

Of Changes and their History: Ideas for New IDEs

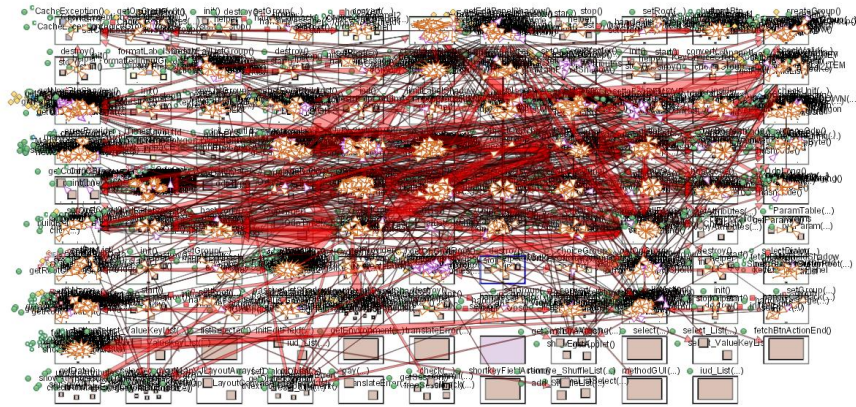
Harald Gall
University of Zurich
Department of Informatics
<http://seal.ifi.uzh.ch>



*"The **study of products** is vastly more important than the study of production, even for understanding production and its methods."*

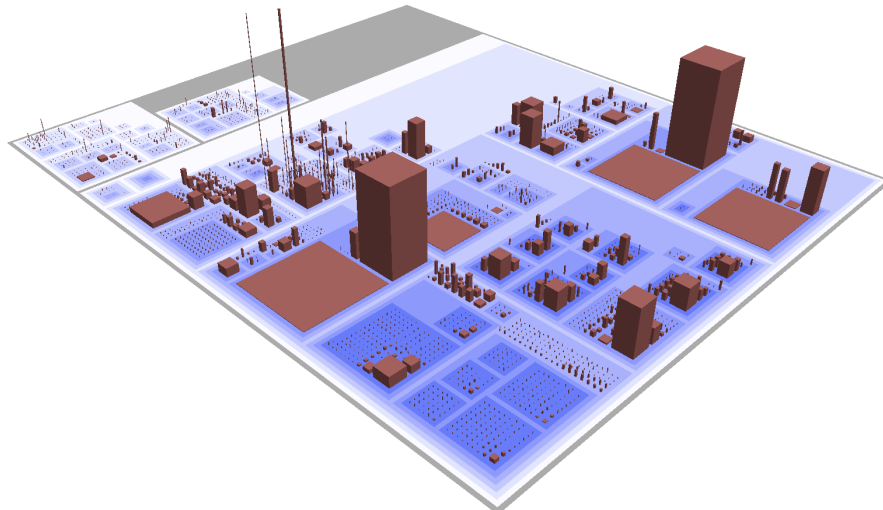
Karl Popper

On the study of products...



3

... for understanding production



Richard Wetzel, Michele Lanza. Visualizing Software Systems as Cities. In VISSOFT 2007

4

Goals & Questions

- What can an IDE learn about
 - Software and its **structure**
 - **Change** impact and propagation
 - **Developer** contributions and efforts
 - **Team** structure and social networks
 - Change **smells**, trends and hot spots
 - **Faults** and defects
-

5

Features of Modern IDEs



Visual Studio Team System 2010

- *“Focusing on an integrated set of tools that make it easy to **collect data** about your software development projects and **enable transparency**, promote **collaboration** and provide the information necessary to make the right **decisions** at the right time”*

source: <http://blogs.msdn.com/bharry/>

7

Some Rosario features

- Improved **work item tracking** capabilities that make it easier to track and understand work and perform impact analysis and other traceability functions.
- New **project scheduling** features that enable better coordination between project management and the rest of the organization.
- **Dashboards and cross project reports** that provide roll up of status across diverse projects.
- A simplified **data warehouse** and reporting experience that makes it easier to build custom reports.
- A TFS based **test case management** system that enables you to track all of your test cases, relate them to requirements and plans, track progress and perform impact analysis.

source: <http://blogs.msdn.com/bharry/>

8

Some cool scripting things you can do with Powershell and TFS

Get a list of all files that changed between two versions

```
Get-TfsItemHistory $/orcas -r -includeitems -version 575800~576111 | select-tfsitem | %{$_.path}
Get-TfsItemHistory $/orcas -r -includeitems -version LBeta1~LBeta2 | select-tfsitem | %{$_.path}
```

Show files in a given shelve set that I have local changes pending against

```
Get-TfsPendingChange -shelveset TfptReview | Get-TfsPendingChange
```

Show files that are locked in a source tree

```
Get-TfsPendingChange $/orcas -r -user * | where {$_.IsLock}
```

Delete shelve sets older than 90 days

```
Get-TfsShelveset | where { $_.CreationDate -le [DateTime]::Now.AddDays(-90) } | Remove-TfsShelveset
```

source: <http://blogs.msdn.com/bharry/>

9

IBM's Jazz

- *“Jazz is an IBM Rational project to build a scalable, **extensible team collaboration platform** for integrating work across the phases of the development lifecycle.”*

<http://jazz.net/>

10

IBM's Rational Team Concert

- **Integrated Work Item Management:** Automatically creates and tracks progress of individual work items in accordance with team process and project rules.
- **Integrated Source Control:** Provides essential software version control, workspace management and parallel development support to individuals and teams.
- **Integrated Build Management:** Schedule and execute software build processes.
- **Team Awareness & Collaboration:** Provides project-integrated presence and messaging

<http://www-01.ibm.com/software/awdtools/rtc/express/features/>

11

The latest and greatest in IDEs

work items

source control

build management

team awareness & collaboration

test management

data warehouse & reporting

project scheduling

12

But what about ...

- all the data already in the software archives?
 - all the release histories?
 - all the types of changes?
 - all the dependencies?
-

13

Mining Software Archives ...

- Code base
 - Which entities co-evolve/co-change?
 - Do code and comments co-evolve?
- Bugs and Changes
 - Who should fix this bug?
 - How long will it take to fix this bug?
 - Predicting bugs from cached bug history
 - When do changes induce fixes?
- Project and Process
 - Project memory for software development
- Software Expertise
 - Identifying expertise from changes and bug reports

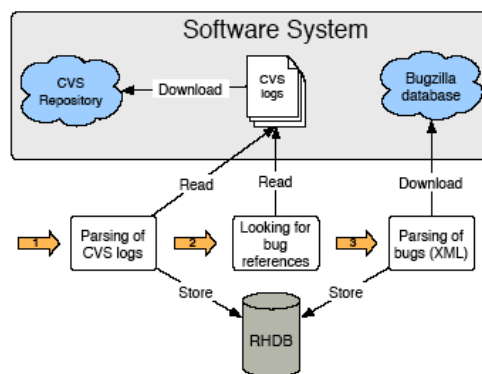


14

Change Analyses & Visualizations



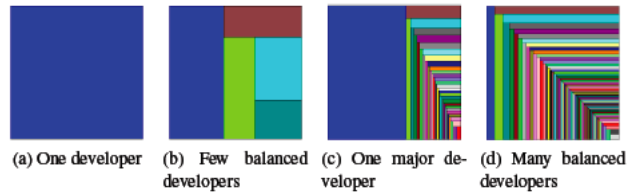
Release History Database



Research tools

- Hipikat, Cubranic et al.
- softChange, German
- Kenyon, Bevan et al.
- Mondrian, Girba et al.
- Evolizer, Gall et al.

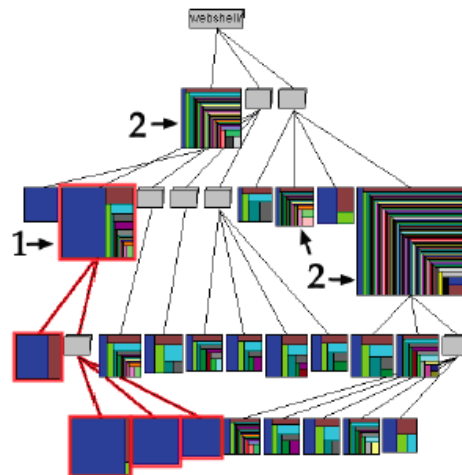
Developer contributions



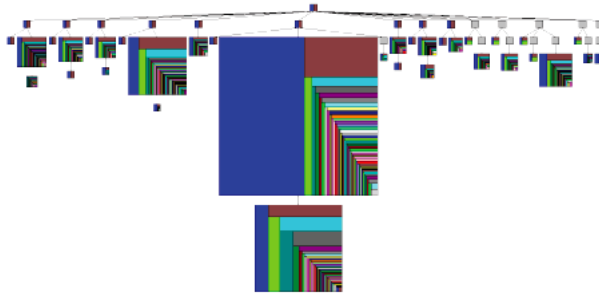
$$\text{Fractal Value} = 1 - \sum_{a_i \in A} \left(\frac{nc(a_i)}{NC} \right)^2, \quad \text{with } NC = \sum_{a_i \in A} nc(a_i)$$

Marco D'Ambros, Michele Lanza and Harald C. Gall, Fractal Figures: Visualizing Development Effort for CVS Entities In Proceedings of International Workshop on Visualizing Software For Understanding and Analysis, 2005.

How many developers per entity?



How many bugs per entity?



19

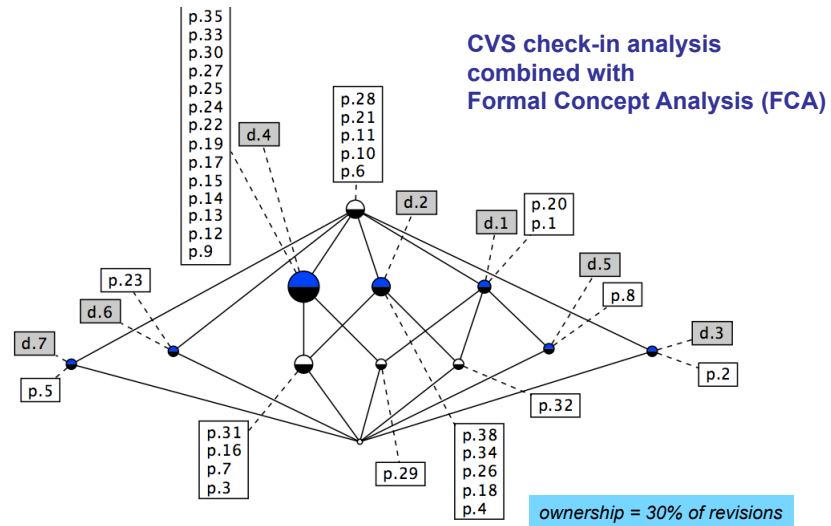
Who should fix this bug?

- Apply machine learning algorithms to open bug repository
- Learn the kinds of reports that each developer resolves
- A classifier suggests developers who should resolve the actual bug
- Precision: 57% in Eclipse, 75% in Firefox

Anvik, J., Hiew, L., and Murphy, G. C. 2006. *Who should fix this bug?*
In Proceeding of the 28th international Conference on Software Engineering, May 20 - 28, 2006.

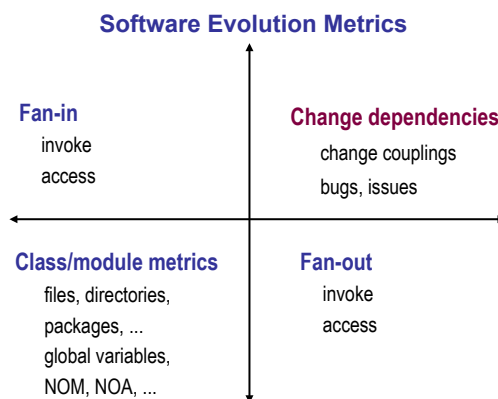
20

Who is the code owner?



21

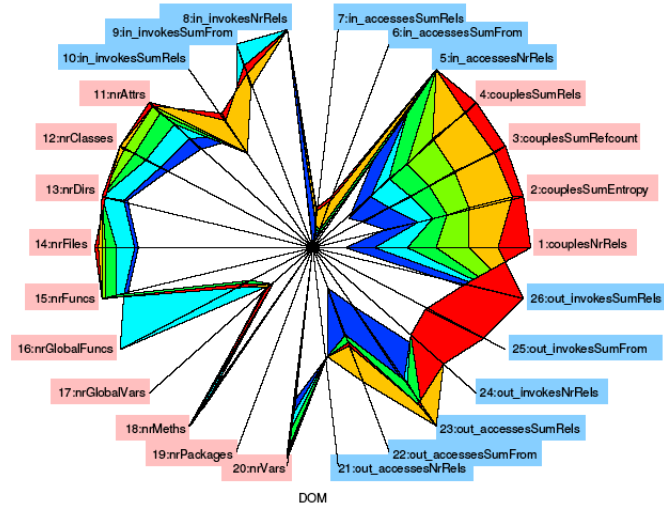
Which entities co-evolve?



Martin Pinzger, Harald C. Gall, Michael Fischer, and Michele Lanza, *Visualizing Multiple Evolution Metrics* In Proceedings of the ACM Symposium on Software Visualization, 2005.

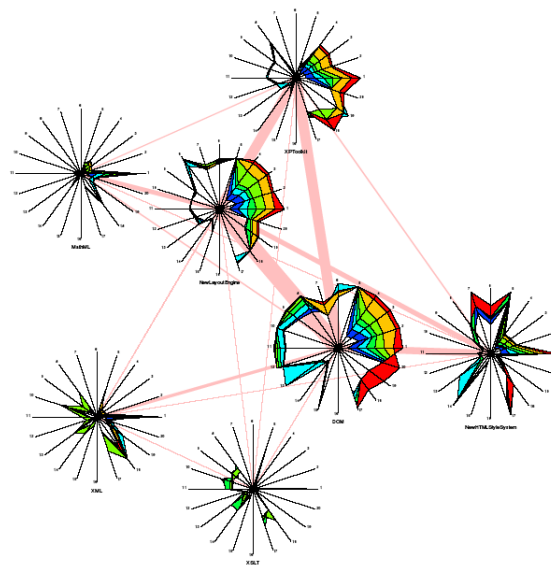
22

Mozilla Module DOM: 0.92 -> 1.7



23

Multiple Evolution Metrics



Kiviati graph:
26 metrics
7 Mozilla modules
7 subsequent releases

24

Change Analysis

CVS diff

```

1967,1970c1964,1965
<   if (d != null) {
<       d.foo();
<       d.bar();
<   }
---
>   d.foo();
>   d.bar();

```

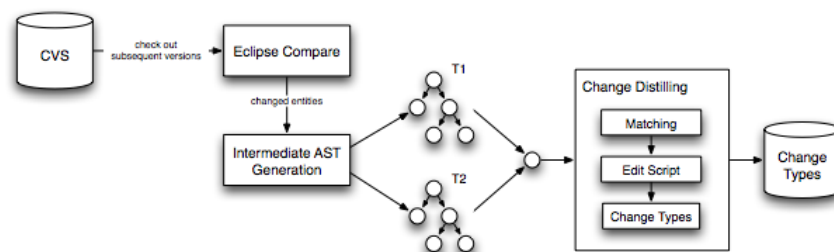
- 3 Body changes
- 2 Statement parent changes
- 1 Statement delete
- Change significance?

CVS log: "lines: +2 -4"

25

Change Distilling

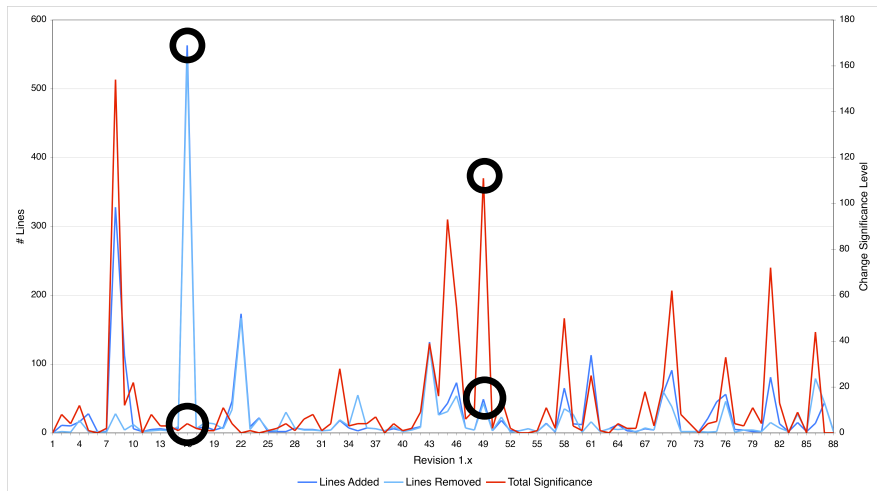
- Identifying change types and change patterns
- Eliminate mass changes and other noise



Beat Fluri, Michael Würsch, Martin Pinzger, Harald C. Gall, Change Distilling: Tree Differencing for Fine-Grained Source Code Change Extraction, IEEE Transactions in Software Engineering (TSE), November 2007

26

Which changes are significant?



Example from ArgoUML

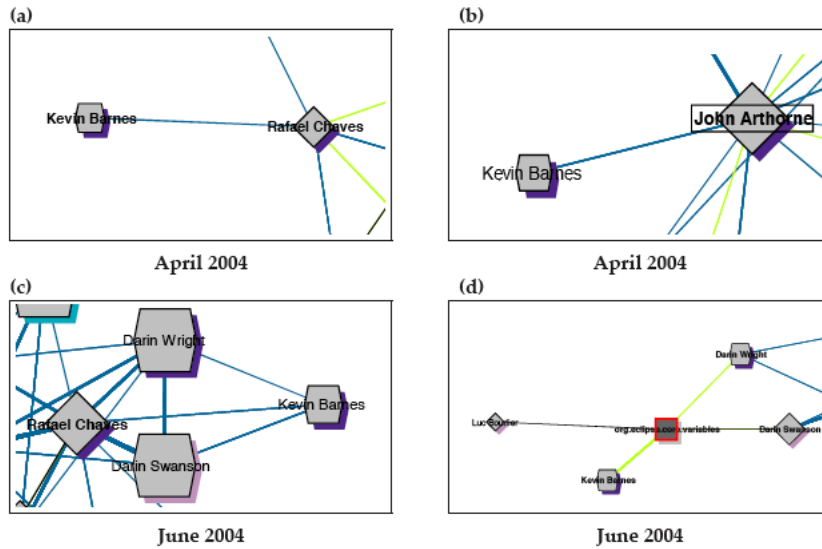
27

What are the developer networks?

- Conway's law
- Inter-team collaboration
- Ownership changes
- Key personalities in social networks
 - *connectors vs. communicators*
 - *gatekeepers, influencers, innovators, leaders and communicators as trendsetters*
- Information for project manager vs. newcomer

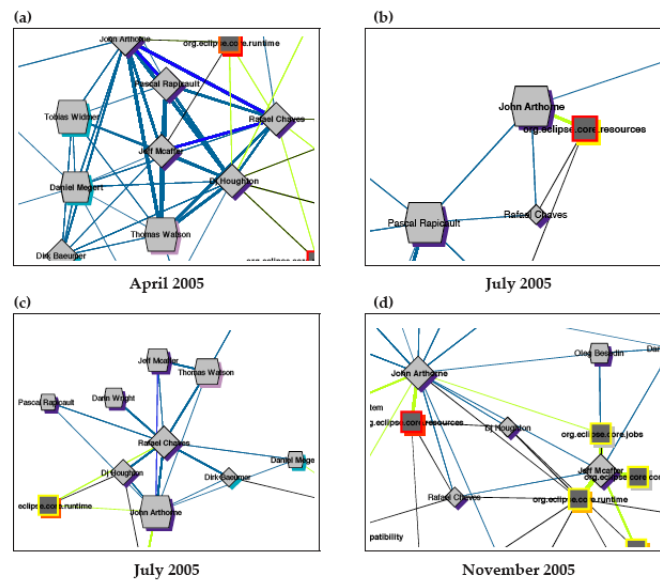
28

Scenario: newcomer Kevin



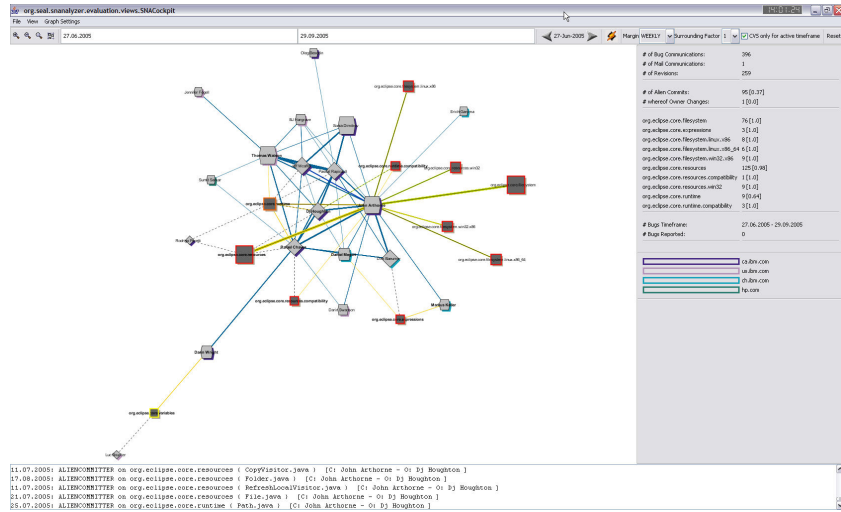
29

Scenario: key person Rafael leaving



30

SNA Cockpit

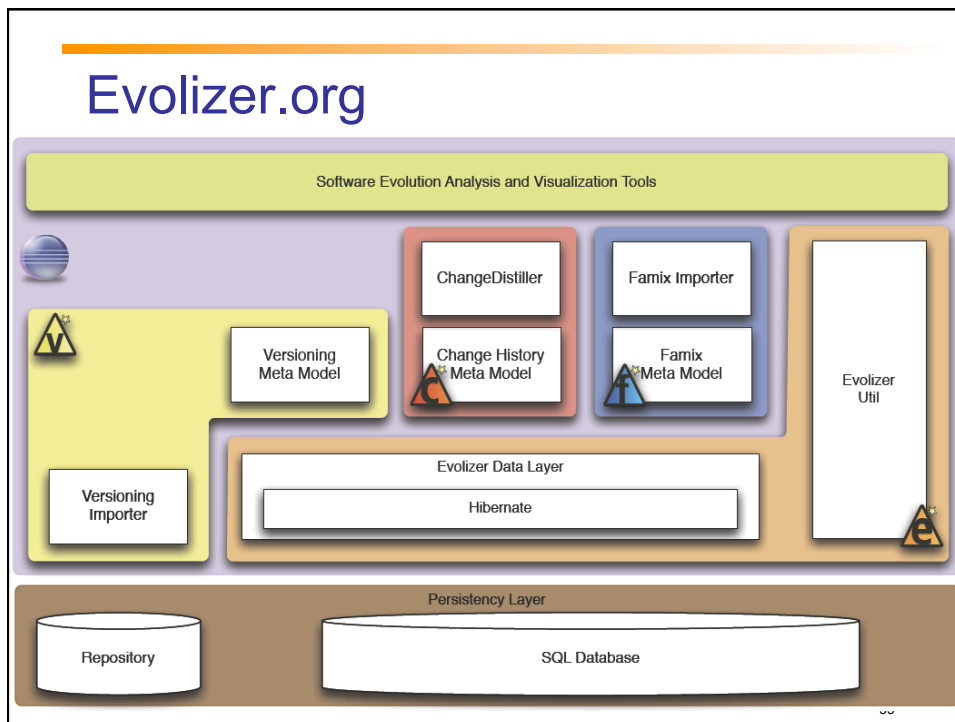


31

Evolizer

Our platform for harvesting and provisioning of software evolution data





Evolizer Tools

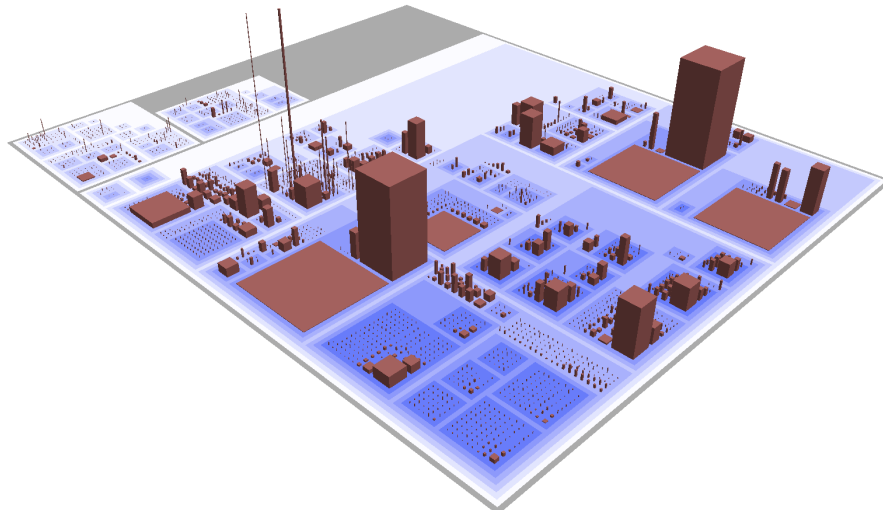
- **ChangeDistiller:** change types and significance
- **ArchView:** evolution metrics
- **SNA Cockpit:** developer networks
- **CocoViz**
- **ChangeCommander:** code and comments
- **Code Clone Evolution**

Ingredients for modern IDEs



35

Metaphors: Code City



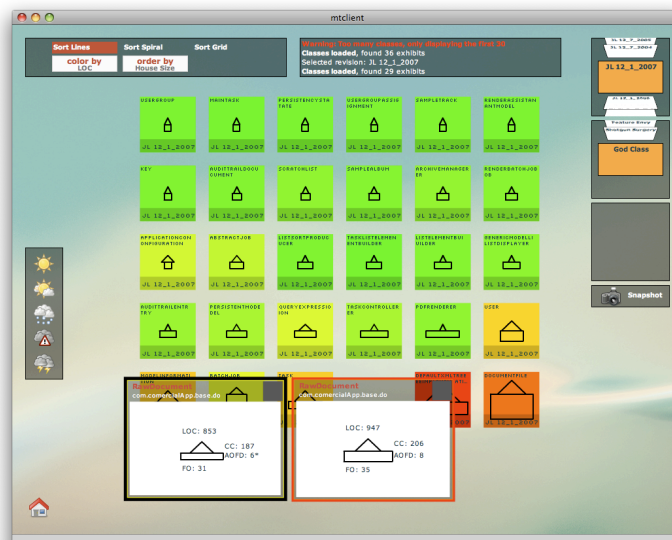
Richard Wettel, Michele Lanza. Visualizing Software Systems as Cities. In VISSOFT 2007

36

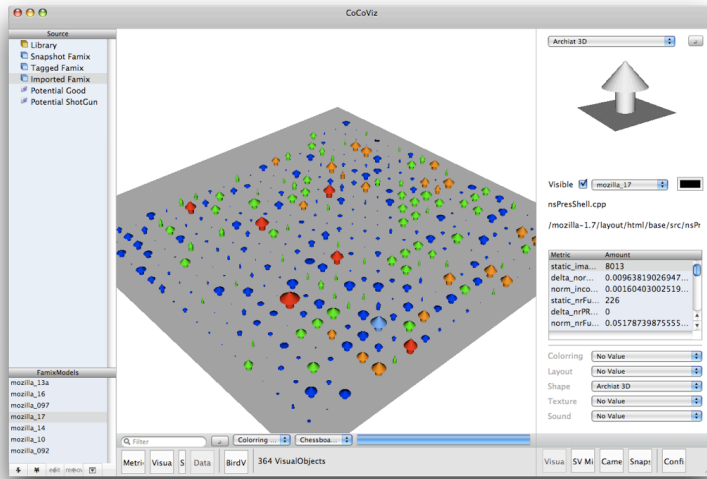
Buildings of ArgoUML



Visual Interfaces



Audio-Visual Exploration



39

Conclusions

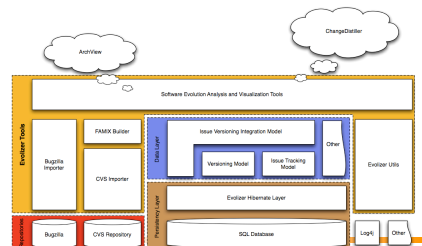
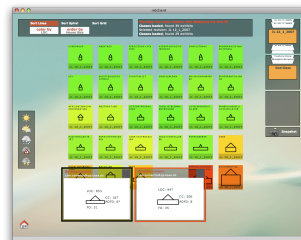
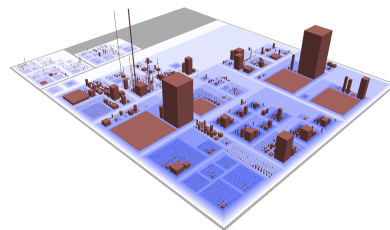
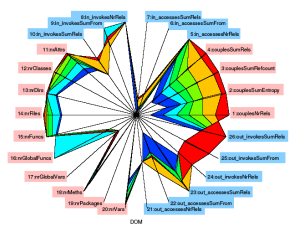


IDEAs: Integrated Development, Evolution and Analysis

- actively include change history and evolution analysis techniques
- have “tabular” project data and graphical views from the history
- provide role-based evolution for the user
- Data warehouses inside IDEs point into that direction...

41

“More of that Jazz”



42