing a physical simulation of springs representing graph edges.

Our solution aims for determinism and incremental stability.

Realizing that the algorithm will be used to layout object graphs, we decided that the algorithm would be a tree layout algorithm with deterministic order of node’s children. The input graph is not necessarily a tree but we layout is as it were a tree (by keeping n-1 tree edges and forgetting the rest of the edges).

The question is, if the graph has m > n-1 edges (n is the number of nodes), how to determine which edges to keep and which to discard. This is done by depth-first traversal of the graph:

Traverse(node):

NodeAlreadySeen[node] = true

For each edge in node.OutgoingEdges

If NodeAlreadySeen[edge.target]

edge.Discard = true // not a tree edge -> discard it

Else

Traverse(edge.Target)

In the end, all the edges for which edge.Discard is true are removed from the graph and the rest of edges is kept. This way we obtain a tree so that we can layout it using a two-pass tree layout algorithm:

First pass: “measure”

Measure(node)

Foreach child in node.Children

Measure(child)

subtreeHeight = node.Children.Sum(n => n.DesiredHeight)

node.DesiredHeight = max(node.OwnHeight, subtreeHeight);

Second pass: “arrange”

Arrange(node, position) // place node at given position

// center node relative to its children

node.Pos = CenterVertically(position, node.DesiredHeight)

childPos = position

Foreach child in node.Children

Arrange(node, childPos)

childPos.Y += child.DesiredHeight // place next child below this child

The important part of this algorithm is its determinism: the first child on input will always be the topmost child and the last child will be the lowermost child. In object graph the order of children corresponds to the order of properties in the class so it makes a lot of sense to order children like this:

Screenshot of object graph node with outgoing three edges

Screenshot of object graph collection node with outgoing three edges

Note that many layout engines do not order children deterministically. In Graphviz it is possible to force deterministic ordering of children by adding artificial edges between the children but this adds computational complexity and it is not completely reliable.

#### Similarity to WPF’s “Measure-Arrange” layout algorithm

It is not a coincidence that the two phases of the algorithm above are named “measure” and “arrange”. This algorithm works on exactly the same principle as the WPF layout algorithm. The same algorithm is probably used in most UI models (XUL, html?).

In WPF, the user interface elements form so called visual tree.

When an UIElement is asked for its size (Measure() method) it asks its children for their size and then determines its own size.

When an UIElement is told its new position (Arrange() method) it also repositions its children, knowing their sizes because Measure() has already been called before Arrange().

In WPF the user can define his own user interface elements and completely control layout of their children by subclassing FrameworkElement and overriding MeasureOverride() and ArrangeOverride() methods.

This is also how existing UIElements are implemented. For example, StackPanel is a panel which stacks its children one next to each other. This logic is implemented in the ArrangeOverride method. The Measure method tells that the StackPanel needs as much space as all its children together.

Measure and Arrange actually each have a parameter. Measure(Size maxSize) specifies the maximum available size which the element can use. Arrange(Size arrangeSize) means what? Read WPF Unleashed

## Graphviz - graph layout and drawing toolkit

## Graph matching and animation

## Spline routing

The problem:

Example of a solution (Graphviz): screenshot from Graphviz

One approach to routing splines is routing the edges one-by-one, that is treating every edge as completely separate input. This is how Graphviz does it. This approach can lead to spline overlaps.

Another approach could be routing the splines globally, i.e. every edge path can affect paths of other edges. Such approach could try to reduce edge overlaps while still maintaining reasonable edge paths. The Graphviz paper mentions this, but they do not provide any ideas for such algorithm.

As said, Graphviz was used for routing the edges and the appearance of resulting graphs was good: screenshot

However, the members of SharpDevelop team expressed a concern about the impact of the size of Graphviz binaries on the size of SharpDevelop installer: Graphviz needs about 10Mb of binaries and won’t run without them even if they are not needed (for example it crashes when a library for writing jpegs is missing even though the output type is set to *text*).

Therefore we decided to implement also the spline routing ourselves and get rid of Graphviz completely.

#### Our algoritm

We realized that when routing edges to avoid rectangles, the crucial points in the plane are the *corners* of the rectangles. Second, to make edges look natural to humans, some sort of shortest path routing would be reasonable. The Graphviz paper confirmed our ideas.

Like Graphviz, we route the edges one-by-one. This is the algorithm:

For each edge e in G:

Determine edge start and end point of edge e: draw a straight line from the center of edge’s source rectangle to the center of edge’s target rectangle. Where this line intersects the source rectangle is the start point es. Analogically end point et.

Build following graph Gv (visibility graph):

V = (every 4 corners every rectangle on input) + (for all e: es) + (for all e: et)

E = (pairs (u, v) from V where u is visible from v: straight line can be drawn from u to v without crossing body of any rectangle)

For each edge e in G:

In Gv find shortest path from es to et (using e.g. A\* or Dijkstra’s algorithm)

Smoothen the joins of obtained path by using Bezier curves

Note that the algorithm deals with the situation when boxes overlap.

Here is an example of a result:

screenshot

This algorithm is O(n3) where n is the number of boxes due to the construction of visibility graph: O(n2) vertex pairs are tested for visibility and each test needs O(n) line-rectangle intersections. It could be probably optimized but for our needs it is good enough.

Note that Graphviz also uses O(n3) algorithm for visibility graph construction.

This algoritm is implemented to be completely reusable. The implementation can be found in /SplineRouting folder of SharpDevelop repository (link). The reusability is achieved by writing the algorithm to work with interfaces. Programming against interface is a common design practice.



#### Edge overlaps

Like with Graphviz, the edges are routed one by one and sometimes can produce long overlaps. We solved this in the user interface by highlighting edge under cursor.

screenshot

#### Multiple edges, self edges

Our algorithm also deals with the situation when there are multiple edges between one pair of boxes. This is solved by distributing the edge starting and ending points along the border of the box. Self edges are solved in a similar way.

Screenshot multi edges

Screenshot self edges

#### Explicitly determined start and end point

Some layout engines (including Graphviz) let the user specify starting and ending point of the edge. This could be useful in object graph visualizer because the edge outgoing from an object property could start directly next to the title of the property. We considered this and concluded that the layout would get problematic when the contents of a node are scrolled – the edge point would have to move and the edge would have to be re-routed when scrolling. Therefore we decided not to implement this feature.

#### Join smoothing

The last step of the algorithm – join smoothing – makes the result visually much more appealing:

Screenshot with join smoothing

Screenshot without join smoothing

The principle is the following:

Given two consecutive line segments, replace the segments by a Bezier curve of order 3 extended by straight line segments on each end:

Screenshot

The control points of the Bezier curve lie on the original lines. The distance of the control points from the original points determines the “smoothness” of the curve:

(Screenshot)

# Collection visualizer

## Lazy loading items when scrolling

Having said that the Collection visualizer obtains values from the debuggee lazily as the user scrolls down the grid, let’s now see how we designed this.

There is a conceptual difference between two types of collections:

* IList collections, allowing random access using the indexer
* IEnumerable collections, allowing only sequential access

We approached these two cases differently. When the visualizer is asked to visualize an instance of a collection it first checks whether the instance implements IList<T>. If the instance does not implement IList<T> it further checks if the instance implements IEnumerable<T>. Based on whether the instance is an IList<T> or IEnumerable<T> we choose between two approaches to display its items:

### IList<T>

We use WPF ListView to display the contents of the collection. ListView (as any ItemsControl) has an ItemsSource property. ItemsSource is of type IEnumerable and the ItemsControl distinguishes whether the actual ItemsSource is an IList or just an IEnumerable: if it is an IList the ItemsControl uses its indexer to query individual items as it needs to render them when they come into view, mostly by scrolling. We take advantage of this.

What we do is that we supply a special collection as the ItemsSource. This collection is a wrapper around some data source (in our case the debuggee collection) and when asked for an item at index *i* it queries the underlying data source for item *i* and returns it (and caches it). We call this collection VirtualizingCollection and this principle is sometimes called data virtualization. We call the underlying data source IListValuesProvider.

The top part of the class diagram is the data virtualization implementation. The bottom part is specific to virtualizing collections of objects in the debuggee. ListValuesProvider is the implementation of IListValuesProvider which wraps the collection in the debuggee and makes calls to the debugger API. The items returned by the indexer of ListValuesProviders are instances of ObjectValue – our representation of a collection item in the debuggee. ObjectValue contains values of item’s properties converted to string by the debuggee-defined ToString() method. It also contains index of the item in the collection.

The last thing left unexplained is the Count property of the VirtualizingCollection. The value of this property is needed for the ItemsControl to display a scrollbar correctly. The VirtualizingCollection asks the data source for the count of its items and in our case the data source queries the debugger API for the value of Count property of the collection in the debuggee.

The good thing about this implementation is that when the user scrolls fast the ItemsControl only queries items which are needed to be rendered, skipping indices which were skipped by the fast scrolling (and saving unnecessary debugger API calls).

Note on data virtualization: As we said an implementation of IListValuesProvider wraps some underlying data source and queries items from it. In some cases getting a group of items from the data source one-by-one is more expensive than getting all the items in one call due to the overhead of each query. One such scenario are relational databases. In such case it is reasonable to implement some sort of paging in the value provider so that when it is asked for an item *i* it queries and caches a number of items around index *i* because it is very probable that these items will be needed right after.

### IEnumerable<T>

As in case of IList<T> we use WPF ListView to visualize contents of IEnumerable<T> collections in the debuggee. The difference lies in the data virtualization. We designed two solutions – the first later replaced by the other. Let’s describe and compare both of them.

##### The first solution

As said, the ItemsControl’s ItemSource property distinguishes between IList and IEnumerable: whereas in case of IList it uses its indexer to query individual items, in case of IEnumerable it always enumerates all the items. Wrapping the IEnumerable in our special IEnumerable and relying on the ItemsControl to query next items as needed is therefore not an option. We will have to do more work.

Our solution is to subclass ListView and implement the “lazy” functionality we need there. The subclass is called LazyListView. LazyListView watches when its scrollbar reaches the bottom margin and queries more subsequent items from the data source and adds them to its ItemsSource collection.



In the beginning only a number of items (given by LazyListView.InitialCount) is loaded from the data source. Then as the user scrolls to the end of the view more items are added to the end of the collection and the scrollbar gets shorter. User can continue pulling the scrollbar down. The overall user experience is quite natural.

##### Alternative to the first solution

This was actually an idea of David Srbecký. In the previous diagram the logic of adding additional items when needed is implemented in the LazyListView class. We could also implement this in the VirtualizingIEnumerable in the following way:

VirtualizingIEnumerable is an ObservableCollection. Its Count initially returns some fixed number (analogy to InitialCount of LazyListView). Its indexer ensures that at least *i* items are loaded from the data source and returns the *i*-th item. When item at index near the value of Count is requested (that means the ListView wants to render an item near end of the collection), Count is incremented and PropertyChanged(“Count”) event is raised. The ListView therefore updates its scrollbar.

This solution is a little more complicated than the first one.

##### The second solution

The second solution is quite different from the first one. Instead of keeping a reference to an IEnumerator in the debuggee and querying subsequent items from it as needed we enumerate all the collection at once in the debuggee:

Say we have a reference to IEnumerable<Foo> variable in the debugee, *v* be the name of the variable in the debuggee. We construct the following expression: “new List<Foo>(v)” and evaluate it. This allocates a List in the debuggee and runs its constructor. The constructor enumerates the IEnumerable and stores its items. The result of the evaluation is the reference to the new List. Finally, we visualize this reference exactly the same way as we visualize ILists (described in the previous chapter).

Note that in most cases enumerating whole collection in the debuggee is almost instant (or takes up to five seconds, then evaluation timeouts). What determines the speed of the visualizer almost completely is the expensiveness of the debugger API. Therefore it rarely makes a difference to enumerate whole collection in the debuggee and fetch first *n* items or fetch first *n* items from an IEnumerable in the debugee.

##### Comparison of the solutions

The first solution and its alternative are essentially the same and differ only in the way of determining when the user has scrolled to the end of the collection. We will therefore compare the first and the second solution.

As we said, the performance difference is in most cases negligible, unless each GetNext() call of the enumerator is quite expensive. In such cases the first solution wins.

The first solution can, unlike the second, visualize “infinite” collections. By infinite we mean IEnumerables such as:

IEnumerable MakeInfiniteCollection()

{

int i = 0;

while(true) {

yield return i++;

}

}

The first solution will allow the user to scroll “infinitely” (until memory runs out) and more importantly it will display the first few items of the infinite collection. The second solution will try to enumerate the whole collection in the debugee and therefore always timeout / run out of memory and display error message.

On the other hand the second solution lets the user immediately see how many items there are in the collection and scrolling is much smoother as many items can be skipped when scrolling fast.

The first solution has some quite significant disadvantage though: We wanted to implement a feature to be able expand each individual item of the collection as a debugger tooltip:

screenshot

This is very useful as the collection visualizer only shows stringified values of item properties and often the item is a composite object and we would like to explore it more deeply. Now, if we look at the first solution – it obtains items (Values) from an enumerator but these Values have no Expression. They are simply Values in the debuggee but there is not a way to describe a “path” (e.g. in C#) to obtain each Value. That means to be able to expand each Value we have to keep a PermanentReference to it. But this means we will be holding a lot of PermanentReferences (adding up as the scrolls through a long collection) and as we know this is discouraged by MS documentation of the ICorDebug API.

On the other hand if we are working with an IList each item has a clear expression – e.g. “list[5]”. Therefore we do not need to keep any PermanentReferences and when used expands item at index *i* we evaluate “list[i]” again and display a debugger tooltip for obtained value.

(Note that our list actually has no name such as “list” because we allocated it dynamically by evaluating expression “new List<Foo>(identifier)”. Indeed, the expression “new List<Foo>( identifier)[i]” is what we will be evaluating and it will work without reallocating the list each time because of caching in the ExpressionEvaluator: “new List<Foo>(identifier)” is cached as a PermanentReference and each evaluation of any expression that contains this expression will use the cached value.)

Expression caching is useful but we have to be careful with it: if we simply evaluated expressions like “new List<Foo>(identifier)[i].Property”, the “new List<Foo>(identifier)[i]” would be cached as a PermanentReference. We have to call the indexer and evaluate the property without using Expressions to avoid the Expression cache.

To sum up, what is important about the second solution is that we have a store for the items in the debuggee (the List) providing named access to any individual item of the collection. In the first solution we had to keep PermanentReference to each item to be able to access it again to expand it.

Visual Studio 2008 probably uses something similar to our second solution. This can be tested by trying to expand a debugger tooltip for an infinite IEnumerable. Instead of displaying the first few items a timeout error is displayed as the debuggee is given a few seconds to try to enumerate the collection.

##### Third solution

In theory it would be possible to combine the two solutions – have lazy access from the first solution and a store of items in the debuggee providing named access to each item. What we would have to do would be to evaluate “list.Add(value)” after obtaining each value from the enumerator. This way the list would grow as the user would scroll and any item could be reobtained and expanded by querying “list[i]”.

This solution would still suffer from the slow scrolling and unability to see the total count of items in the IEnumerable collection. We feel that not being able to visualize infinite collections is not such a big disadvantage of our second solution.

##### Note on garbage collection

Note that all the described difficulties are caused only by the fact that the Garbage collector can move instances in memory when compacting the heap. If the address of each instance stayed always fixed we would just enumerate the items, remember memory address for each of them and then accessed any item by its address. In garbage collected environment, PermanentReference brings us the same comfort with the exception that we should keep the number of PermanentReferences reasonably row (in order of hundreds).

## Expanding the items in the grid

## Generic vs. non-generic collections

## Lazy loading columns (object properties)

Show also IQueryable, ObservableCollection, IParallelEnumerable.

## Heterogeneous collections

# Debugger tooltips