Grip on software: understanding development progress of SCRUM sprints and backlogs

Helwerda, L.S.

Citation
Helwerda, L. S. (2024, September 13). *Grip on software: understanding development progress of SCRUM sprints and backlogs*. *SIKS Dissertation Series*. Retrieved from https://hdl.handle.net/1887/4092508

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/4092508

Note: To cite this publication please use the final published version (if applicable).

Grip on Software: Understanding development progress of SCRUM sprints and backlogs

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op vrijdag 13 september 2024
klokke 11:30 uur
door

Leon Sebastiaan Helwerda

geboren te Voorburg, Nederland
in 1992

Promotores:
Prof.dr.ir. F.J. Verbeek
Dr. W.A. Kosters

Promotiecommissie:
Prof.dr. M.M. Bonsangue
Prof.dr. H.C.M. Kleijn
Prof.dr. S. Manegold
Prof.dr. M.R.V. Chaudron (Eindhoven University of Technology)
Dr. C. Soomlek (Khon Kaen University)
Dr. F. Niessink (Stichting ICTU)

Copyright © 2024 Leon Helwerda
Cover art by Marian Helwerda, based on photograph by Leon Helwerda
Printed by: NBD Biblion

Funded by Stichting ICTU as part of a collaboration between Leiden University and ICTU

SIKS Dissertation Series No. 2024-28
The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

Contents

1 Introduction **1**
- 1.1 Preface . . . 3
- 1.2 Context . . . 3
  - 1.2.1 Software development, Agile and SCRUM . . . 4
  - 1.2.2 Machine learning, pattern recognition and predictive analytics . . . 7
  - 1.2.3 Case studies of workflows . . . 9
- 1.3 Design scope . . . 11
- 1.4 Problem statement . . . 14
- 1.5 Pipeline components . . . 17
  - 1.5.1 Instances . . . 18
  - 1.5.2 Non-functional requirements . . . 19
- 1.6 Structure of this thesis . . . 20

2 Data pipeline **23**
- 2.1 Introduction . . . 25
  - 2.1.1 Ecosystem . . . 26
  - 2.1.2 Structure . . . 27
- 2.2 Design . . . 27
  - 2.2.1 Distributed data systems . . . 27
  - 2.2.2 Agent-based communication . . . 28
  - 2.2.3 Organizational approaches . . . 28
- 2.3 Method . . . 29
  - 2.3.1 Data acquisition . . . 30
  - 2.3.2 Further pipeline steps . . . 33
- 2.4 Technical considerations . . . 36
  - 2.4.1 Generalizability . . . 36
  - 2.4.2 Continuous integration . . . 37
  - 2.4.3 Documentation . . . 37
  - 2.4.4 Novelty . . . 38
- 2.5 Results . . . 39
- 2.6 Discussion . . . 40

3 Database construction **41**
- 3.1 Introduction . . . 43
- 3.2 Relevant work . . . 45

3.3 Method . . . 45
3.3.1 Data model . . . 47
3.3.2 Linking data sources . . . 57
3.4 Architecture . . . 59
3.5 Experiments . . . 63
3.5.1 Setup . . . 63
3.5.2 Results . . . 64
3.6 Discussion . . . 66

4 Pattern recognition methods **69**
4.1 Proposition . . . 71
4.2 Background . . . 72
4.2.1 Framework . . . 73
4.2.2 Story points and adaptations . . . 74
4.3 Related work . . . 75
4.4 Approach . . . 76
4.4.1 Feature extraction . . . 77
4.4.2 Data set . . . 79
4.4.3 Models . . . 80
4.5 Analysis strategy . . . 83
4.6 Results . . . 84
4.6.1 Sprint classification and estimation . . . 84
4.6.2 Backlog size estimation . . . 85
4.7 Conclusions . . . 89
4.7.1 Threats to validity . . . 90
4.7.2 Proposed additions . . . 91

5 Information visualization **93**
5.1 Preamble . . . 95
5.2 Purpose . . . 96
5.3 Relevant concepts . . . 97
5.4 Dashboard framework . . . 99
5.5 Visualizations for analytical decision support . . . 102
5.5.1 Sprint report . . . 102
5.5.2 Prediction results . . . 111
5.5.3 Timeline . . . 114
5.5.4 Leaderboard . . . 118
5.6 Visualizations for ecosystem management . . . 121
5.6.1 Collaboration graph . . . 121
5.6.2 Process flow . . . 125
5.6.3 Heat map . . . 128
5.6.4 Platform status . . . 131
5.7 Novel backlog visualizations . . . 134
5.7.1 Product backlog burndown chart . . . 134
5.7.2 Product backlog progression chart . . . 135
5.7.3 Product backlog relationship chart . . . 137

5.8 Evaluation . . . 138
5.8.1 Assessment . . . 138
5.8.2 Adoption . . . 141
5.8.3 Conclusion . . . 142

6 Discussion **145**
6.1 Retrospective . . . 147
6.1.1 Technical overview . . . 147
6.1.2 Main contributions . . . 149
6.2 Overall conclusion . . . 151
6.2.1 Problem statement . . . 151
6.2.2 Research questions . . . 152
6.3 Future work . . . 159
6.3.1 Further research . . . 159
6.3.2 Generalizability . . . 161

Glossary **163**

Bibliography **169**

Appendices **183**

A Code repositories of the Grip on Software pipeline **185**

B Queries used in database performance experiments **189**

Summary **203**

Samenvatting **207**

Curriculum Vitae **211**

Acknowledgments **213**

SIKS Dissertation Series **215**