

# **Semantic Clustering**

Adrian Kuhn

Software Composition Group  
University of Berne, Switzerland

# Topic Location

Source code

Identifier names

Latent Semantic Indexing

Hierarchical Clustering

Distribution Map

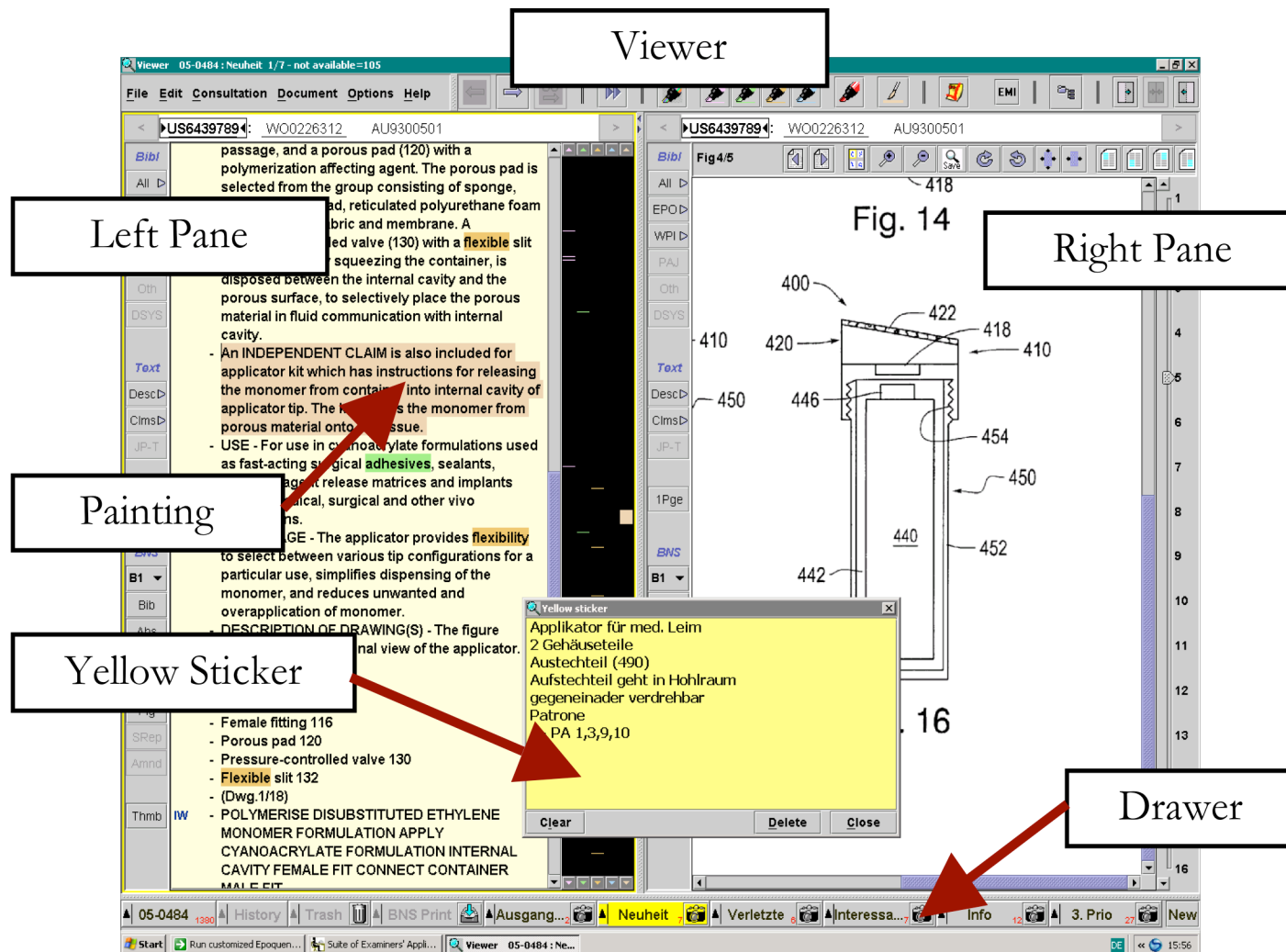
Searching

# Assumptions

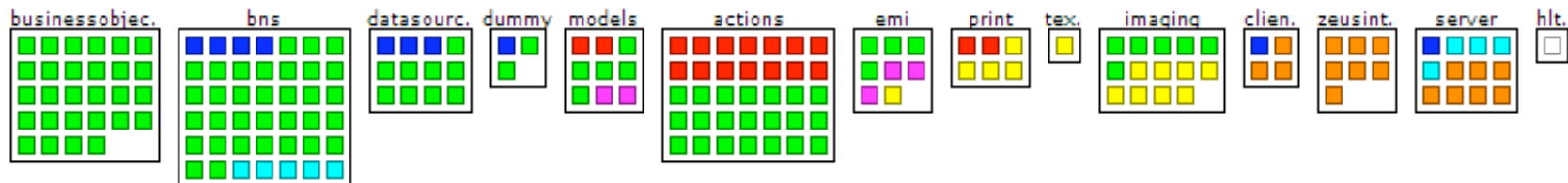
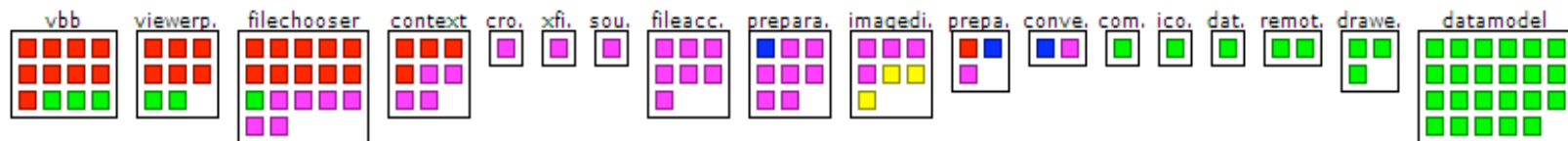
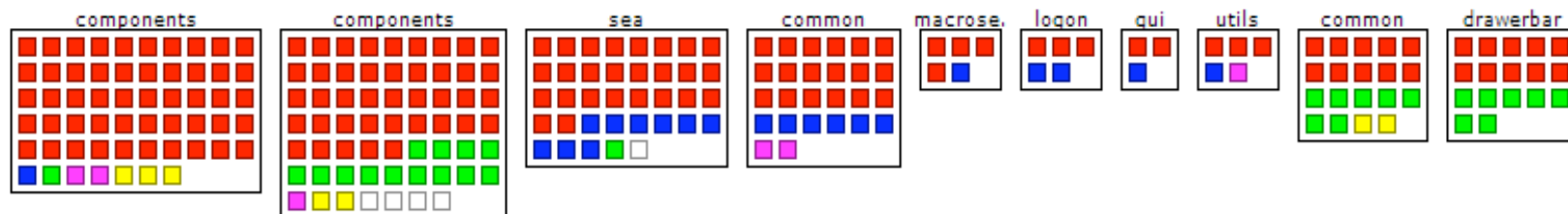
- Developers use meaningful names.
- Classes that use similar identifier names belong to the same topic.

**Applied for real...**








# Customer asked me to add a "Save as XML" menu item for them.

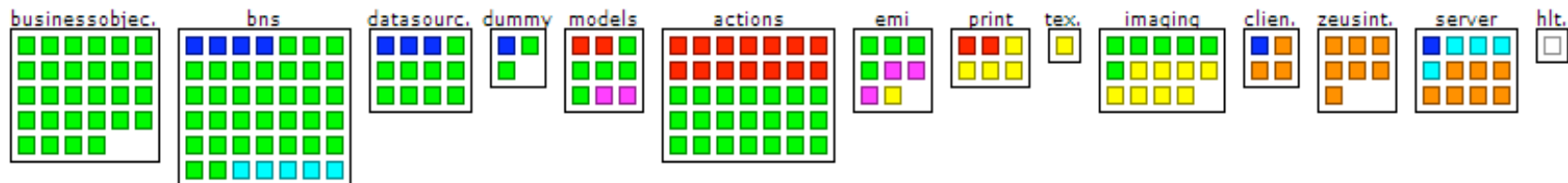
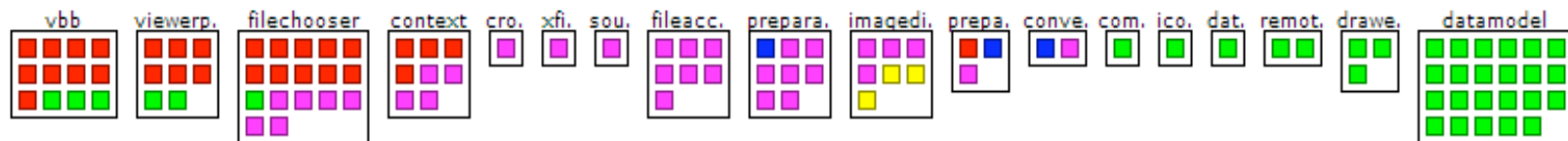
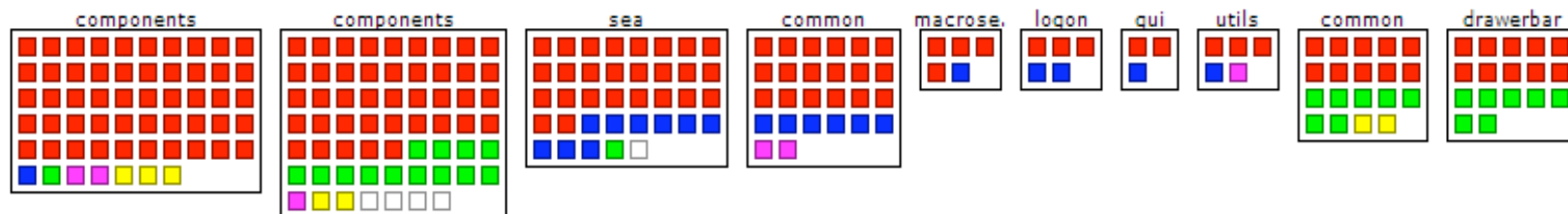
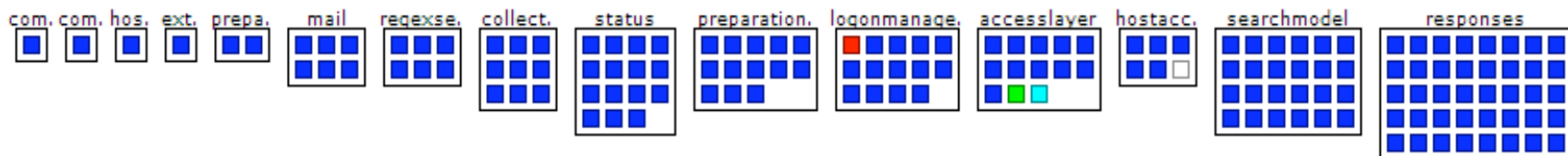


**Where should I start?**



# The System's Topics

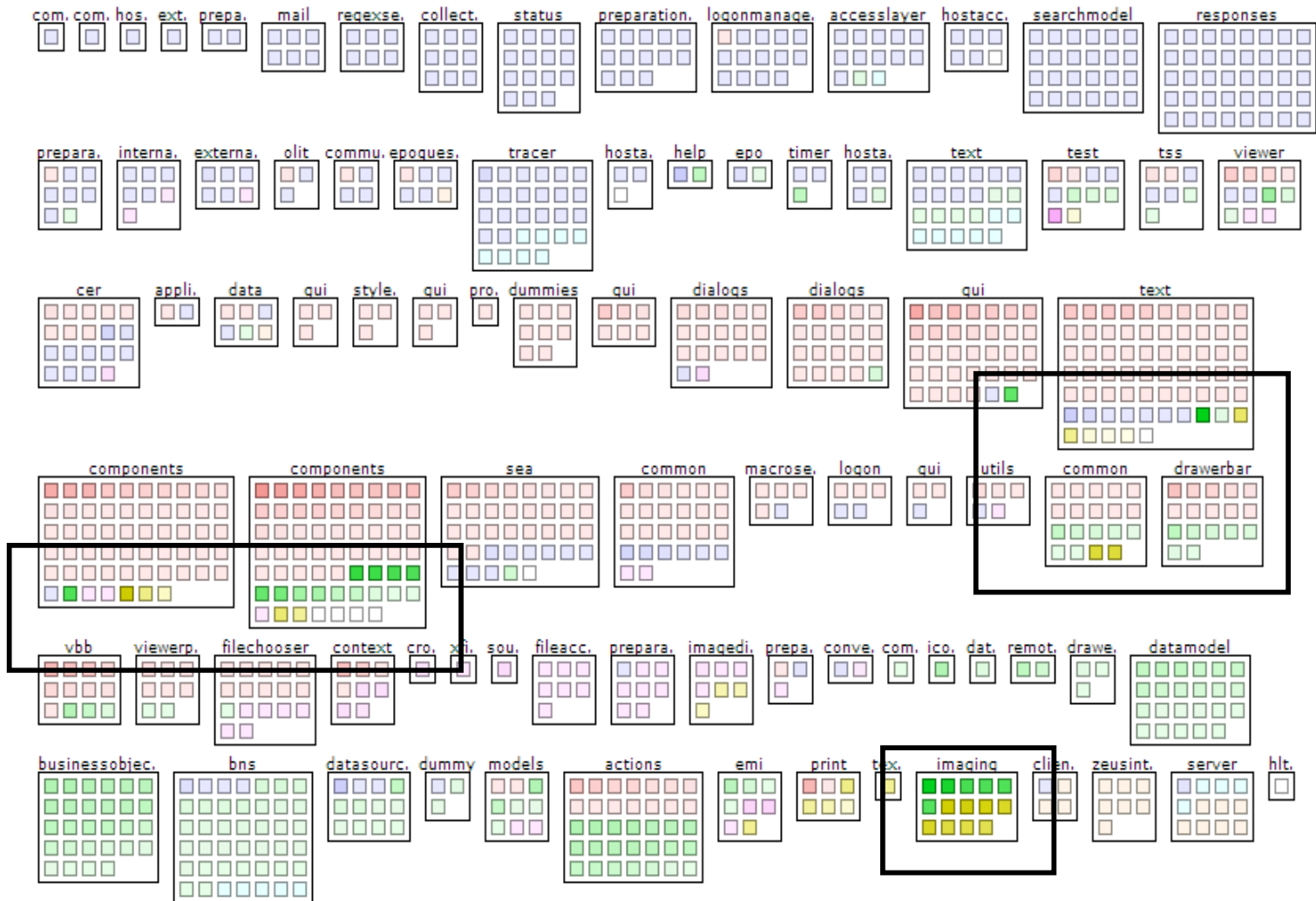
-  Red, GUI layer
-  Green, business model
-  Blue, database layer
-  Cyan, RMI remote access
-  Orange, server access
-  Magenta, file access
-  Yellow, image processing



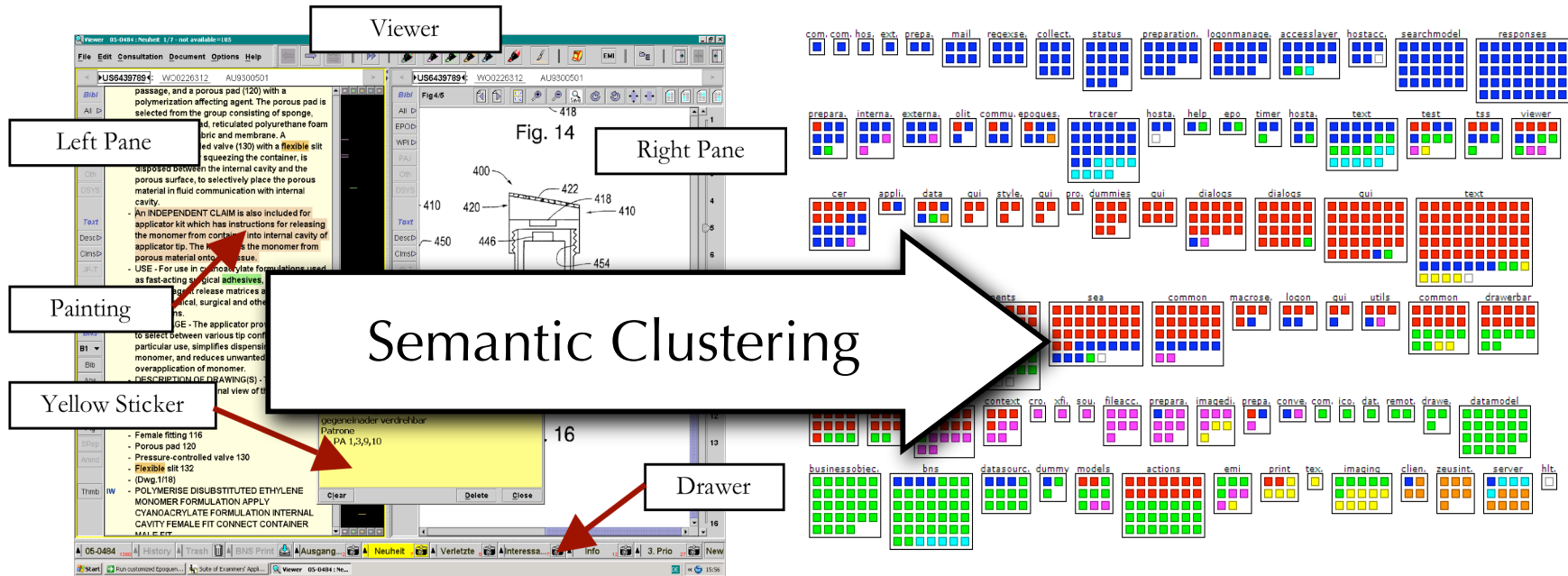
# Search for "drawer"



# “save cropped image”



# Topic Location



Questions? Comments? Feedback?