

Mining Software Repositories for Understanding

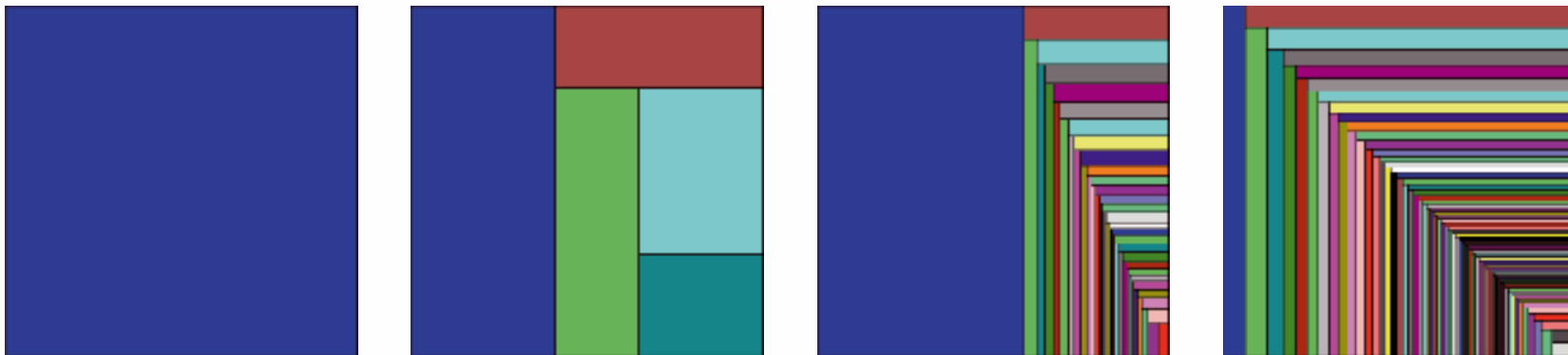
Software Evolution – L6T2

Dr. Vadim Zaytsev aka @grammarware, March 2021



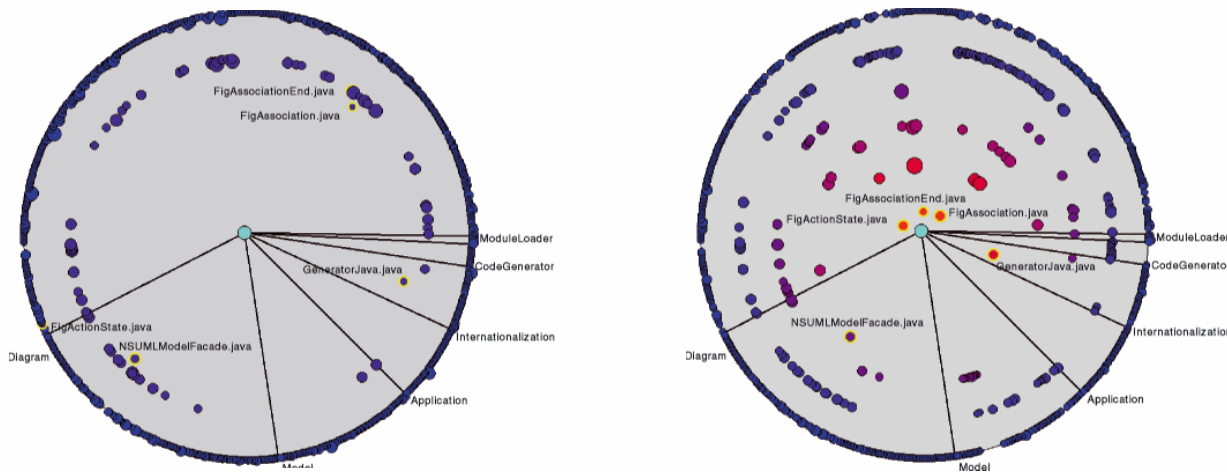
Developer Effort

- Questions:
 - how hard was/is it to develop feature X?
 - what is the maintenance rate of the project?
 - what is the distribution of effort in the team?



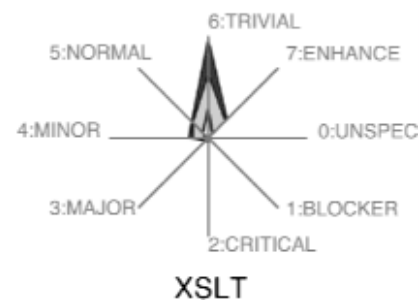
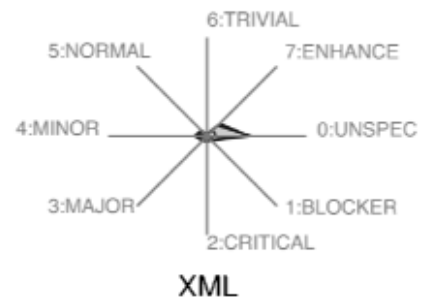
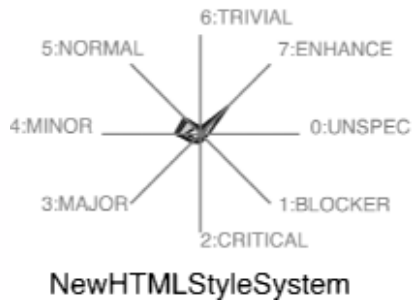
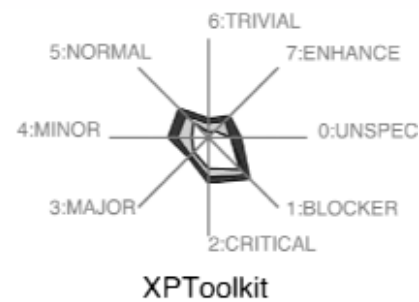
Change Coupling

- Questions:
 - which components change together often?
 - what are the modules with the strongest coupling?
 - which files are responsible for the coupling?



Hotspots

- Traditional hotspot: an outlier component
- Evolutional hotspot: an evolutionary outlier



Unstructured: Code

- Code organisation
 - modules, packages, dependencies
- Comments
 - linguistic data, internal documentation, emotions
 - `// TODO`
- Names
 - linguistic data, conventions
- String constants

Unstructured: Bug Databases

- aka Issue Trackers
 - title, description, comments
 - can reference commits & branches
- Explicit link between a problem and a solution
 - with timestamps & documentation
- Occasionally
 - contain heated discussions

Unstructured: Archives

- Mailing lists
- Chat logs
- Requirements documents
- Design documents
 - – all require NLP
 - tokenisation
 - topic discovery
 - concept matching

Conclusion

- Resource relocation
- Reengineering aide
- Occasionally fundamental results
- Q&A Sessions @ Canvas
 - ⇒ v.zaytsev@utwente.nl
 - ⇒ <https://discord.gg/n7VQAPNBPD>