# Fundamentale Organisations- & Architekturprobleme durch Cross-Project Clone Detection finden

Dr. Elmar Jürgens

CQSE
Continuous Quality in Software Engineering

# Über Mich

## Forschung

- Clone Detection, Architekturanalyse
- Effektivität und Effizient von QS-Maßnahmen

## Beratung

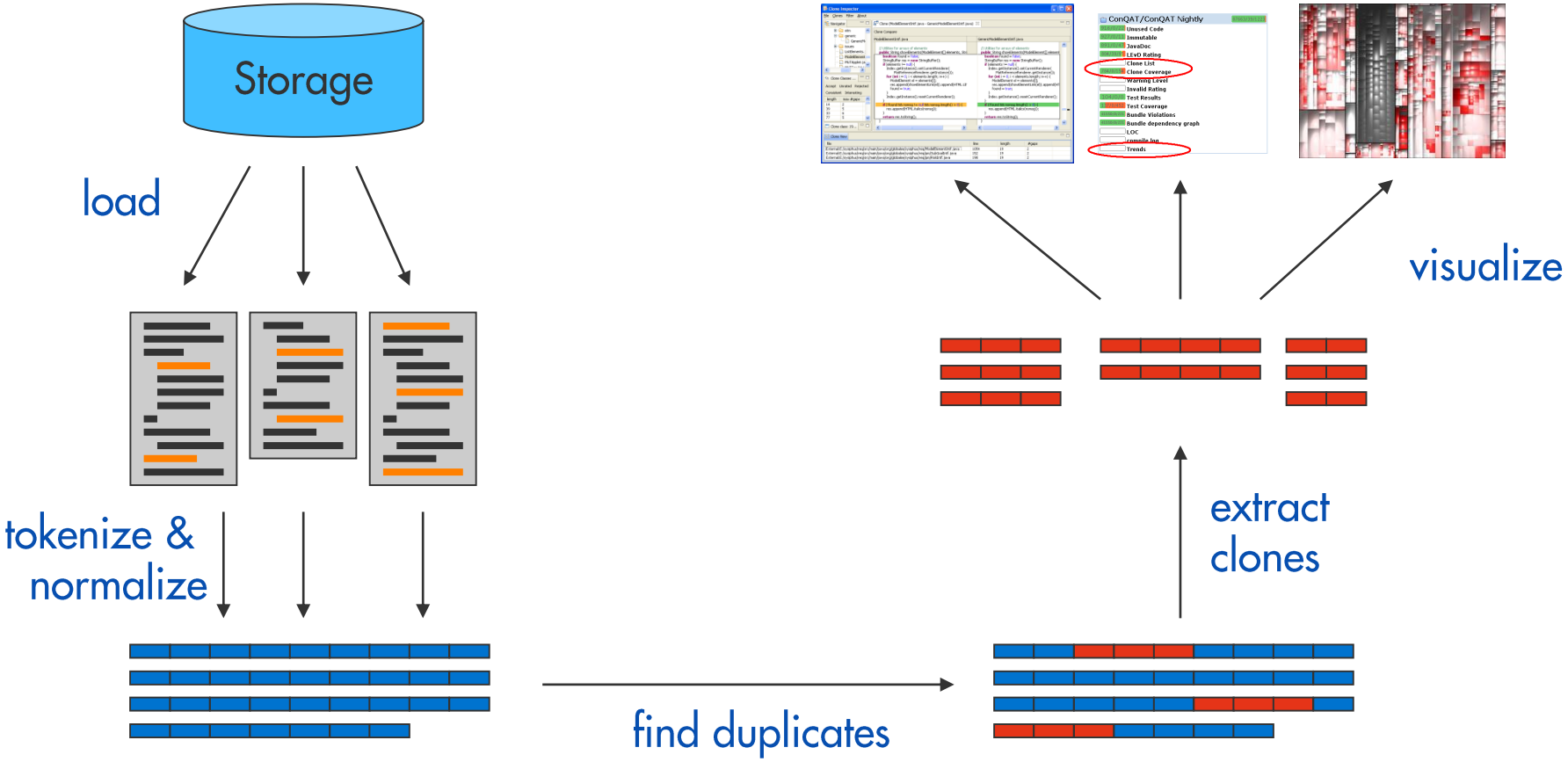
- Gründer
- Qualitäts-Bewertung & Qualitäts-Controlling

## Gesellschaft für Informatik

- Zum Junior-Fellow ernannt
- Erfahruzngsaustausch Forschung <-> Praxis

```java
// Utilities for arrays of elements
public String showElements(ModelElement[] elements, String nomsg) {
   boolean found = false;
   StringBuffer res = new StringBuffer();
   if (elements != null) {
      Index.getInstance().setCurrentRenderer(
            FlatReferenceRenderer.getInstance());
      for (int i = 0; i < elements.length; i++) {
         ModelElement el = elements[i];
         res.append(showElementLink(el)).append(HTML.LINE_BREAK);
         found = true;
      }
      Index.getInstance().resetCurrentRenderer();
   }
   if (!found && nomsg != null && nomsg.length() > 0) {
      res.append(HTML.italics(nomsg));
   }
   return res.toString();
}
```

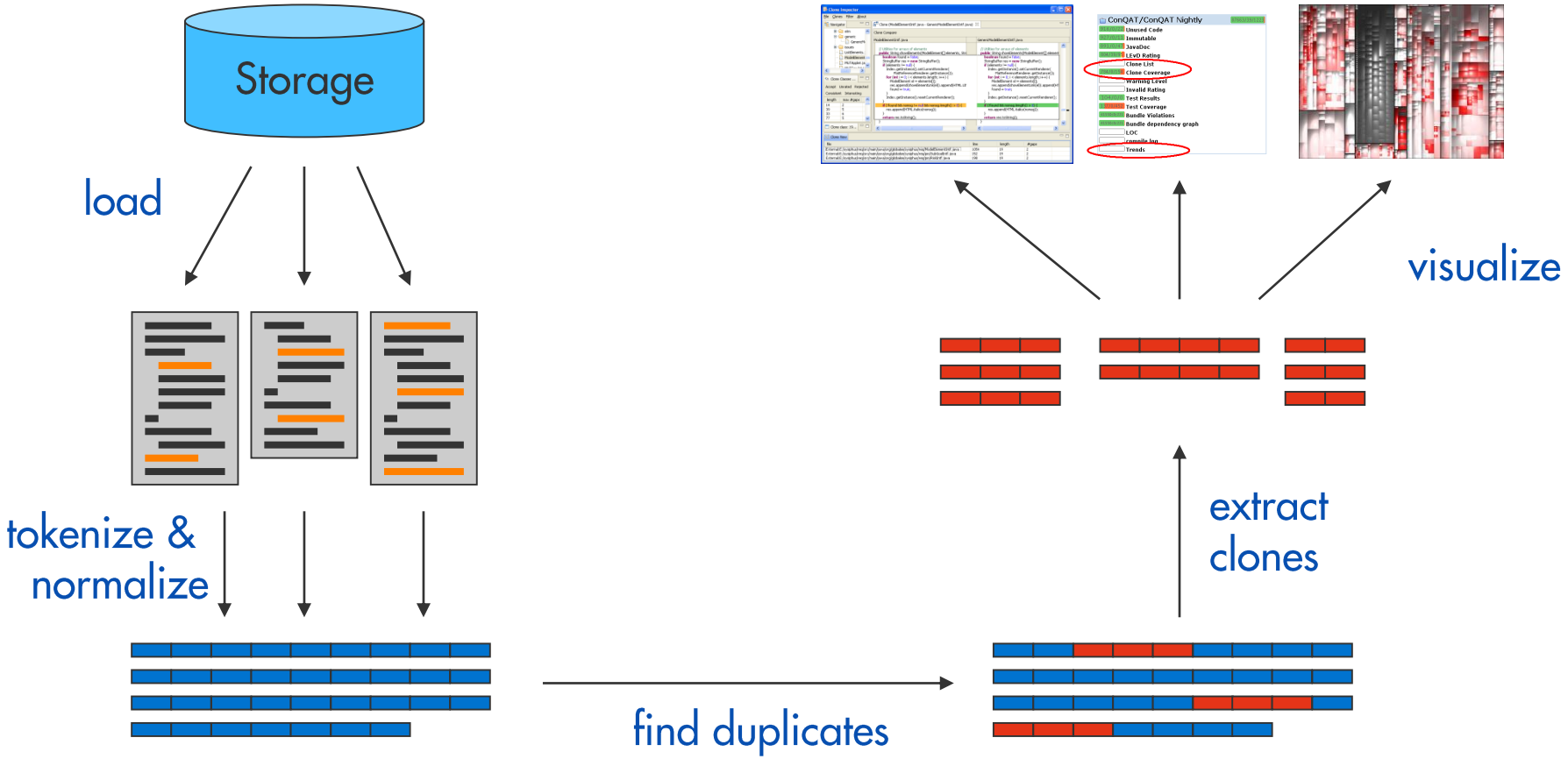
```java
// Utilities for arrays of elements
public String showElements(ModelElement[] elements, String nomsg) {
   boolean found = false;
   StringBuffer res = new StringBuffer();
   if (elements != null) {
      Index.getInstance().setCurrentRenderer(
            FlatReferenceRenderer.getInstance());
      for (int i = 0; i < elements.length; i++) {
         ModelElement el = elements[i];
         res.append(showElementLink(el)).append(HTML.LINE_BREAK);
         found = true;
      }
      Index.getInstance().resetCurrentRenderer();
   }
   if (!found && nomsg.length() > 0) {
      res.append(HTML.italics(nomsg));
   }
   return res.toString();
}
```


Storage
load
tokenize &
normalize
find duplicates
extract
clones
visualize
ConQAT/ConQAT Nightly
Unused Code
Immutable
JavaDoc
LEvD Rating
Clone List
Clone Coverage
Warning Level
Invalid Rating
Test Results
Test Coverage
Bundle Violations
Bundle dependency graph
LOC
Trends

# Normalisierung

```
String readFileUtf8(File file) {
    FileInputStream in = new FileInputStream(file);
    byte[] buffer = new byte[file.length()];
    in.read(buffer); in.close();
    return new String(buffer, „UTF-8“);
}

String readFileUtf16(File file) {
    FileInputStream in = new FileInputStream(file);
    byte[] buffer = new byte[file.length()];
    in.read(buffer); in.close();
    return new String(buffer, „UTF-16“);
}
```

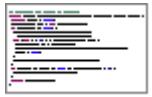
```
id0 id1(id2 id3) {
    id0 id2 = new id0(id4);
    id0[] id1 = new id0[id2.id3()];
    id0.id1(id2); id0.id3();
    return new id0(id1, lit0);
}

id0 id1(id2 id3) {
    id0 id2 = new id0(id4);
    id0[] id1 = new id0[id2.id3()];
    id0.id1(id2); id0.id3();
    return new id0(id1, lit0);
}
```

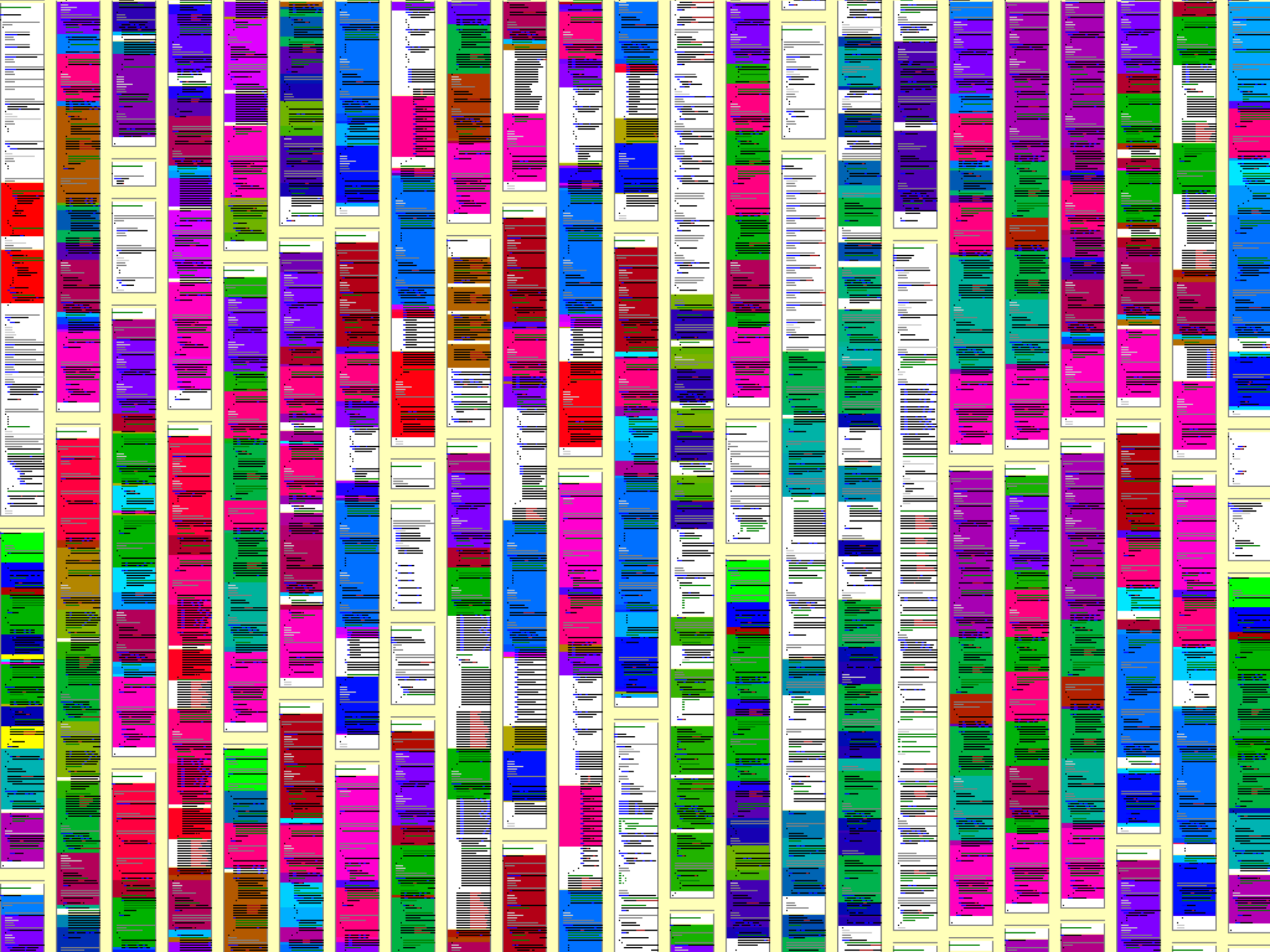


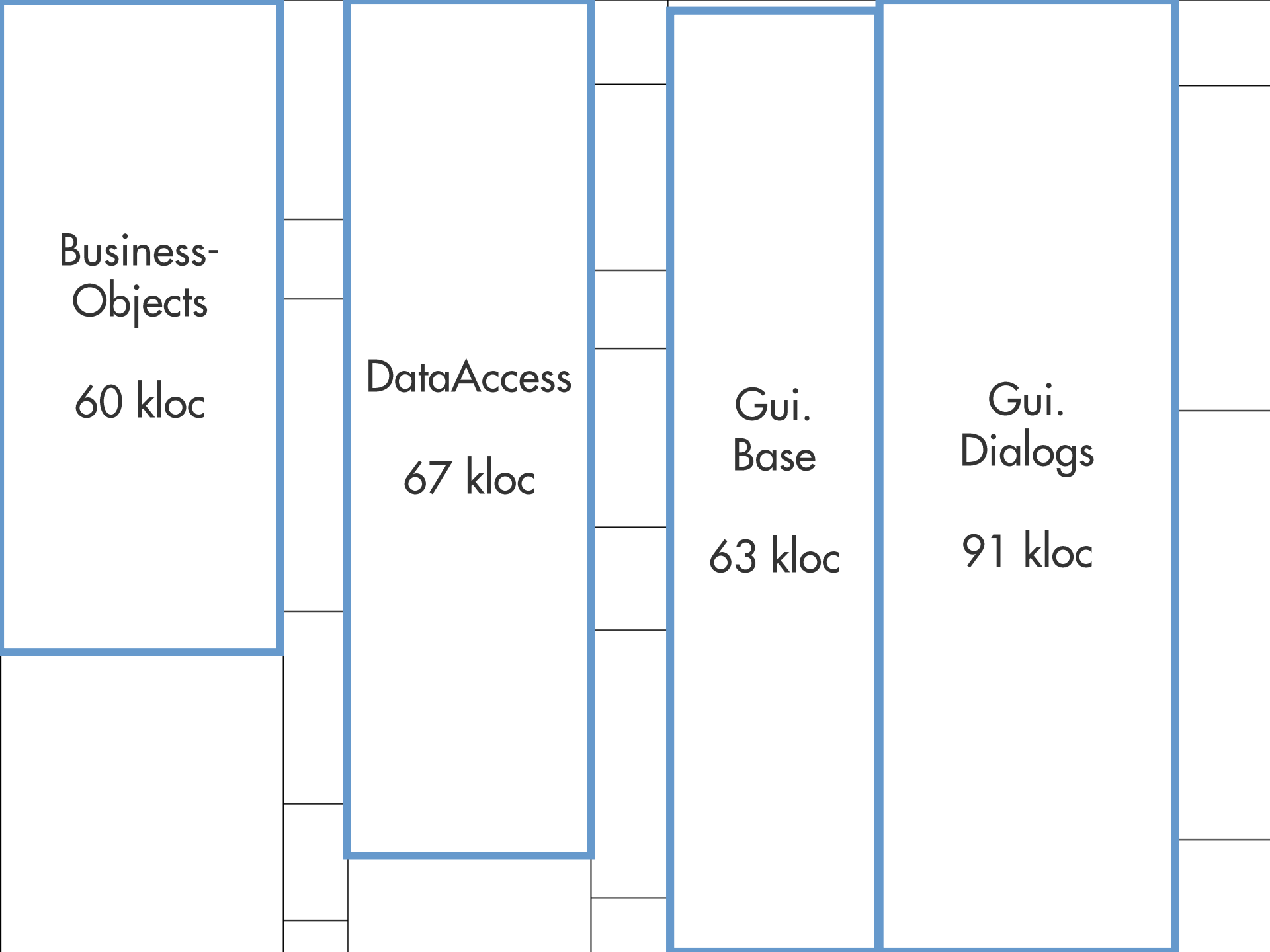

ABAP, ADA, C, C#, C++, Cobol, Java, JavaScript, Matlab, PHP, PL/I, PL/SQL, Python, T-SQL, VB, VB.NET, XML, …

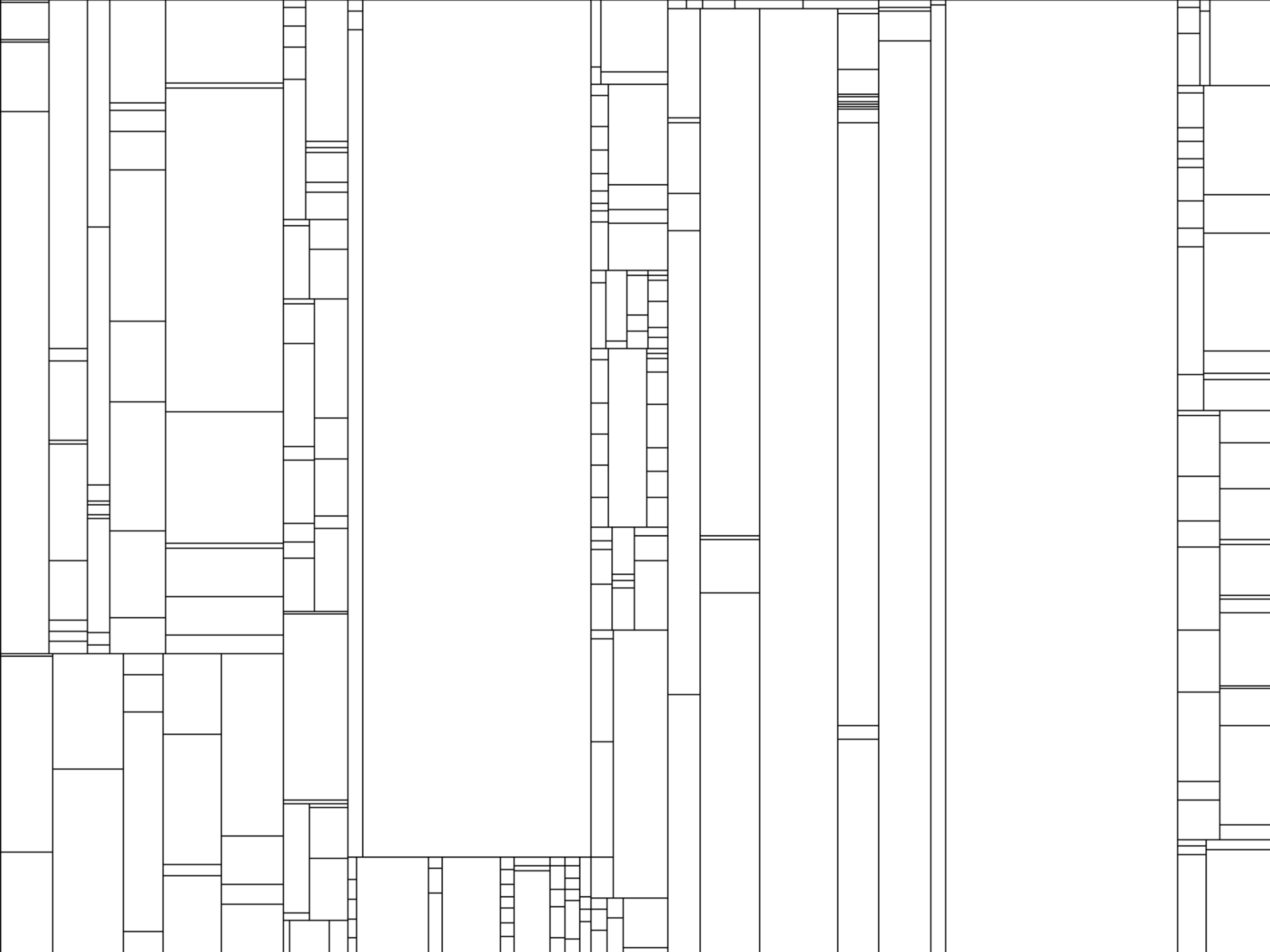
conQAT

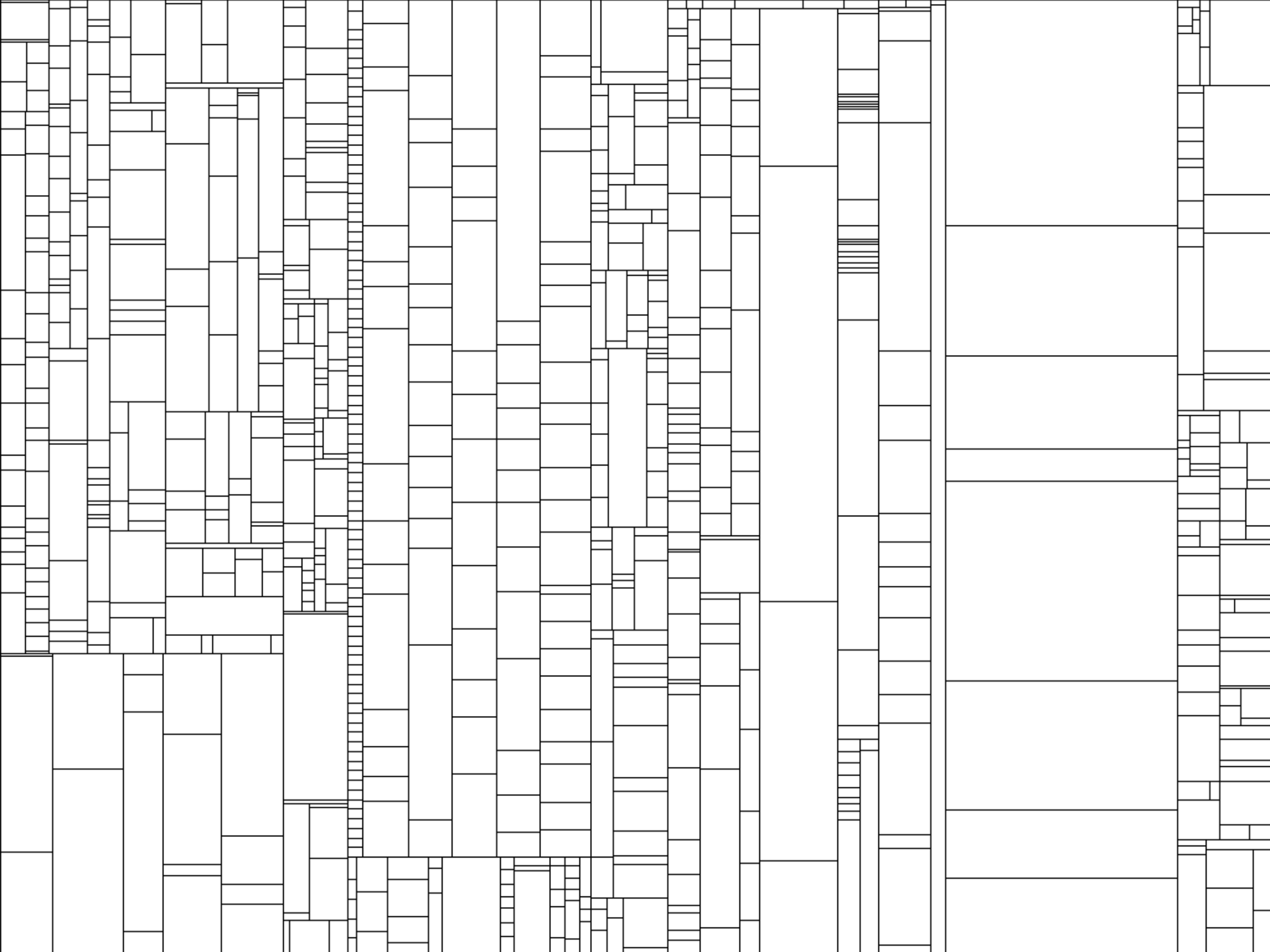
```java
// Utilities for arrays of elements
public String showElements(ModelElement[] elements, String nomsg) {
   boolean found = false;
   StringBuffer res = new StringBuffer();
   if (elements != null) {
      Index.getInstance().setCurrentRenderer(
            FlatReferenceRenderer.getInstance());
      for (int i = 0; i < elements.length; i++) {
         ModelElement el = elements[i];
         res.append(showElementLink(el)).append(HTML.LINE_BREAK);
         found = true;
      }
      Index.getInstance().resetCurrentRenderer();
   }
   if (!found && nomsg != null && nomsg.length() > 0) {
      res.append(HTML.italics(nomsg));
   }
   return res.toString();
}
```

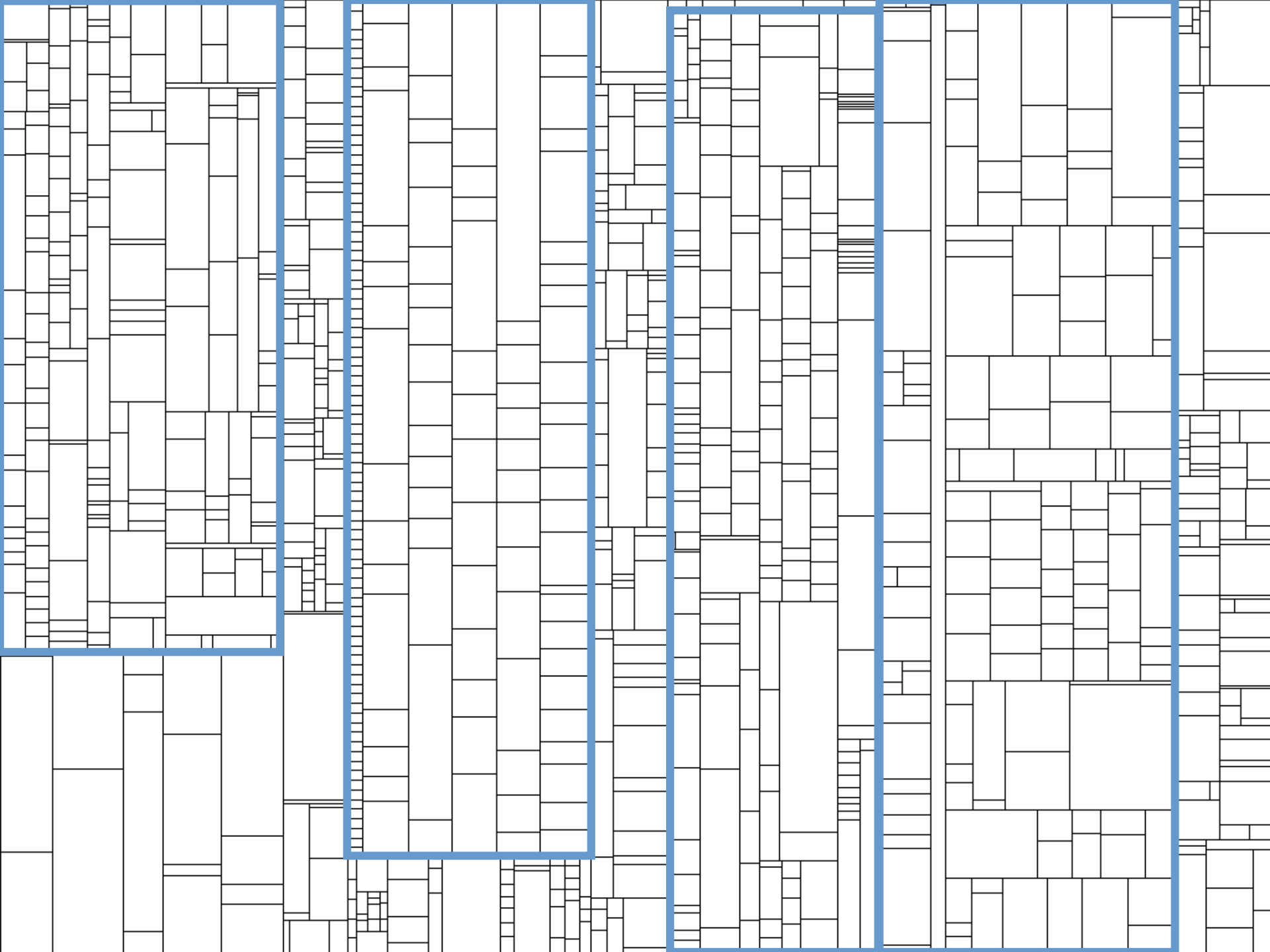


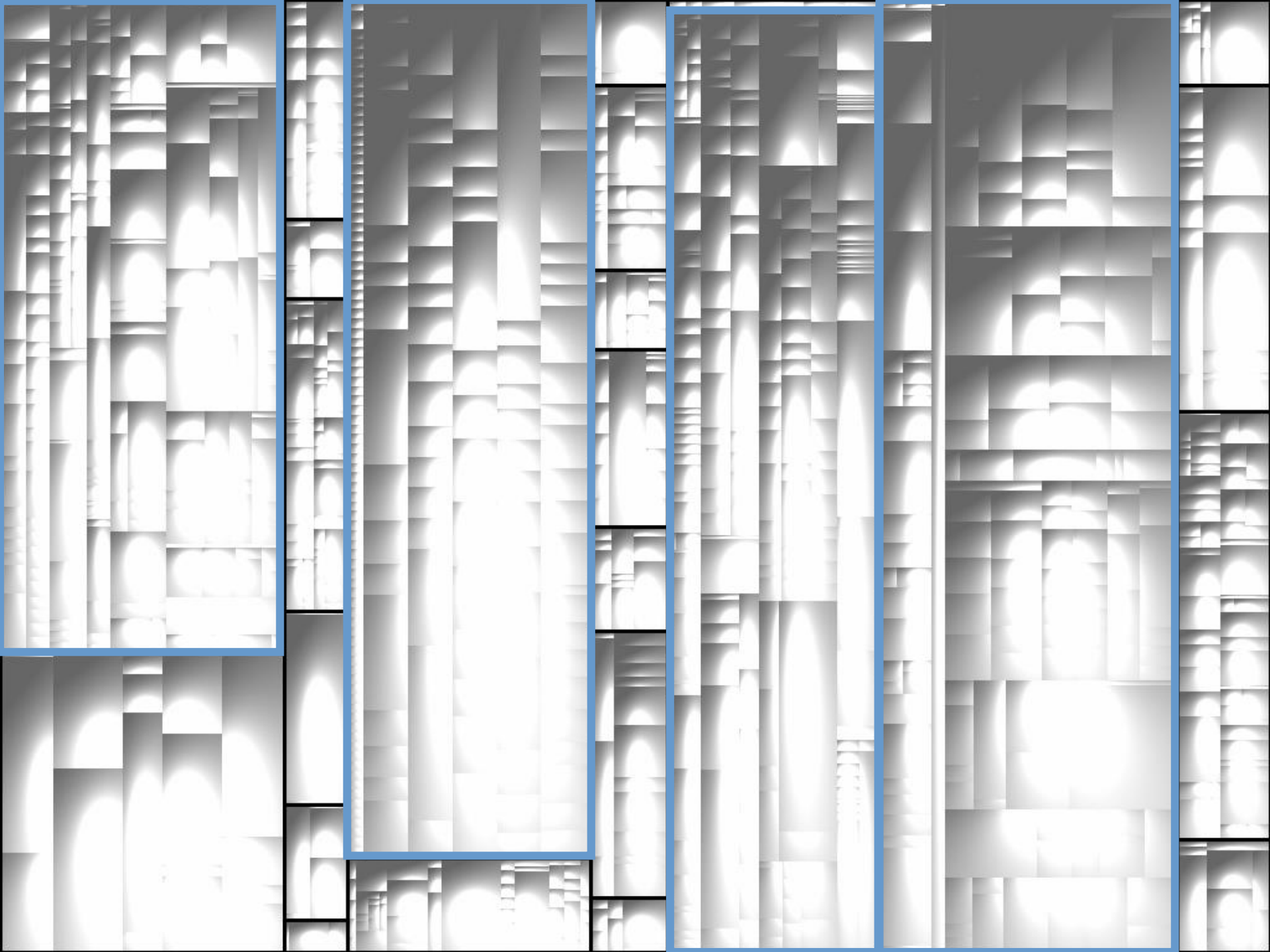


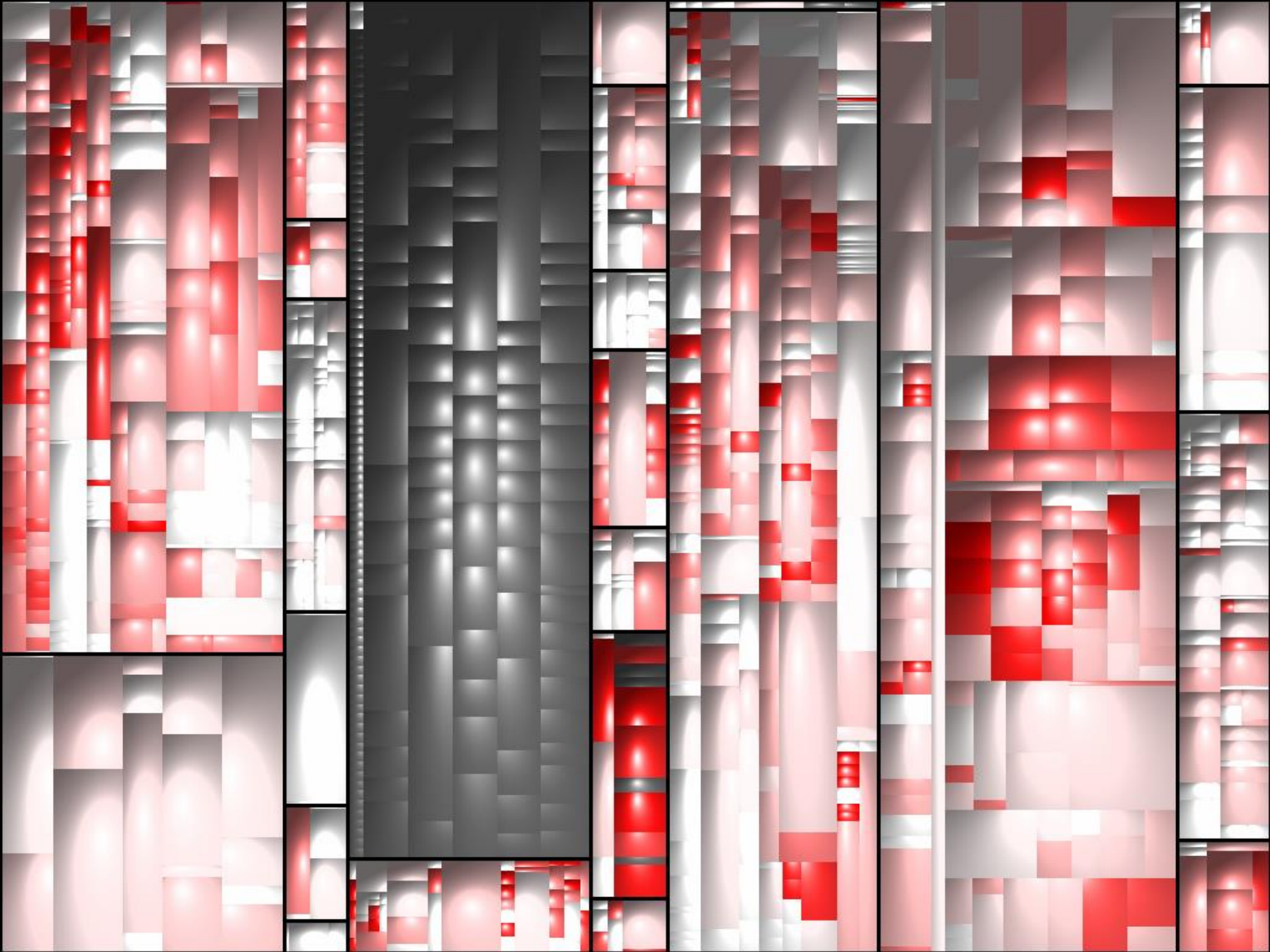

Business-
Objects
60 kloc
DataAccess
67 kloc
Gui.
Base
63 kloc
Gui.
Dialogs
91 kloc



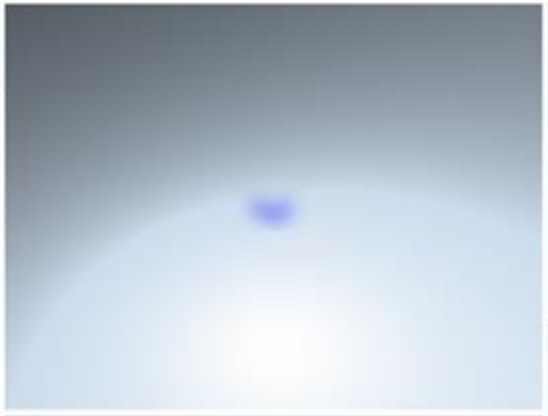


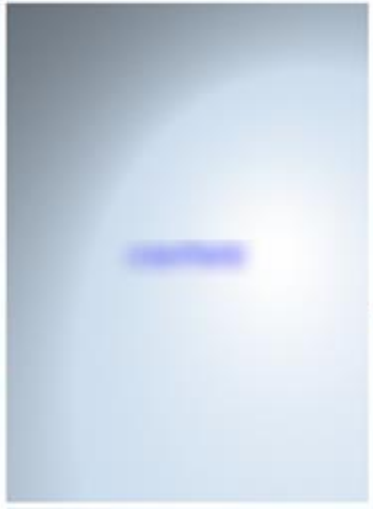
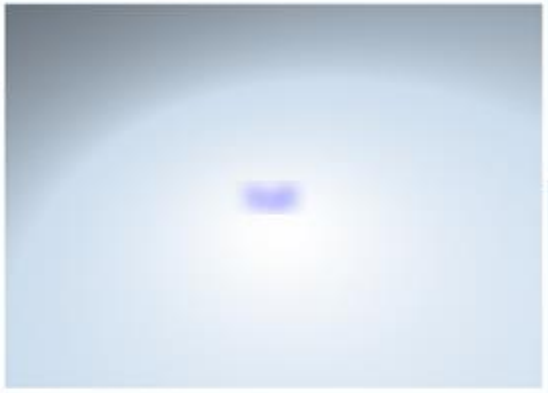


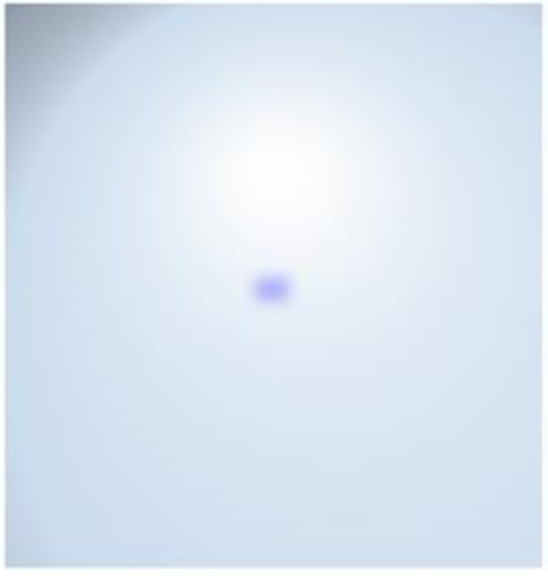


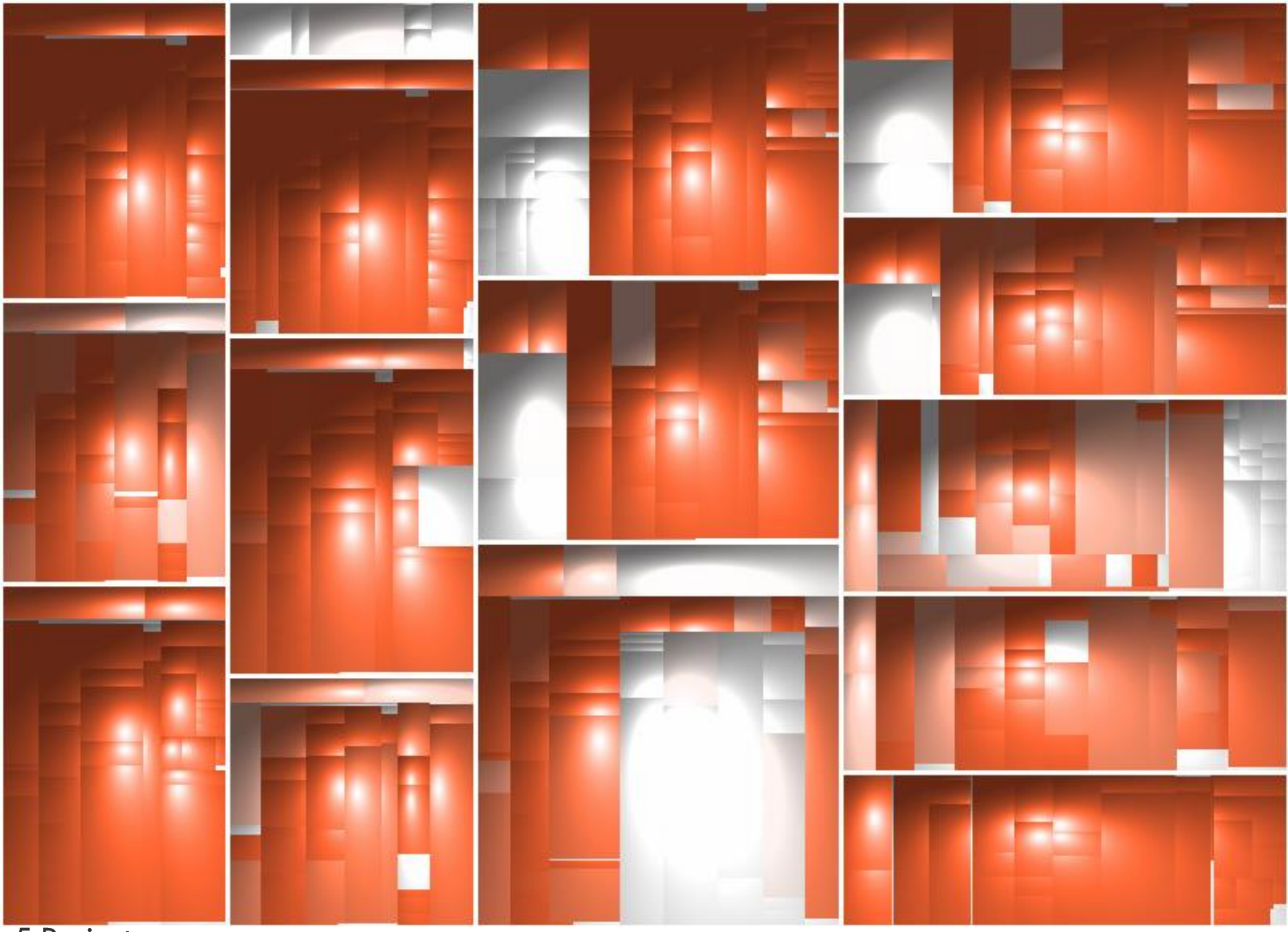


# Produkte einer Produktfamilie

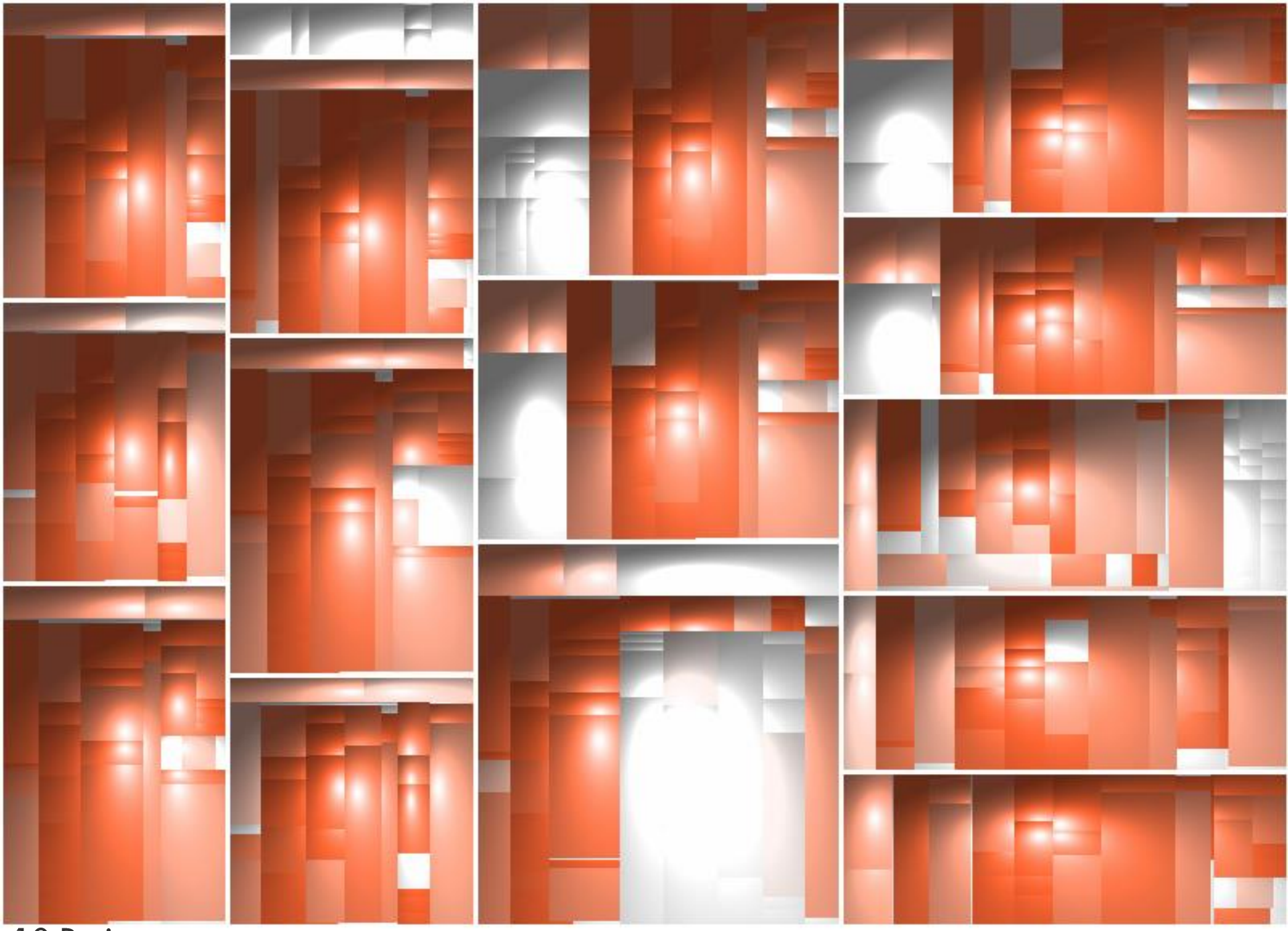


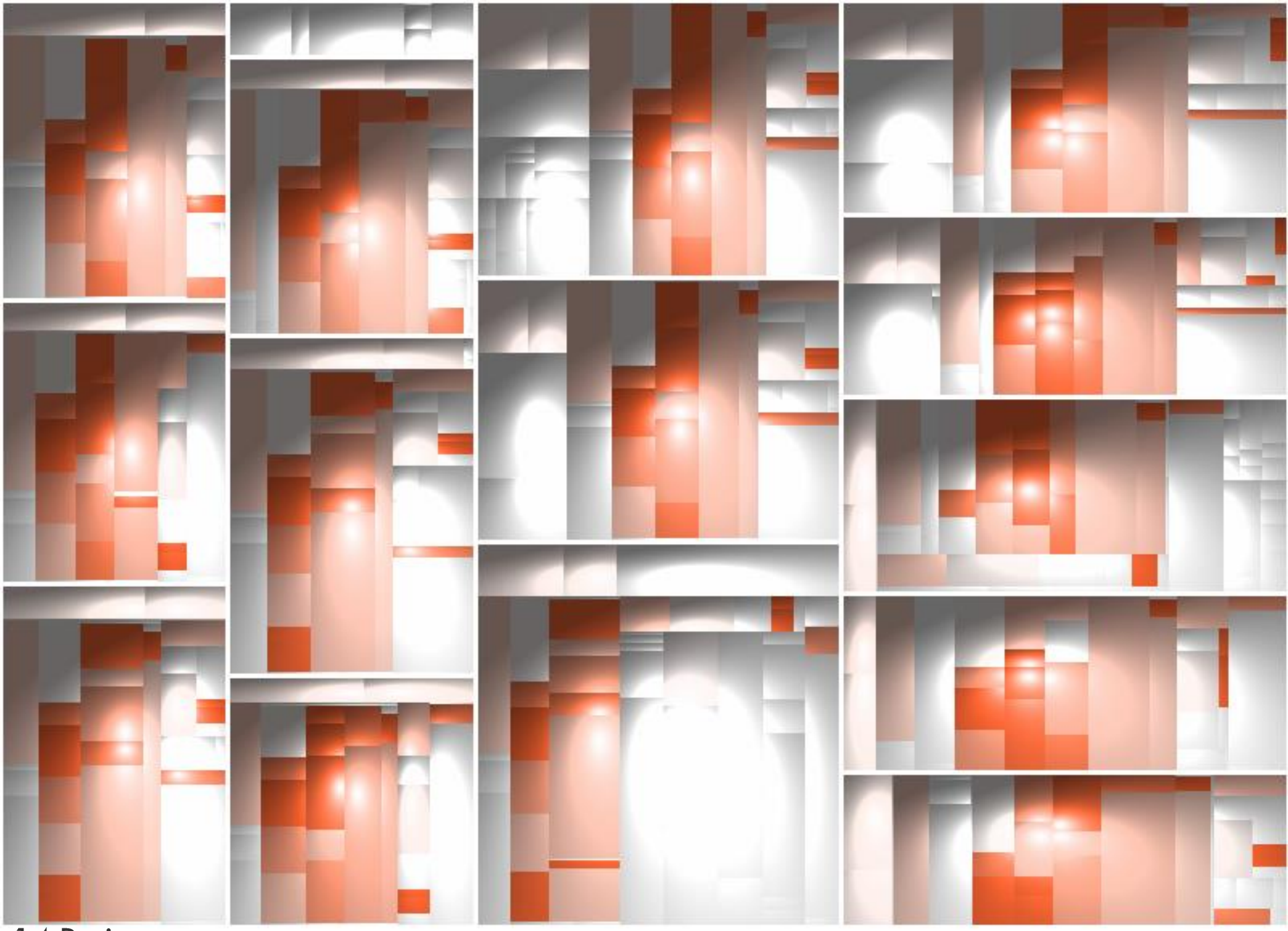




5 Projects

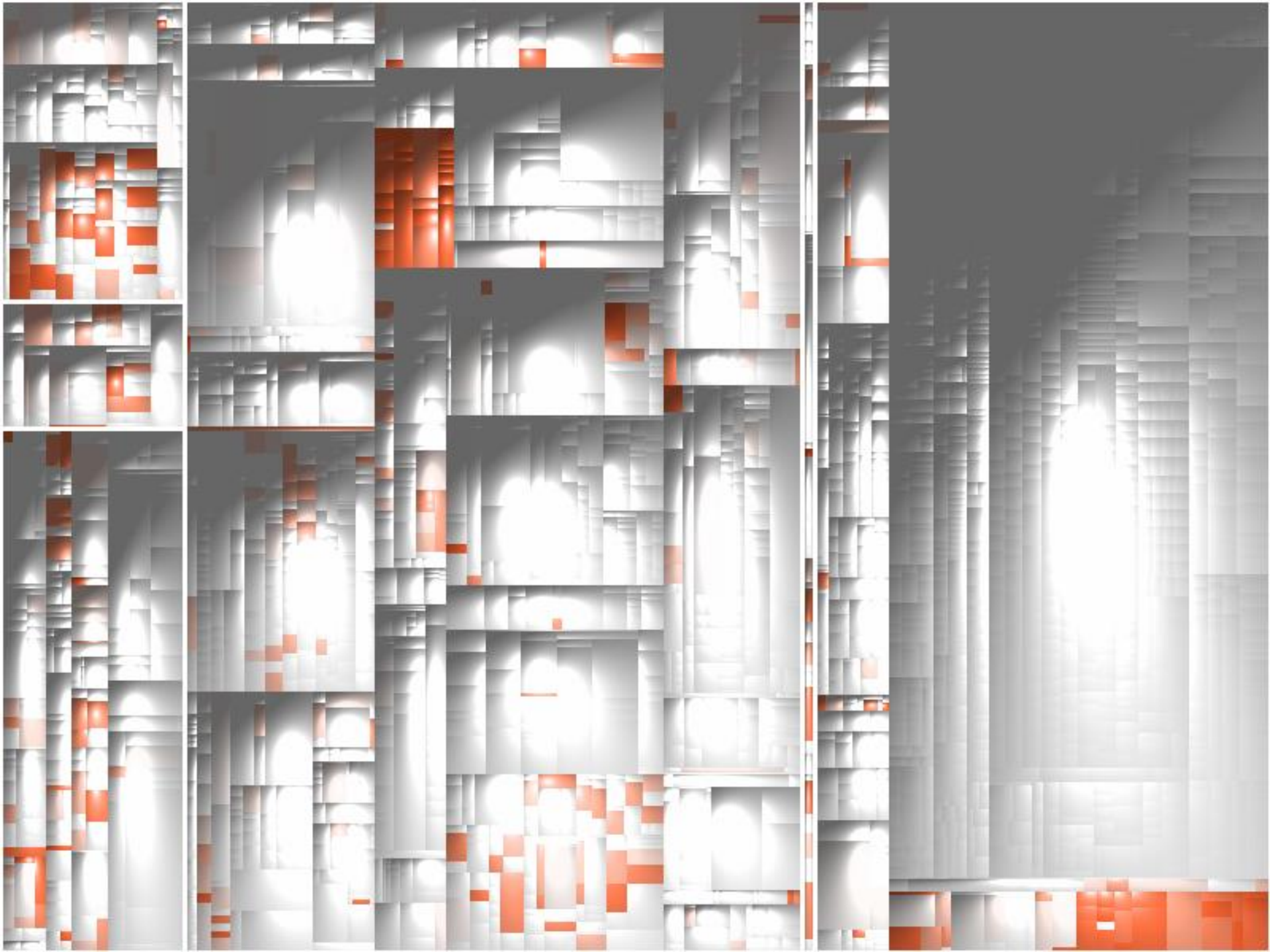


10 Projects

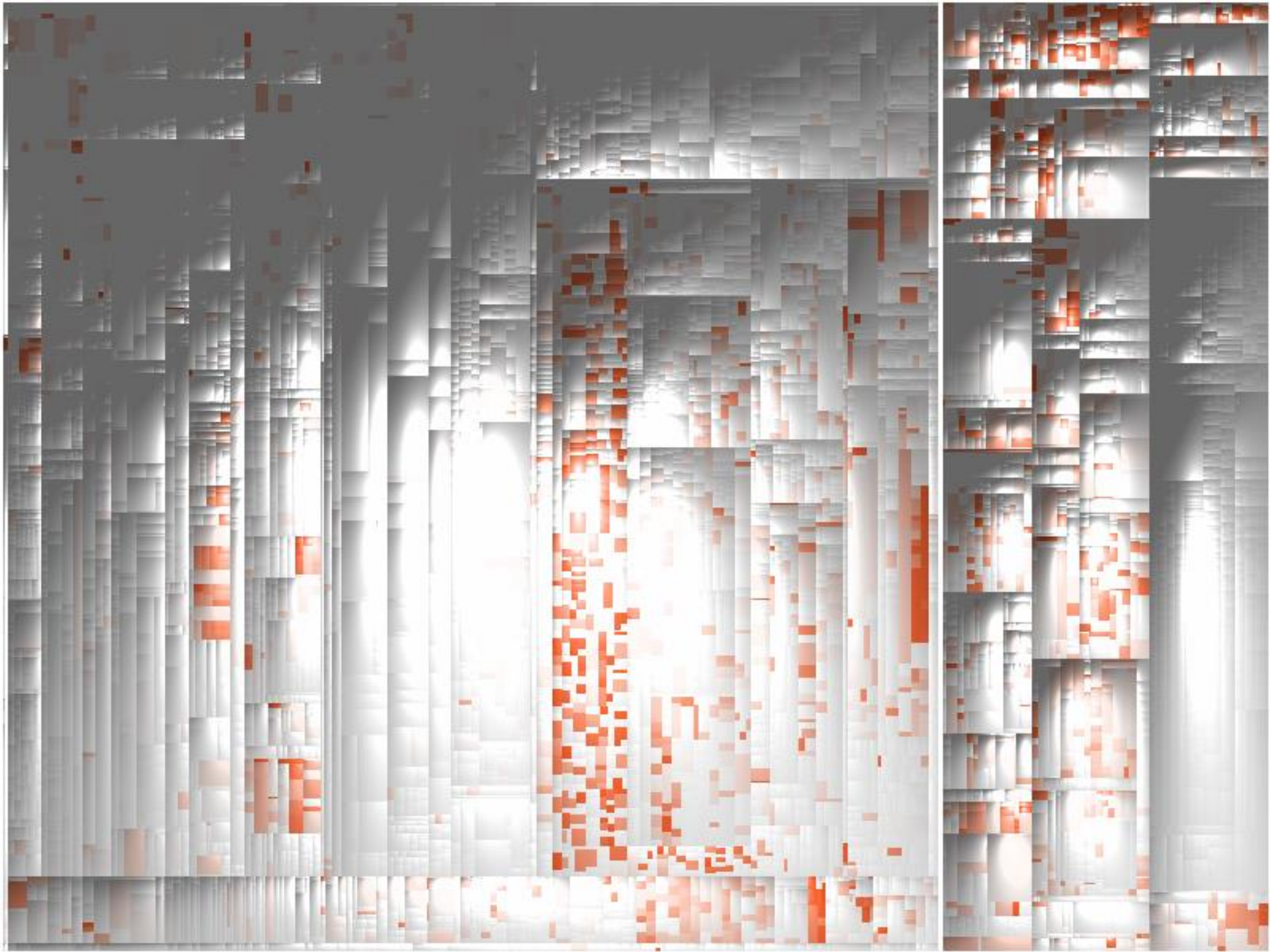


14 Projects

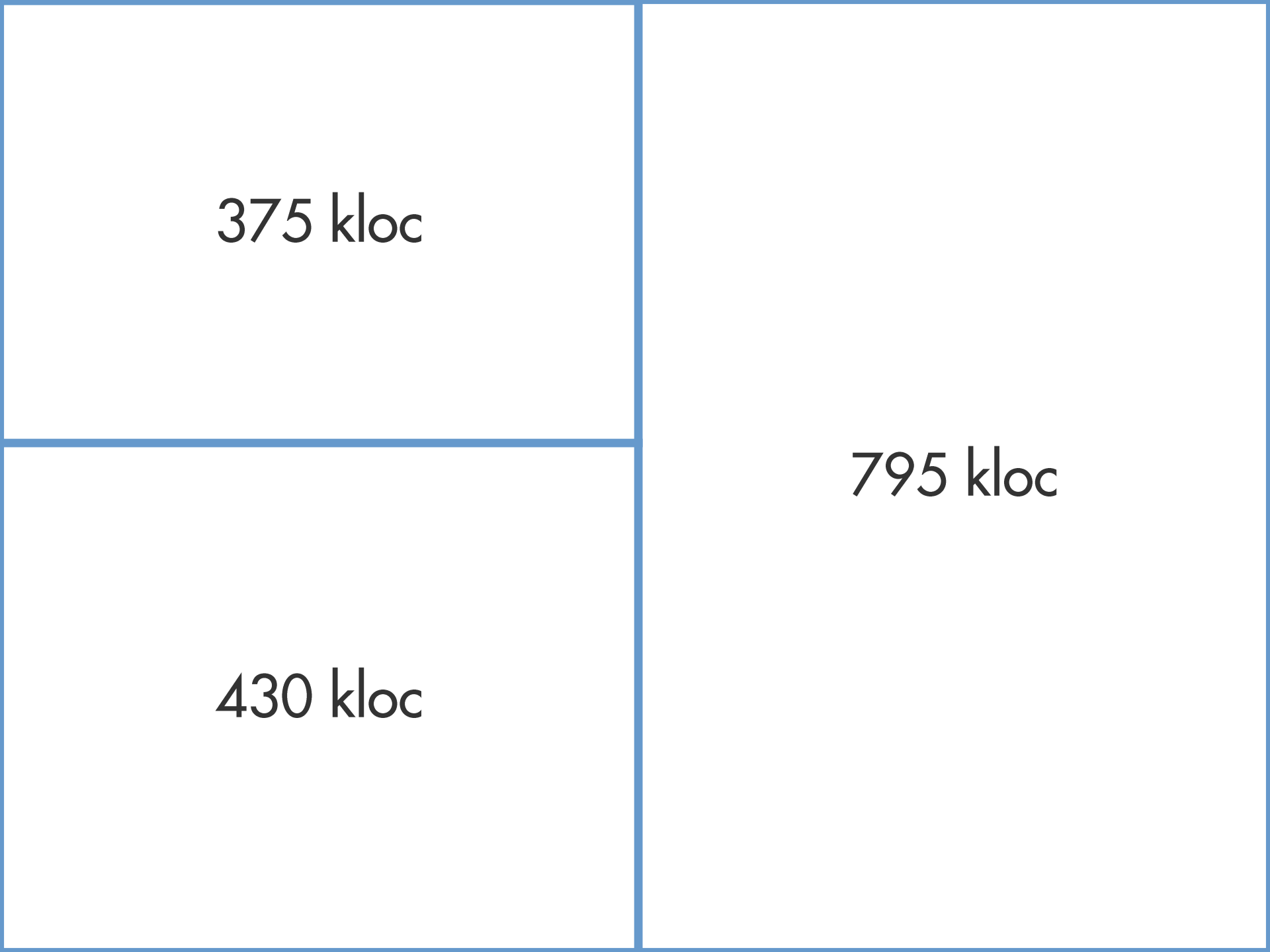
# Produkte einer Produktfamilie II

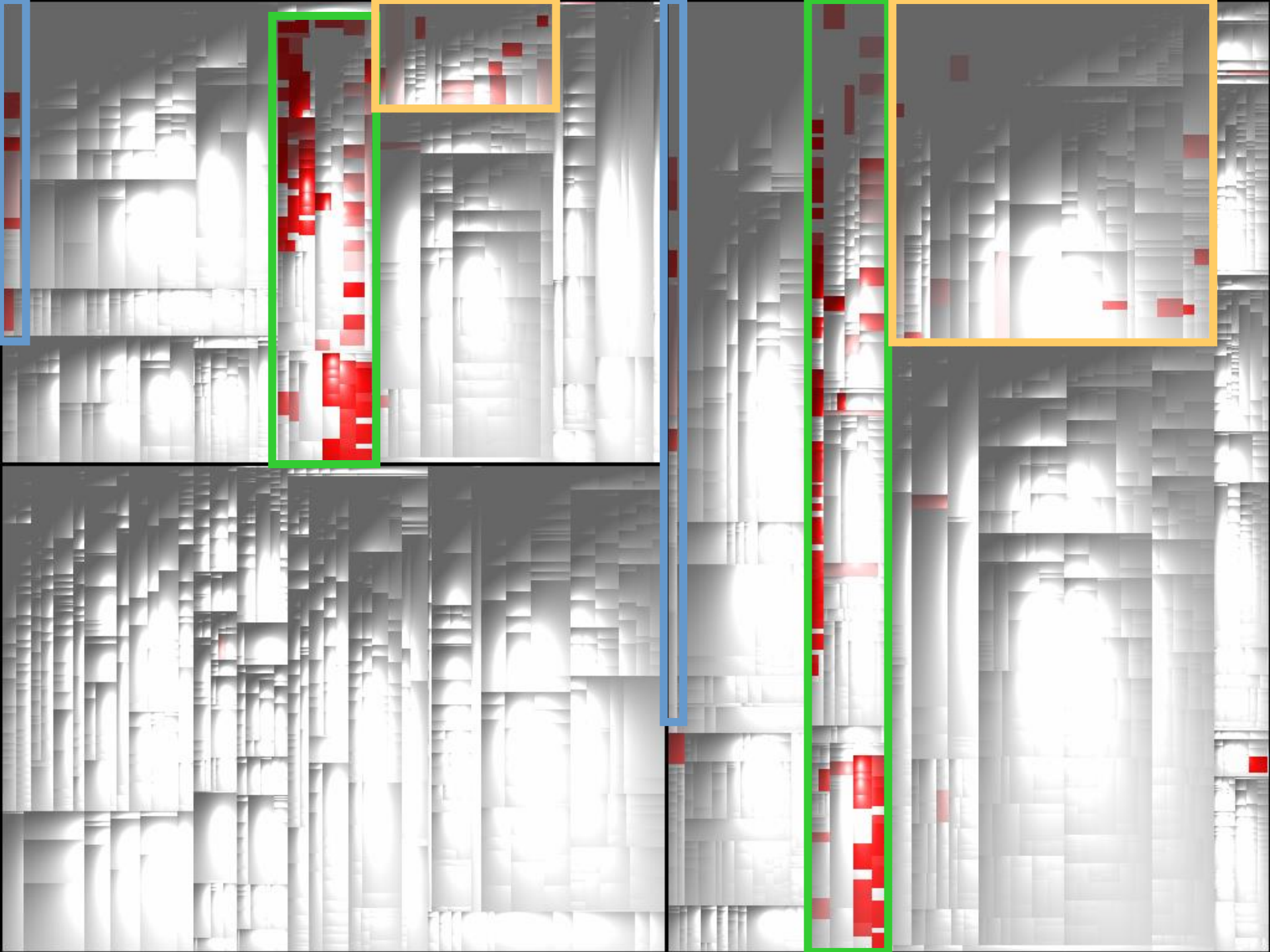


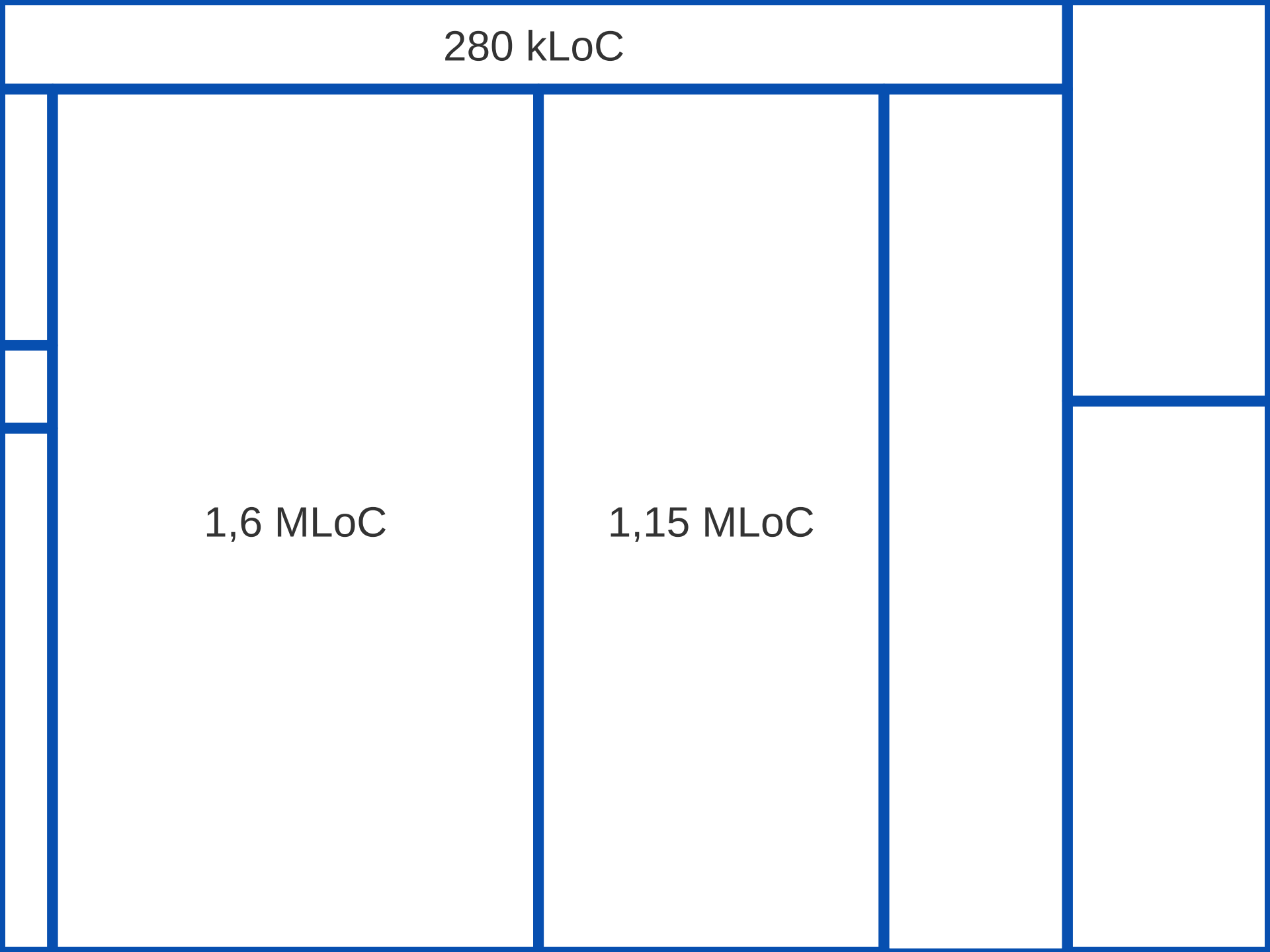
# Architektur einer Produktlinie

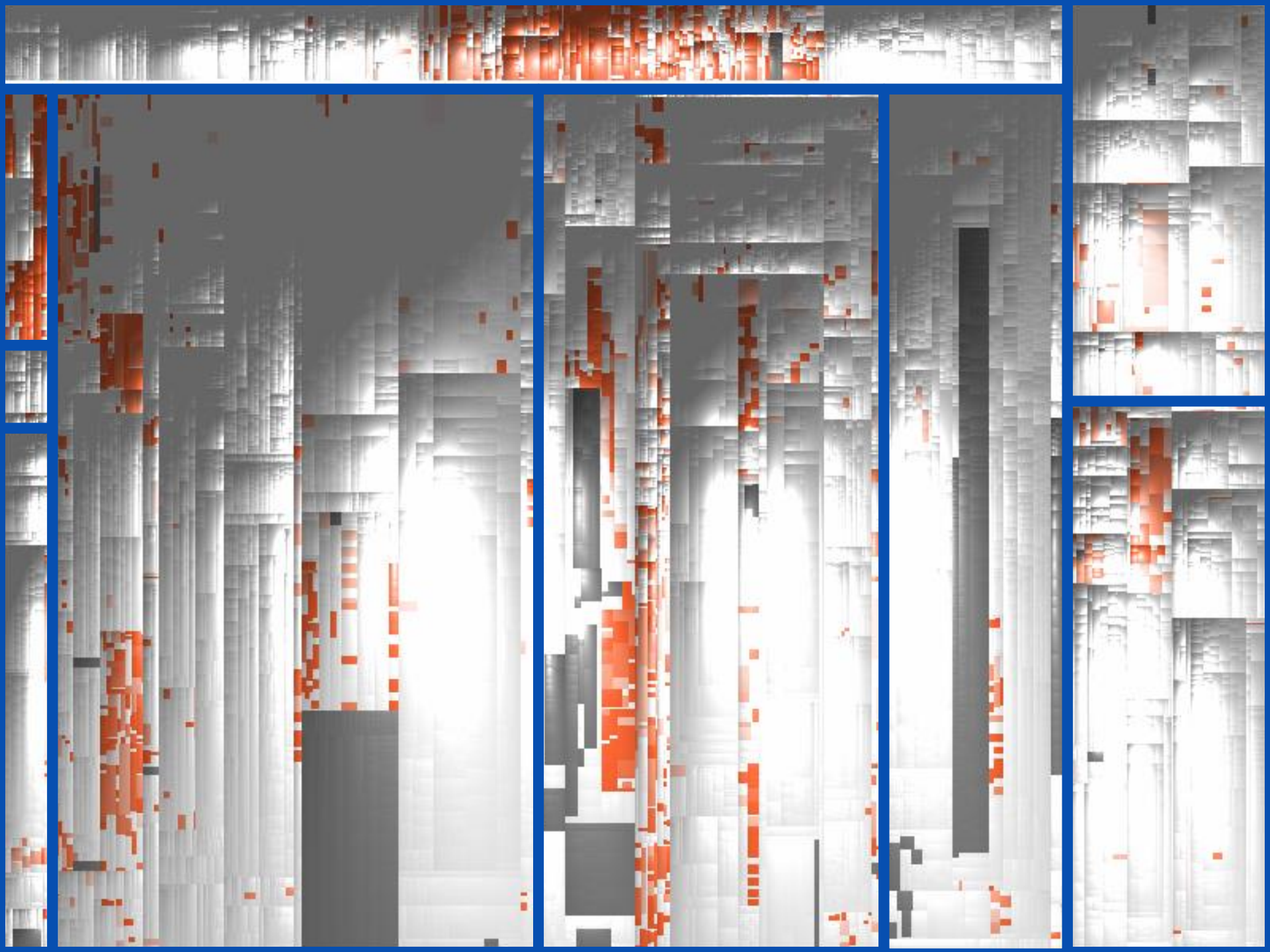


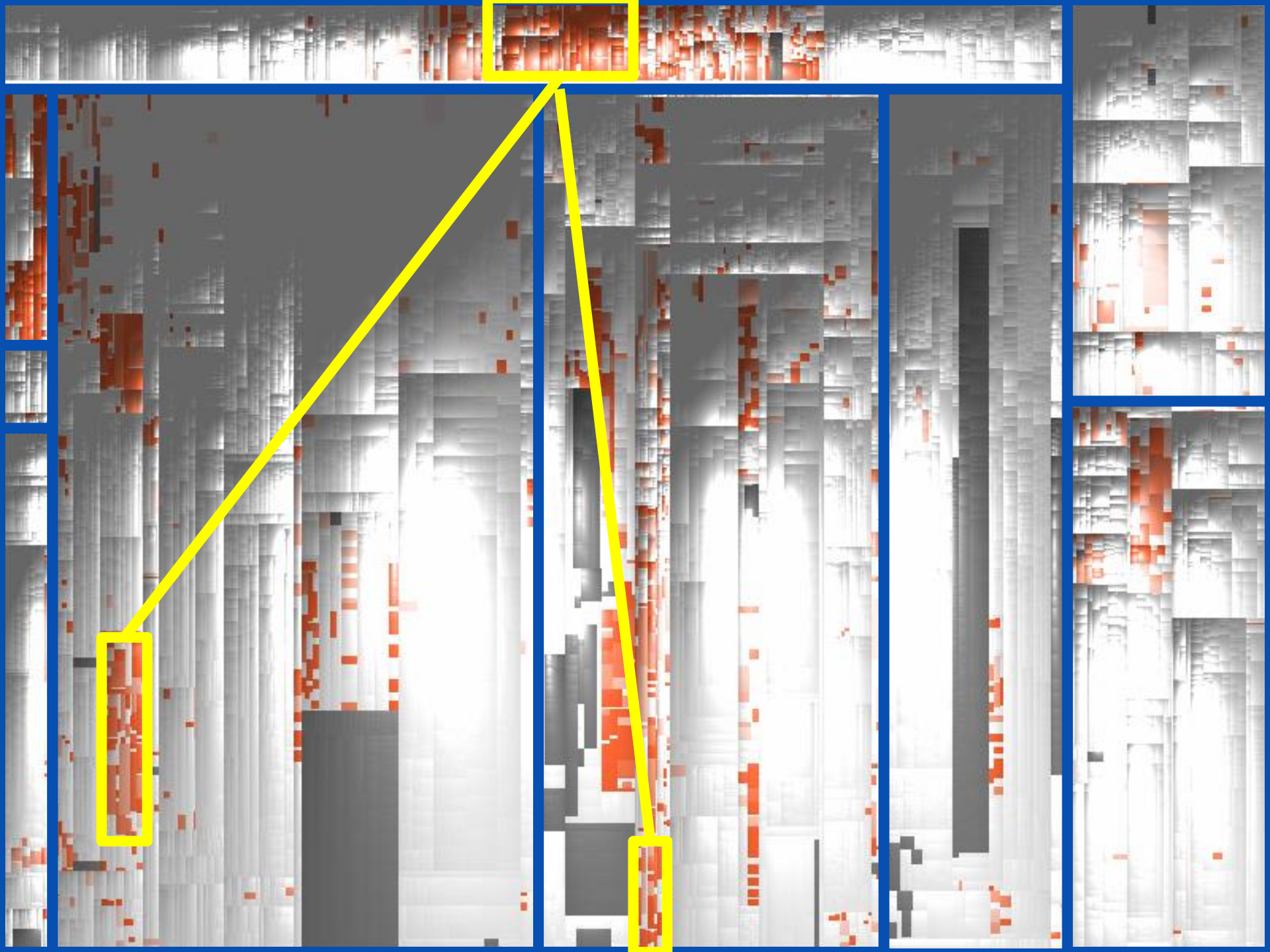
# Mehrere Anwendungen in einer Abteilung

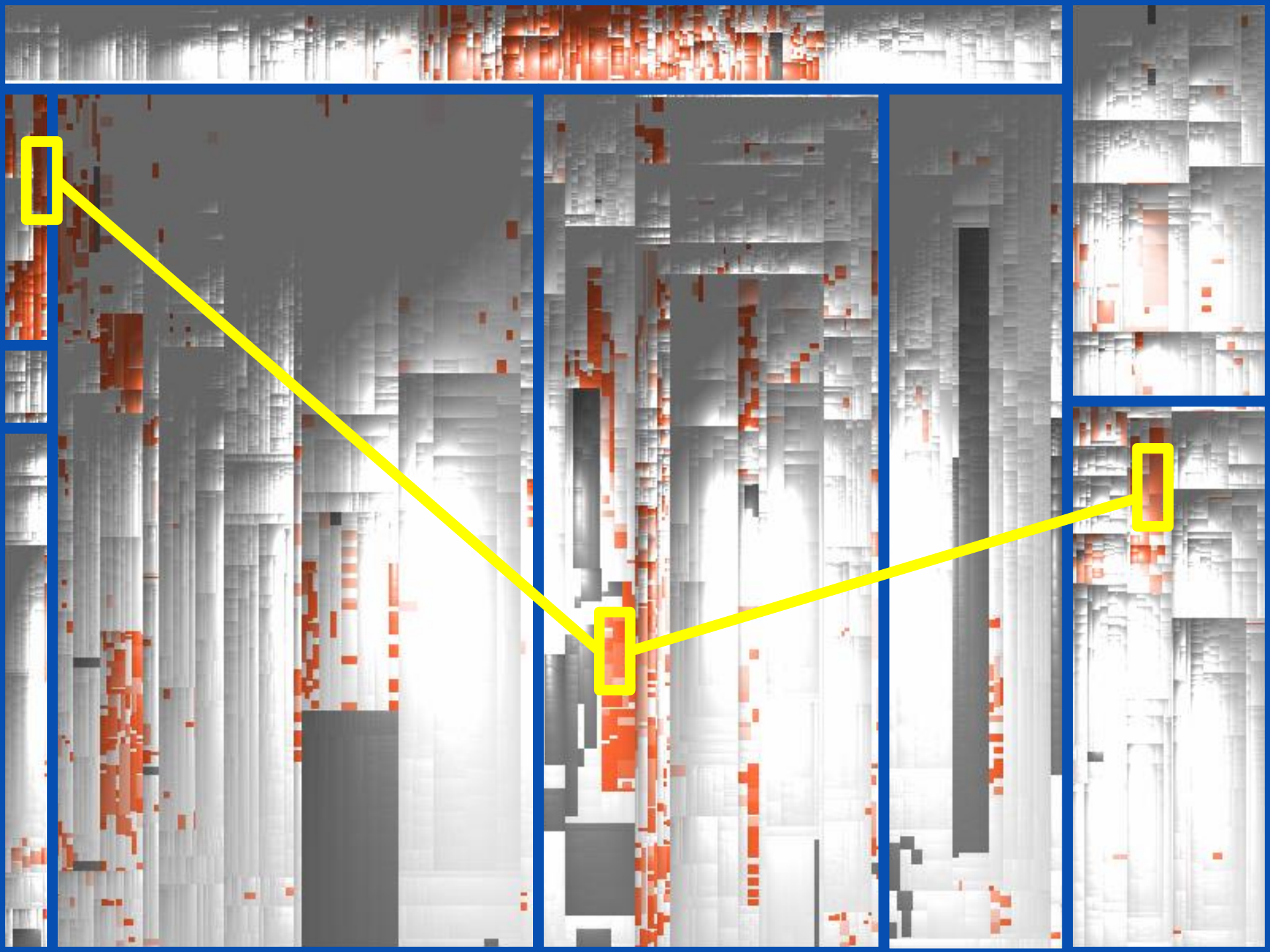

375 kloc
795 kloc
430 kloc




280 kLoC
1,6 MLoC
1,15 MLoC







```
    /* Process the input string received prior to the
    newline. */
    do
    {
        /* Pass the string to FreeRTOS+CLI. */
        xMoreDataToFollow = FreeRTOS_CLIProcessCommand( cInputString, cO

        /* Send the output generated by the command's
        implementation. */
        sendto( xSocket, cOutputString,  strlen( cOutputString ), 0, ( S

    } while( xMoreDataToFollow != pdFALSE ); /* Until the command does n

    /* All the strings generated by the command processing
    have been sent.  Clear the input string ready to receive
    the next command. */
    cInputIndex = 0;
    memset( cInputString, 0x00, cmdMAX_INPUT_SIZE );

    /* Transmit a spacer, just to make the command console
    easier to read. */
    sendto( xSocket, "\r\n",  strlen( "\r\n" ), 0, ( SOCKADDR * ) &xClie
}
else
{
    if( cInChar == '\r' )
    {
        /* Ignore the character.  Newlines are used to
        detect the end of the input string. */
    }
    else if( cInChar == '\b' )
    {
        /* Backspace was pressed.  Erase the last character
        in the string - if any. */
        if( cInputIndex > 0 )
        {
            cInputIndex--;
            cInputString[ cInputIndex ] = '\0';
        }
    }
    else
    {
```

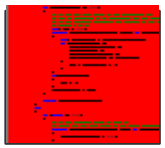
```
    /* Process the input string received prior to the
    newline. */
    do
    {
        /* Pass the string to FreeRTOS+CLI. */
        xMoreDataToFollow = FreeRTOS_CLIProcessCommand( cInputString,

        /* Send the output generated by the command's
        implementation. */
        sendto( xSocket, cOutputString,  strlen( cOutputString ), 0,

    } while( xMoreDataToFollow != pdFALSE ); /* Until the command does

    /* All the strings generated by the command processing
    have been sent.  Clear the input string ready to receive
    the next command. */
    cInputIndex = 0;
    memset( cInputString, 0x00, cmdMAX_INPUT_SIZE );

    /* Transmit a spacer, just to make the command console
    easier to read. */
    sendto( xSocket, "\r\n",  strlen( "\r\n" ), 0, ( SOCKADDR * ) &xC
}
else
{
    if( cInChar == '\r' )
    {
        /* Ignore the character.  Newlines are used to
        detect the end of the input string. */
    }
    else if( ( cInChar == '\b' ) || ( cInChar == cmdASCII_DEL ) )
    {
        /* Backspace was pressed.  Erase the last character
        in the string - if any. */
        if( cInputIndex > 0 )
        {
            cInputIndex--;
            cInputString[ cInputIndex ] = '\0';
        }
    }
    else
    {
```

JabRef/java/net/sf/jabref/Util.java

```
if (fileParts.length > 1) {

    for (int i = 0; i < fileParts.length - 1; i++) {

        String dirToProcess = fileParts[i];

        dirToProcess = expandBrackets(dirToProcess, entry, databas

        if (dirToProcess.matches("^.:$")) { // Windows Drive Lette
            directory = new File(dirToProcess + "/");
            continue;
        }
        if (dirToProcess.equals(".")) { // Stay in current directo
            continue;
        }
        if (dirToProcess.equals("..")) {
            directory = new File(directory.getParent());
            continue;
        }
        if (dirToProcess.equals("*")) { // Do for all direct subdi

            File[] subDirs = directory.listFiles();
            if (subDirs == null)
                return null; // No permission?

            String restOfFileString = join(fileParts, "/", i + 1,

            for (int sub = 0; sub < subDirs.length; sub++) {
                if (subDirs[sub].isDirectory()) {
```

JabRef/java/net/sf/jabref/external/RegExpFileSearch.java

```
if (fileParts.length > 1) {

    for (int i = 0; i < fileParts.length - 1; i++) {

        String dirToProcess = fileParts[i];
        dirToProcess = Util.expandBrackets(dirToProcess, entry,

        if (dirToProcess.matches("^.:$")) { // Windows Drive Le
            directory = new File(dirToProcess + "/");
            continue;
        }
        if (dirToProcess.equals(".")) { // Stay in current dire
            continue;
        }
        if (dirToProcess.equals("..")) {
            directory = new File(directory.getParent());
            continue;
        }
        if (dirToProcess.equals("*")) { // Do for all direct su

            File[] subDirs = directory.listFiles();
            if (subDirs != null) {
                String restOfFileString = Util.join(fileParts,
                for (int sub = 0; sub < subDirs.length; sub++)
                    if (subDirs[sub].isDirectory()) {
                        res.addAll(findFile(entry, database, su
                            restOfFileString, extensionRegExp))
                    }
                }
```



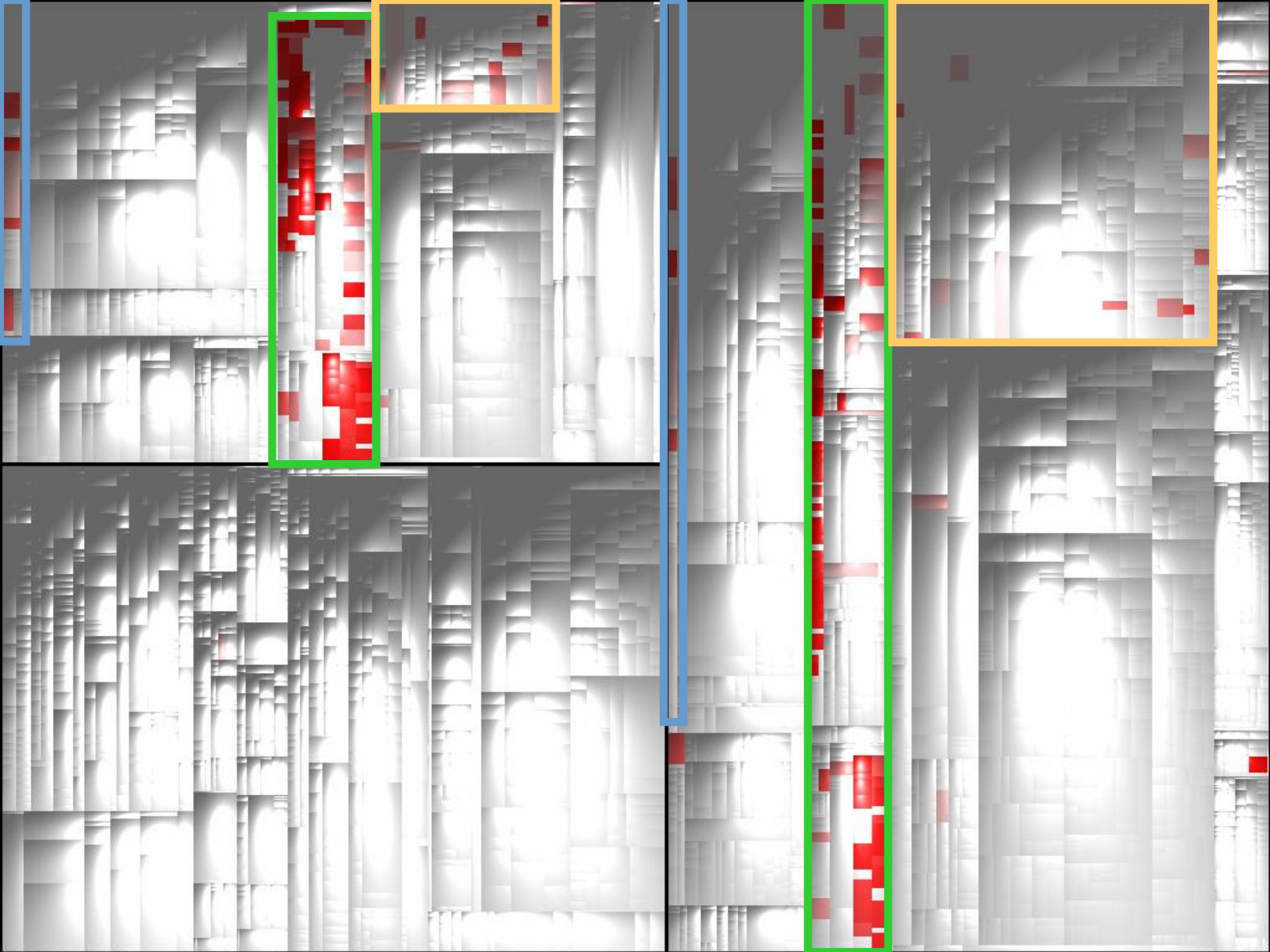
## Studie

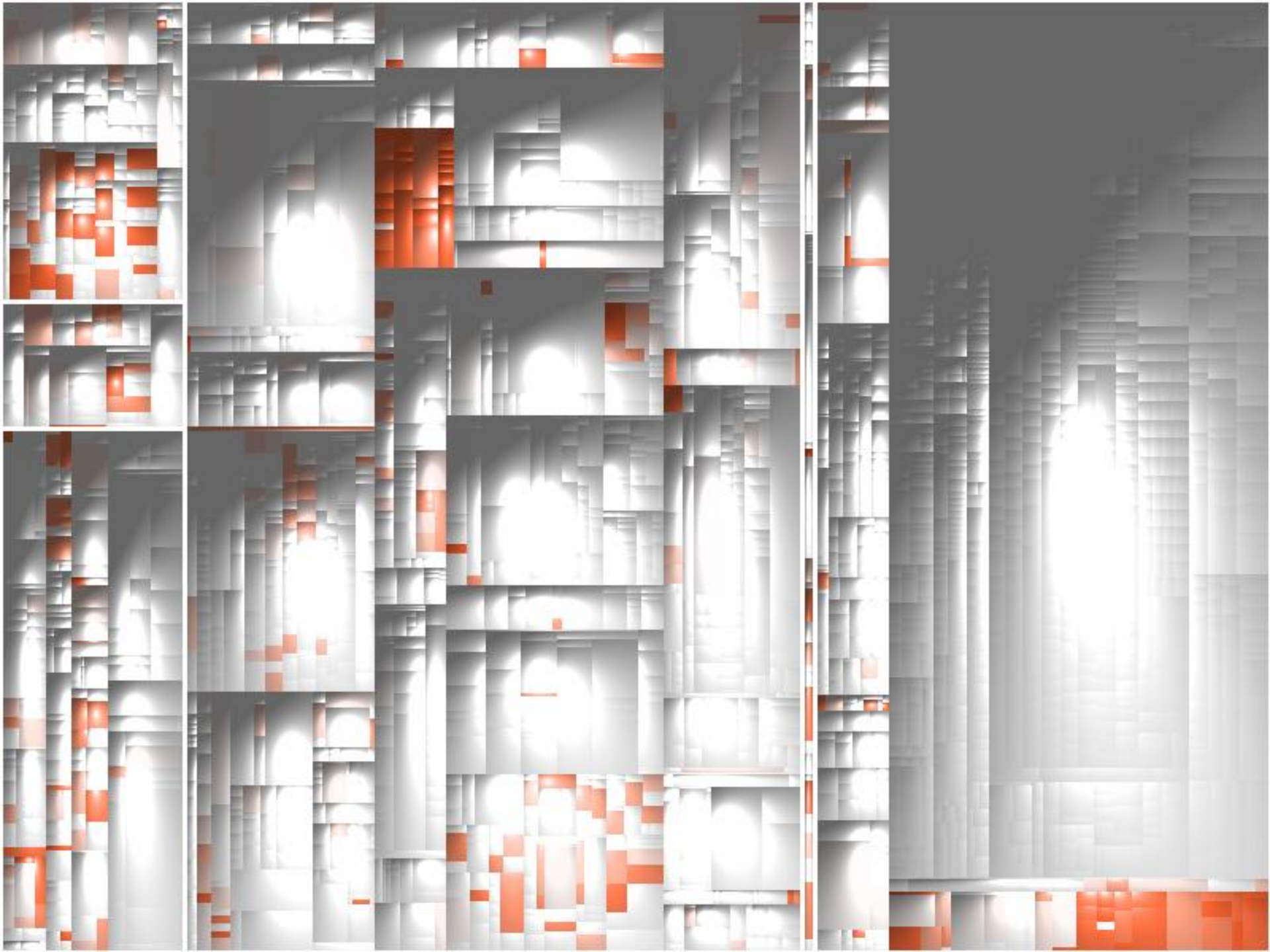
Munich Re LV 1871

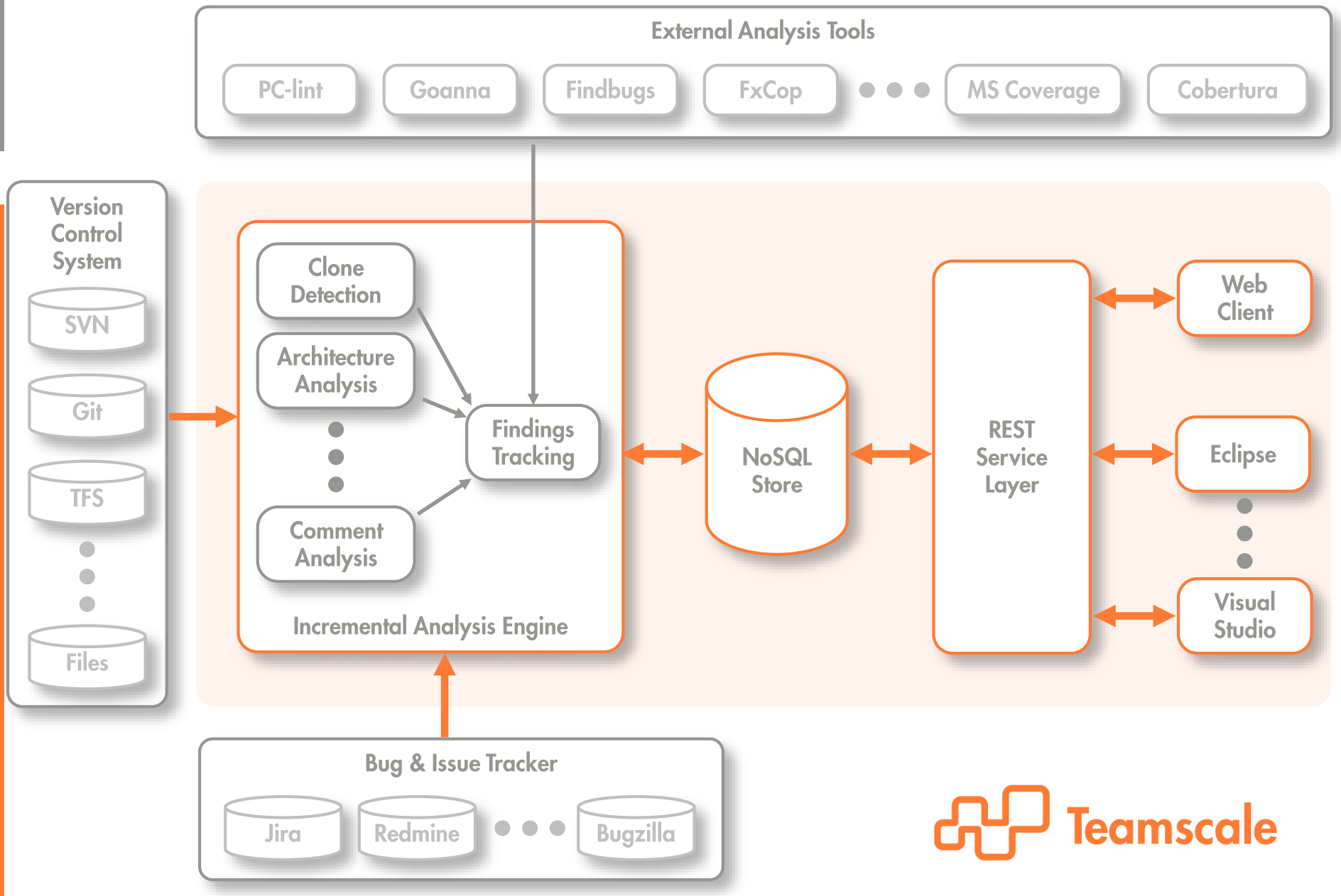
- Über 100 Fehler in produktiver Software

Juergens, Deissenboeck et al: *Do Code Clones Matter?* ICSE 2009

# Und nun?






External Analysis Tools
PC-lint
Goanna
Findbugs
FxCop
MS Coverage
Cobertura
Version Control System
SVN
Git
TFS
Files
Clone Detection
Architecture Analysis
Comment Analysis
Findings Tracking
Incremental Analysis Engine
NoSQL Store
REST Service Layer
Web Client
Eclipse
Visual Studio
Bug & Issue Tracker
Jira
Redmine
Bugzilla
Teamscale

ACT-1270 Fixing Inconsistent handling of serializable process variables

by Victoria King in revision e1aa41b4b133d269980fff3f81d008da8f21a109 (git)

Jun 29 2012
16:05

changed 2 files
-4 findings

ACT-1258 Merging Pablo's work into trunk

by Jacob Nelson in revision 9e664a1f0676cedcbe03415a253e8c3e4a58944c (git)

Jun 29 2012
14:41

added 3 files, changed 2 files
-1 findings

Fix for ACT-1059: Task#setDelegationState(DelegationState) was not saved in database

by Michael Harris in revision 1f48dcad04bc4a621e60af047fb121ae161bca30 (git)

Jun 28 2012
21:45

changed 3 files
+2 findings

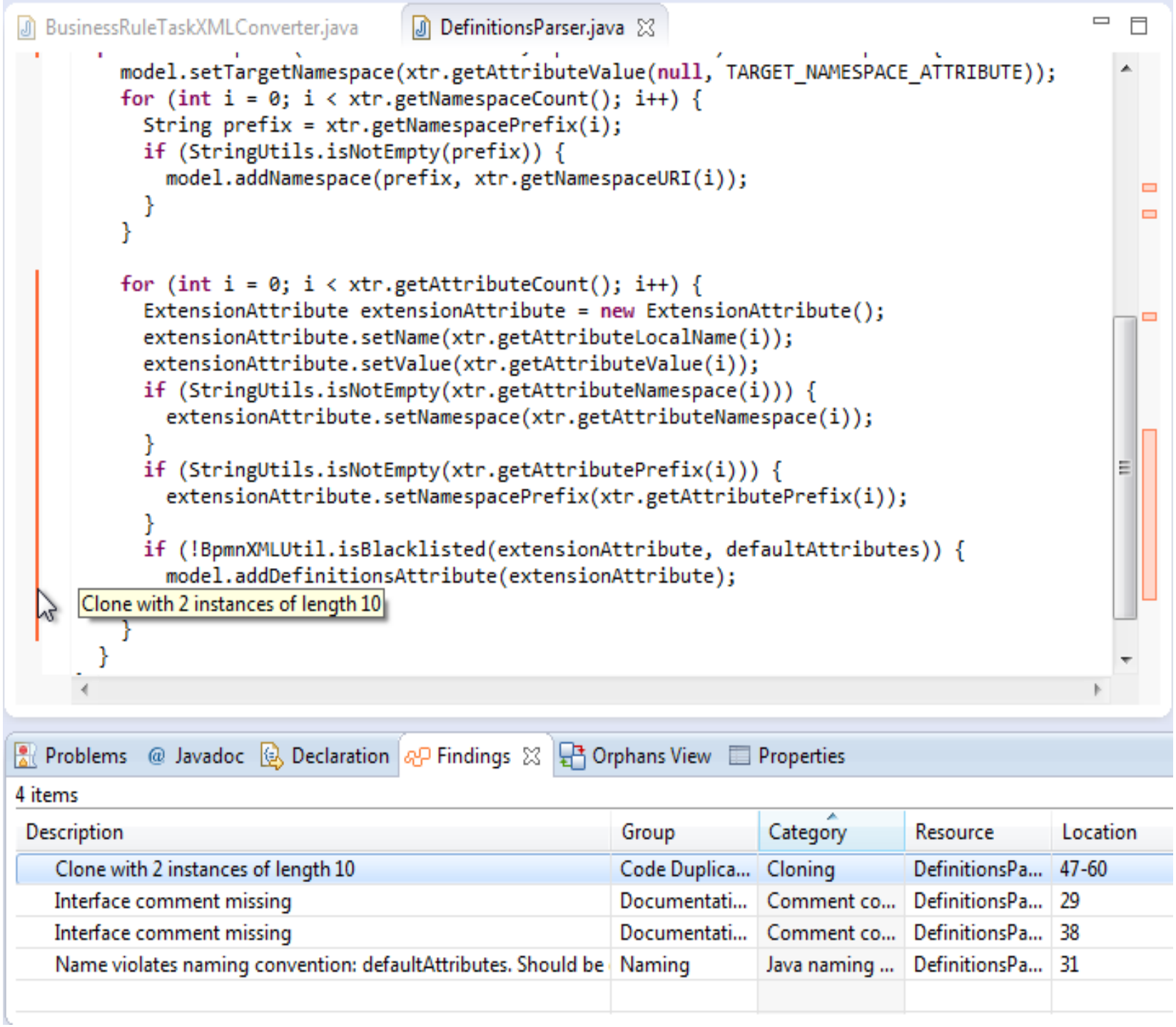
ACT-991 Removed user id from exception message in order not to leak sensitive information

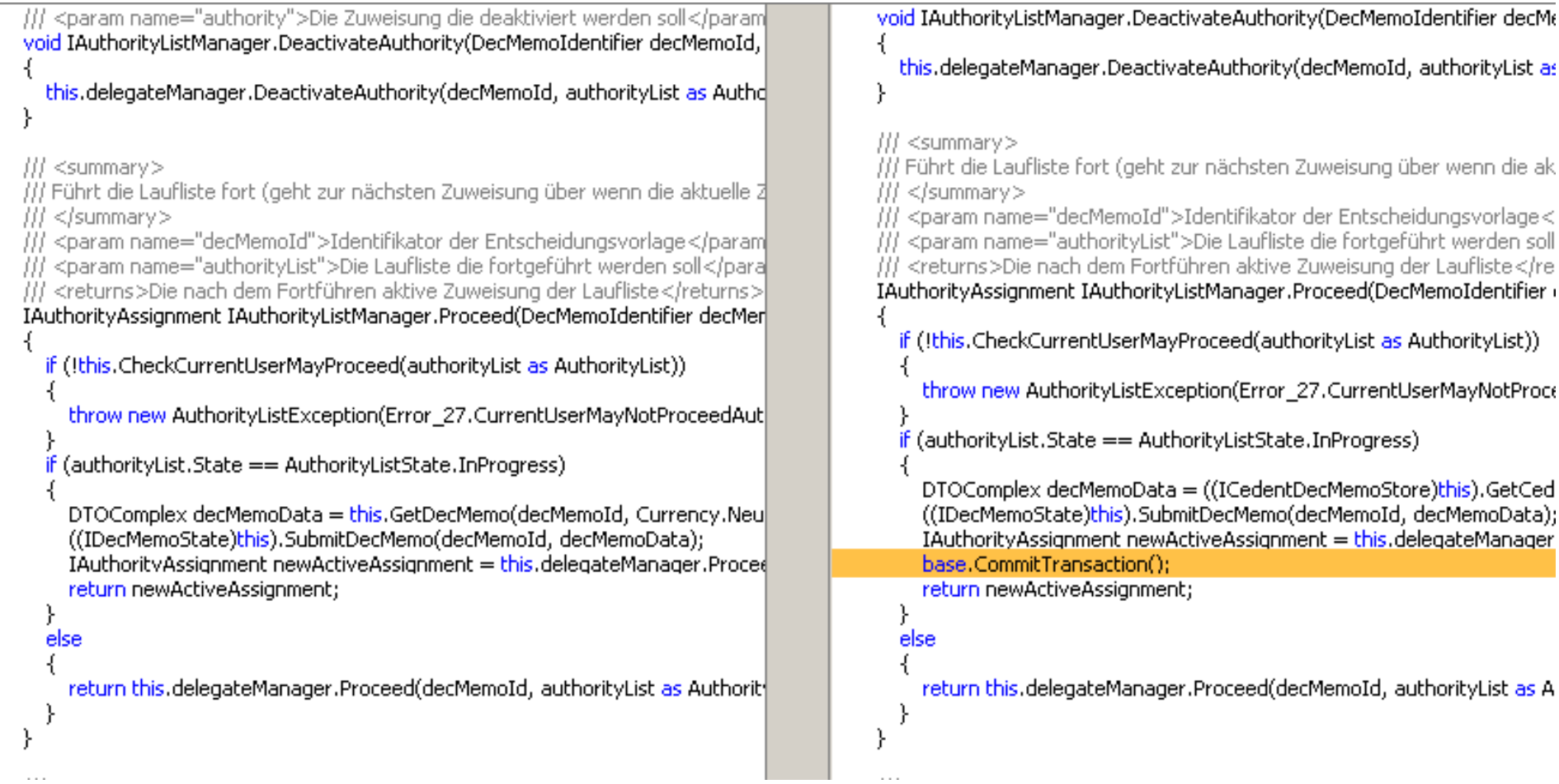
by Michael Harris in revision e9a09424e6309c854c44ac5d08740a8ffb082fc9 (git)

Jun 28 2012
15:26

changed 2 files
-2 findings

Teamscale


BusinessRuleTaskXMLConverter.java
DefinitionsParser.java
model.setTargetNamespace(xtr.getAttributeValue(null, TARGET_NAMESPACE_ATTRIBUTE));
for (int i = 0; i < xtr.getNamespaceCount(); i++) {
String prefix = xtr.getNamespacePrefix(i);
if (StringUtils.isNotEmpty(prefix)) {
model.addNamespace(prefix, xtr.getNamespaceURI(i));
}
}
for (int i = 0; i < xtr.getAttributeCount(); i++) {
ExtensionAttribute extensionAttribute = new ExtensionAttribute();
extensionAttribute.setName(xtr.getAttributeLocalName(i));
extensionAttribute.setValue(xtr.getAttributeValue(i));
if (StringUtils.isNotEmpty(xtr.getAttributeNamespace(i))) {
extensionAttribute.setNamespace(xtr.getAttributeNamespace(i));
}
if (StringUtils.isNotEmpty(xtr.getAttributePrefix(i))) {
extensionAttribute.setNamespacePrefix(xtr.getAttributePrefix(i));
}
if (!BpmnXMLUtil.isBlacklisted(extensionAttribute, defaultAttributes)) {
model.addDefinitionsAttribute(extensionAttribute);
Clone with 2 instances of length 10
}
}
Problems
Javadoc
Declaration
Findings
Orphans View
Properties
4 items
Description
Group
Category
Resource
Location
Clone with 2 instances of length 10
Code Duplica...
Cloning
DefinitionsPa...
47-60
Interface comment missing
Documentati...
Comment co...
DefinitionsPa...
29
Interface comment missing
Documentati...
Comment co...
DefinitionsPa...
38
Name violates naming convention: defaultAttributes. Should be
Naming
Java naming ...
DefinitionsPa...
31



```
/// <param name="authority">Die Zuweisung die deaktiviert werden soll</param
void IAuthorityListManager.DeactivateAuthority(DecMemoIdentifier decMemoId,
{
   this.delegateManager.DeactivateAuthority(decMemoId, authorityList as Autho
}

/// <summary>
/// Führt die Laufliste fort (geht zur nächsten Zuweisung über wenn die aktuelle Z
/// </summary>
/// <param name="decMemoId">Identifikator der Entscheidungsvorlage</param
/// <param name="authorityList">Die Laufliste die fortgeführt werden soll</para
/// <returns>Die nach dem Fortführen aktive Zuweisung der Laufliste</returns>
IAuthorityAssignment IAuthorityListManager.Proceed(DecMemoIdentifier decMem
{
   if (!this.CheckCurrentUserMayProceed(authorityList as AuthorityList))
   {
      throw new AuthorityListException(Error_27.CurrentUserMayNotProceedAut
   }
   if (authorityList.State == AuthorityListState.InProgress)
   {
      DTOComplex decMemoData = this.GetDecMemo(decMemoId, Currency.Neu
      ((IDecMemoState)this).SubmitDecMemo(decMemoId, decMemoData);
      IAuthorityAssignment newActiveAssignment = this.delegateManager.Proce
      return newActiveAssignment;
   }
   else
   {
      return this.delegateManager.Proceed(decMemoId, authorityList as Authorit
   }
}
...
```

```
void IAuthorityListManager.DeactivateAuthority(DecMemoIdentifier decMe
{
   this.delegateManager.DeactivateAuthority(decMemoId, authorityList as
}

/// <summary>
/// Führt die Laufliste fort (geht zur nächsten Zuweisung über wenn die ak
/// </summary>
/// <param name="decMemoId">Identifikator der Entscheidungsvorlage<
/// <param name="authorityList">Die Laufliste die fortgeführt werden soll
/// <returns>Die nach dem Fortführen aktive Zuweisung der Laufliste</re
IAuthorityAssignment IAuthorityListManager.Proceed(DecMemoIdentifier
{
   if (!this.CheckCurrentUserMayProceed(authorityList as AuthorityList))
   {
      throw new AuthorityListException(Error_27.CurrentUserMayNotProce
   }
   if (authorityList.State == AuthorityListState.InProgress)
   {
      DTOComplex decMemoData = ((ICedentDecMemoStore)this).GetCed
      ((IDecMemoState)this).SubmitDecMemo(decMemoId, decMemoData);
      IAuthorityAssignment newActiveAssignment = this.delegateManager
      base.CommitTransaction();
      return newActiveAssignment;
   }
   else
   {
      return this.delegateManager.Proceed(decMemoId, authorityList as A
   }
}
...
```

fixed: latest change is no longer lost when assigning entry to a keyword group while it is being edited

by jzieren in revision e0ca9a51b50c8b01f579f4eef79028bff6c34028 (git)

May 26 2005
15:58

1 alerts:

| Message | Context |
|---|---|
| Found potential inconsistent clone change in RightClickMenu.java | [Broken clone] [Old clone finding] [Code change] |

2 removed findings:

| Message | Location | Finding Group |
|---|---|---|
| Clone with 2 instances of length 10 | src/java/net/sf/.../RightClickMenu.java:366-380 | Code Duplication / Cloning |
| Clone with 2 instances of length 10 | src/java/net/sf/.../RightClickMenu.java:340-354 | Code Duplication / Cloning |

# www.conqat.org

| Parameter | Attribute value |
| --- | --- |
| input | |
| project | ConQAT |
| dir | D:\code\conqat\engine |
| include | |
| pattern | **/*.java |
| clone | |
| minlength | 50 |
| output | |
| dir | E:\temp\crossclonedetection |
| project | |
| prefix | ConQAT/org.conqat.engine.core |
| project | |
| prefix | ConQAT/org.conqat.engine.code_clones |

# Visit our Teamscale online demo

Get a quick impression of Teamscale and what it can do to help you create high quality code.

Read our brief tutorial to get you started.

VISIT TEAMSCALE ONLINE DEMO

https://www.cqse.eu/en/products/teamscale/try/

Lizenz: Mail an juergens@cqse.eu

Ich schicke Euch eine Lizenz für ½ Jahr

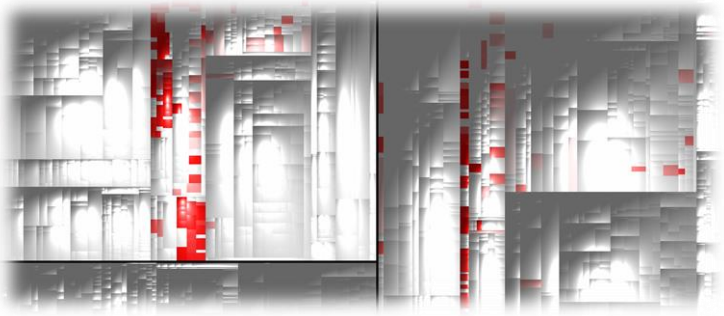
# Fazit

Clone Detection kann fundamentale Probleme in der Team-Organisation oder Architektur aufzeigen.

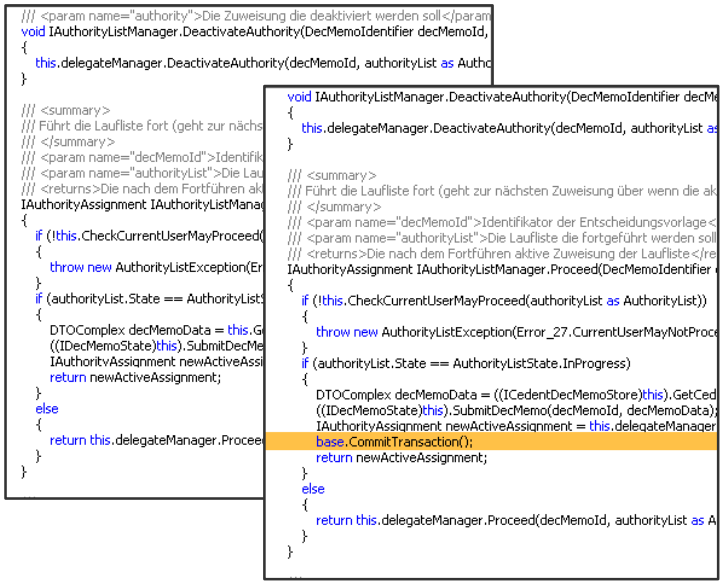
Behebung der Ursachen ist eine Lösung, aber oft auch die aufwändigste. Eine Linderung der Symptome (Clone-Management) schafft Abhilfe, falls das (noch) nicht möglich ist.

# Wie schlimm ist es bei uns?

Schritt 1: Cross-Project Clone Detection auf eigenem Code

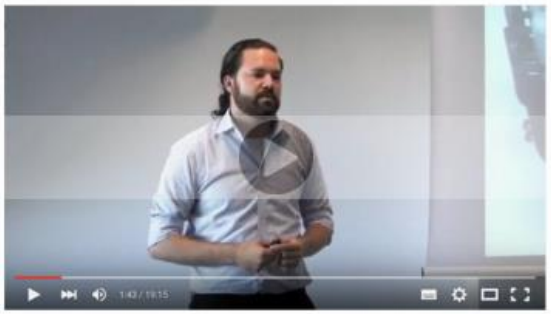


Schritt 2: Optional, falls viel gefunden: Erkennung von inkonsistenten Bug-Fixes, die schon hereingeschlüpft sind.



# Wie sag ichs meinem Chef?

Mit Vorträgen über Softwarequalität sind wir regelmäßig auf Industriekonferenzen oder Kunden-internen Workshops vertreten.



## Impulsvorträge

Gerne kommen wir auch zu Ihnen ins Haus, beispielsweise für interne Konferenzen oder Workshops. Unsere Themen reichen von Qualitätsanalysen über Qualitätscontrolling bis hin zu Testcontrolling oder der Einführung von Reviews. Oder aber Sie schlagen uns ein Thema Ihrer Wahl vor.

**DAS ANGEBOT**

- 60-90 MIN VORTRAG
- SOFTWAREQUALITÄT ALS THEMA
- BEI IHNEN IM HAUS
- € NUR UNSERE ANREISEKOSTEN
- TERMIN NACH VEREINBARUNG

IMPULSVORTRAG ANFRAGEN

# Software Quality Blog

## Practical Guide to Code Clones (Part 1)

*Posted on* 07/16/2014 *by* Dr. Benjamin Hummel

One well known principle in software engineering states *don't repeat yourself*, also known as the DRY principle. A very obvious violation of DRY is the application of copy/paste to create duplicates of large portions of source code within the same code base. These duplicate pieces of code, also known as *code clones*, have been subject to lots of research in the last two decades. In this two-part post I want to summarize those parts of the current knowledge that I find most relevant to the practitioner, especially the impact of clones on software dev

## Practical Guide to Code Clones (Part 2)

*Posted on* 07/30/2014 *by* Dr. Benjamin Hummel

In the previous part we introduced the notion of code clones and discussed, whether and under which circumstances cloning in your code base can be a problem for development and maintenance. In this post, I will introduce ways and tools to deal with code clones in your code base. After reading this, you should be able to select and apply a detection tool to inspect the clones in your own code base.

https://www.cqse.eu/en/blog/practical-guide-to-code-clones-part1/

# Kontakt

Ich bin noch länger hier und freue mich auf Diskussionen!

Bei Interesse an einem Job in diesem Bereich bitte melden ☺

Dr. Elmar Juergens · juergens@cqse.eu · +49 179 6753863

@ElmarJuergens
www.cqse.eu/en/blog

CQSE GmbH
Lichtenbergstraße 8
85748 Garching bei München

CQSE
Continuous Quality in Software Engineering