

Support Scrub Meetings in Distributed Teams by Detecting Duplicates of Software Defect Reports in Issue Management Systems

Maximilian Flis, 02.09.2019, Master Thesis Kickoff

Chair of Software Engineering for Business Information Systems (sebis)
Faculty of Informatics
Technische Universität München
www.matthes.in.tum.de

Outline



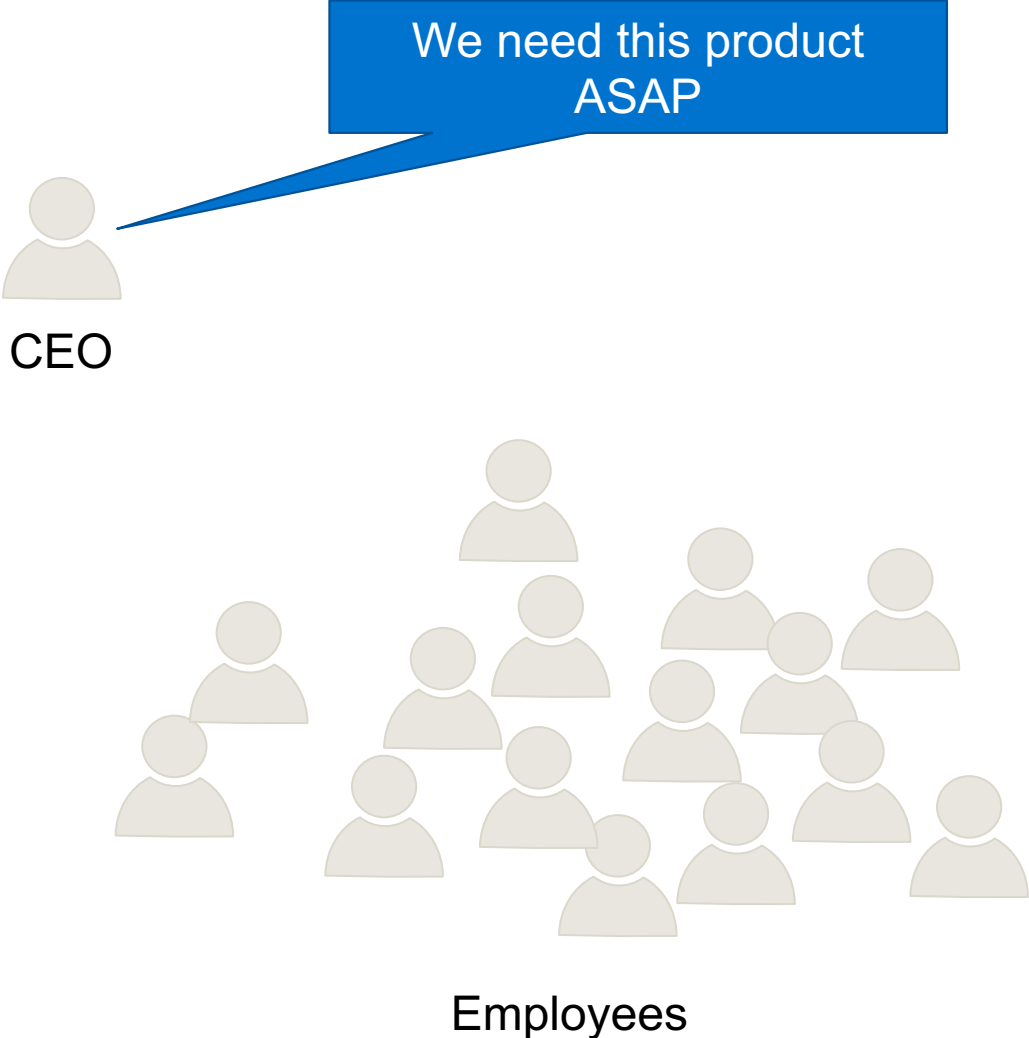
Motivation

Research Focus

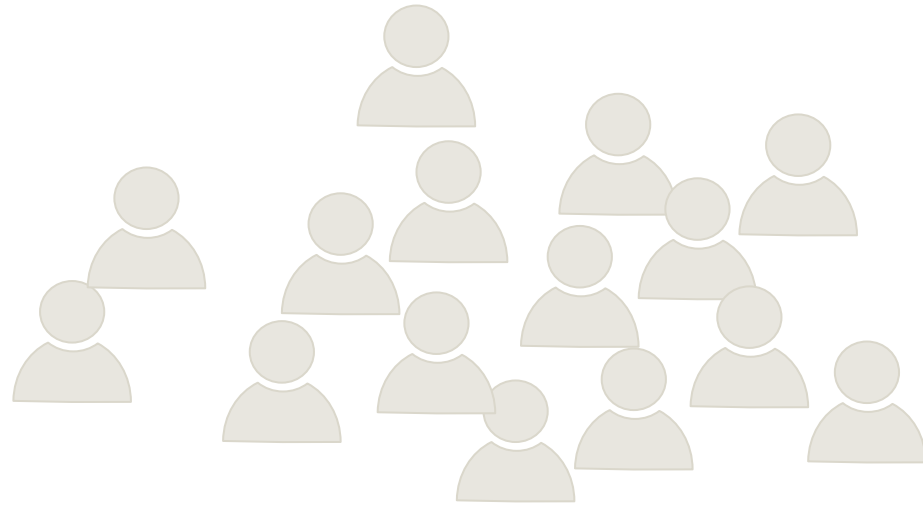
Approach

Current Status

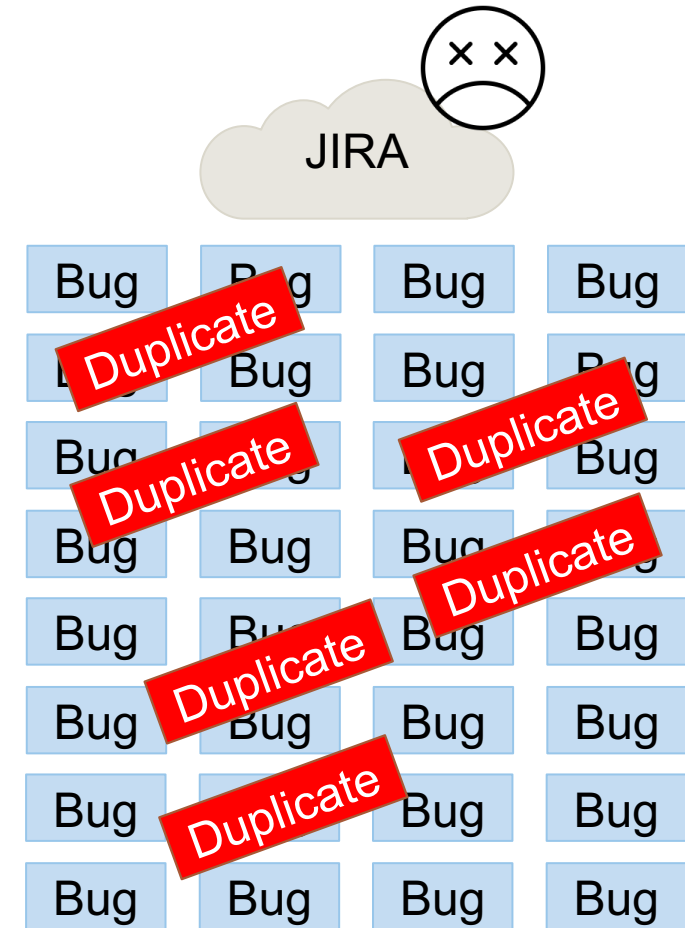
Timeline



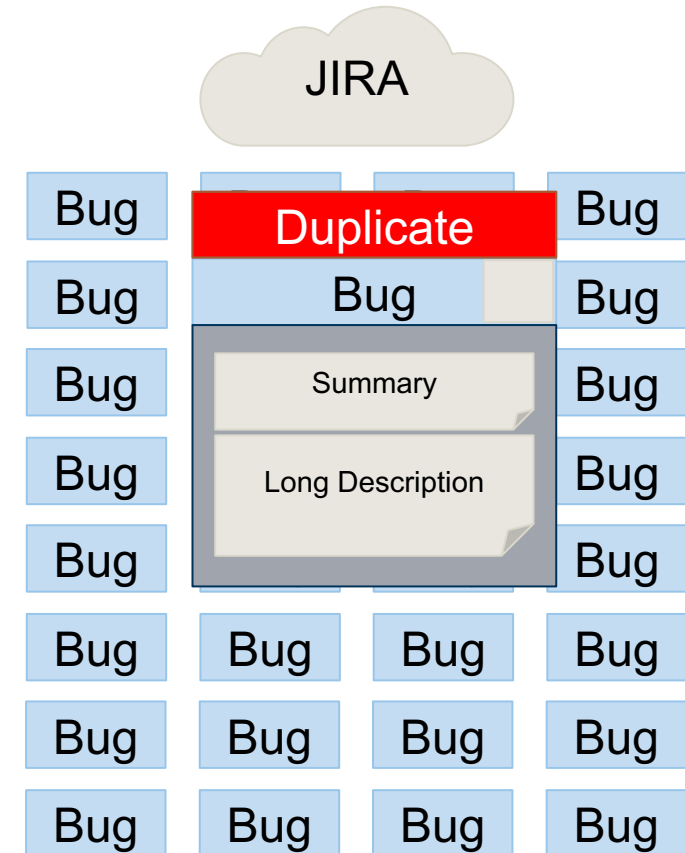
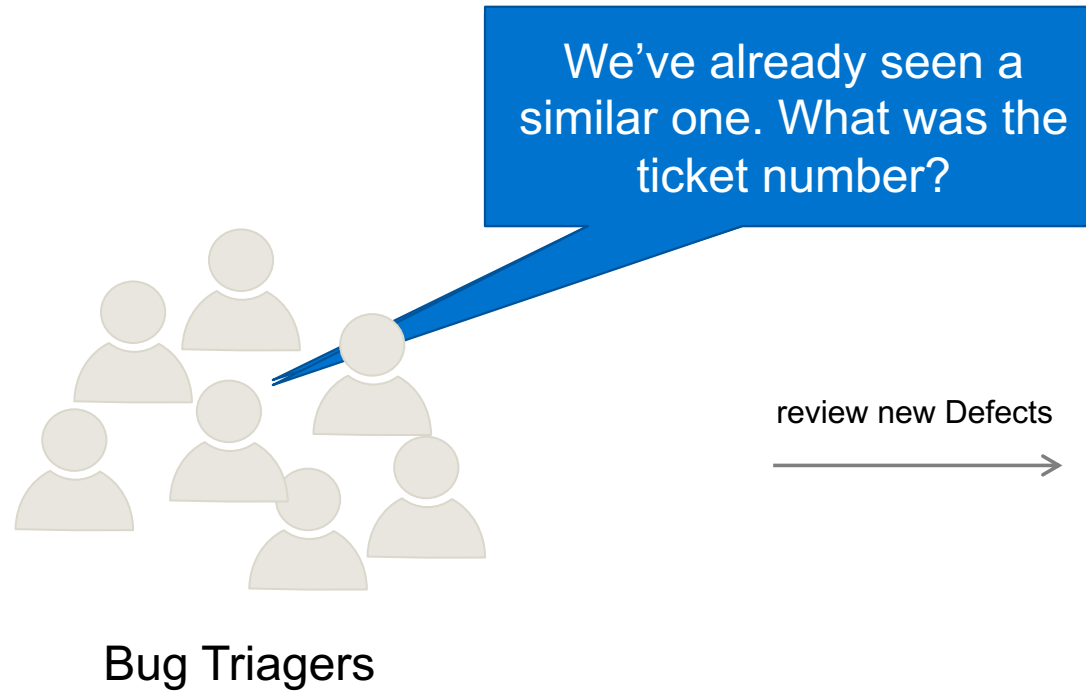




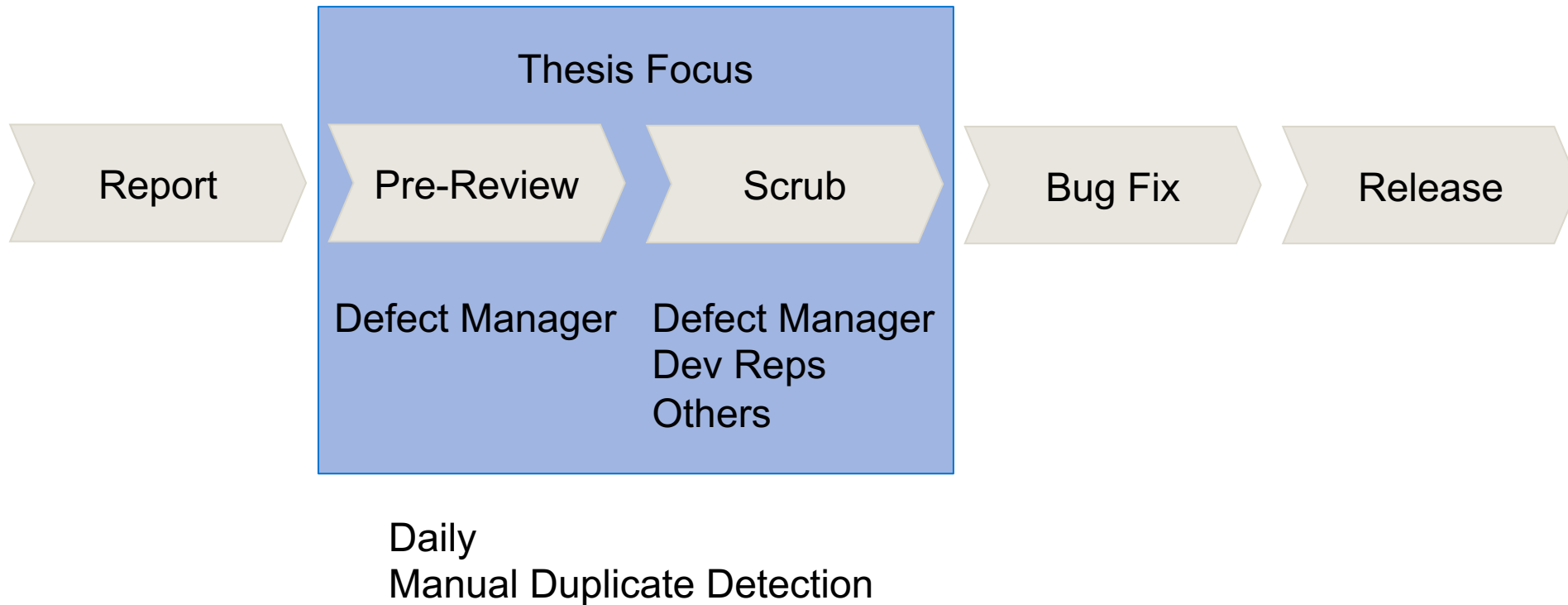
New Testers



Testers used “Duplicate Defect Report”
It’s super effective!



Defect Management today (at Sky)



RQ1	Do duplicate defect reports pose problems to developers in software engineering?	Survey
RQ2	What are the reasons for not incorporating tools for duplicate detection?	Interviews
RQ3	How does the detection tool help in finding duplicate defect reports before and during the scrub meeting?	Interviews
RQ4	What are the similarities and differences in the approach described in this thesis and existing tools to facilitate duplicate defect detection in bug triaging?	Compare to existing tools / Literature
RQ5	How generalizable is the duplicate detection tool to support bug triaging processes in other companies / projects?	Apply on open source project and compare

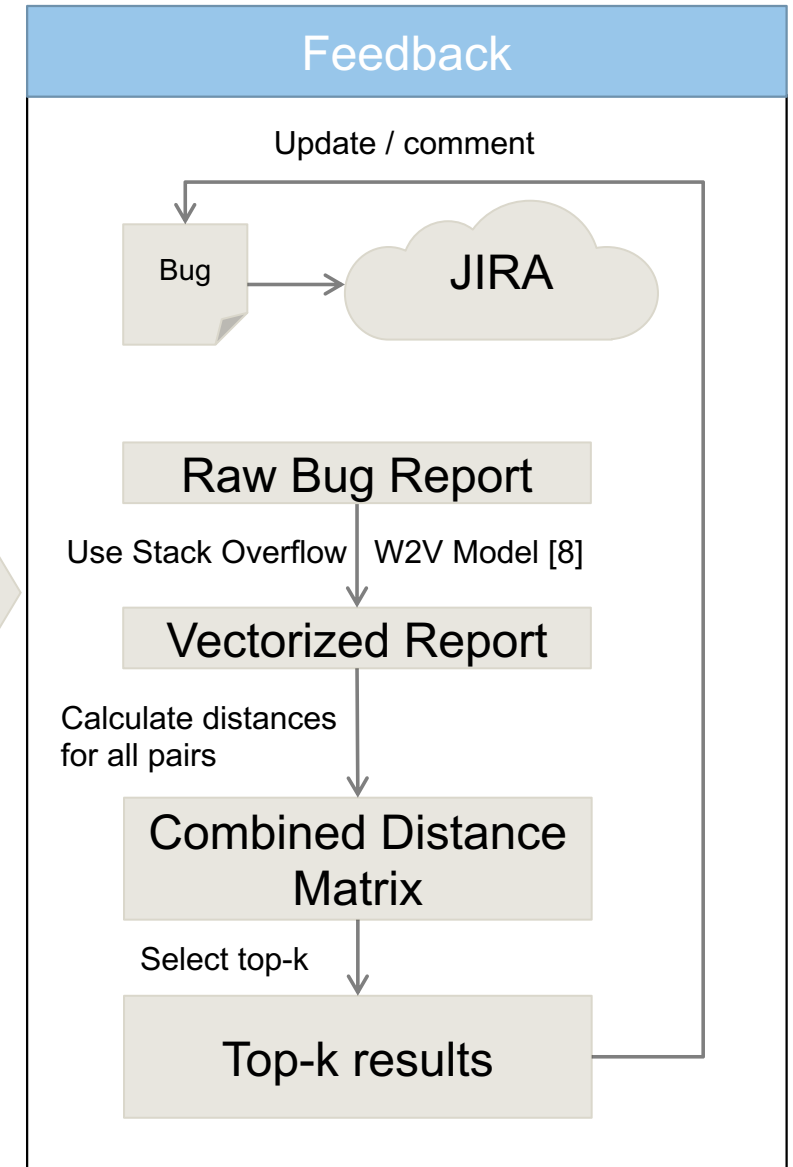
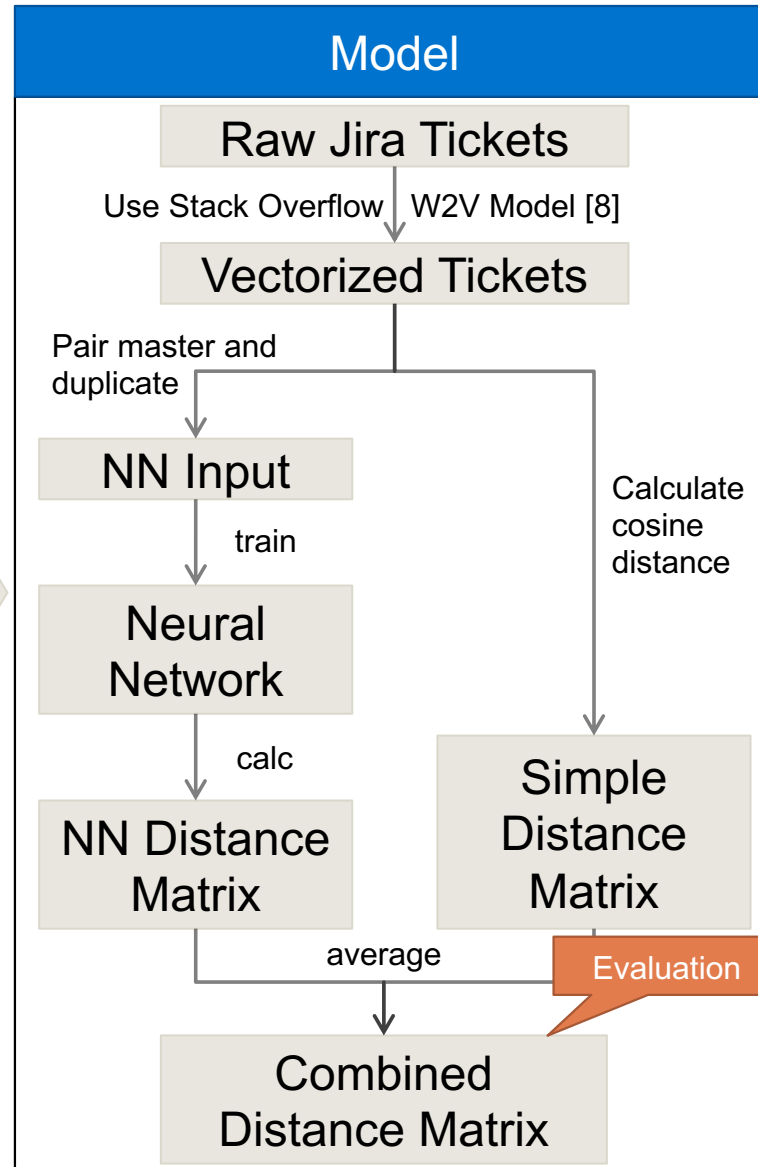
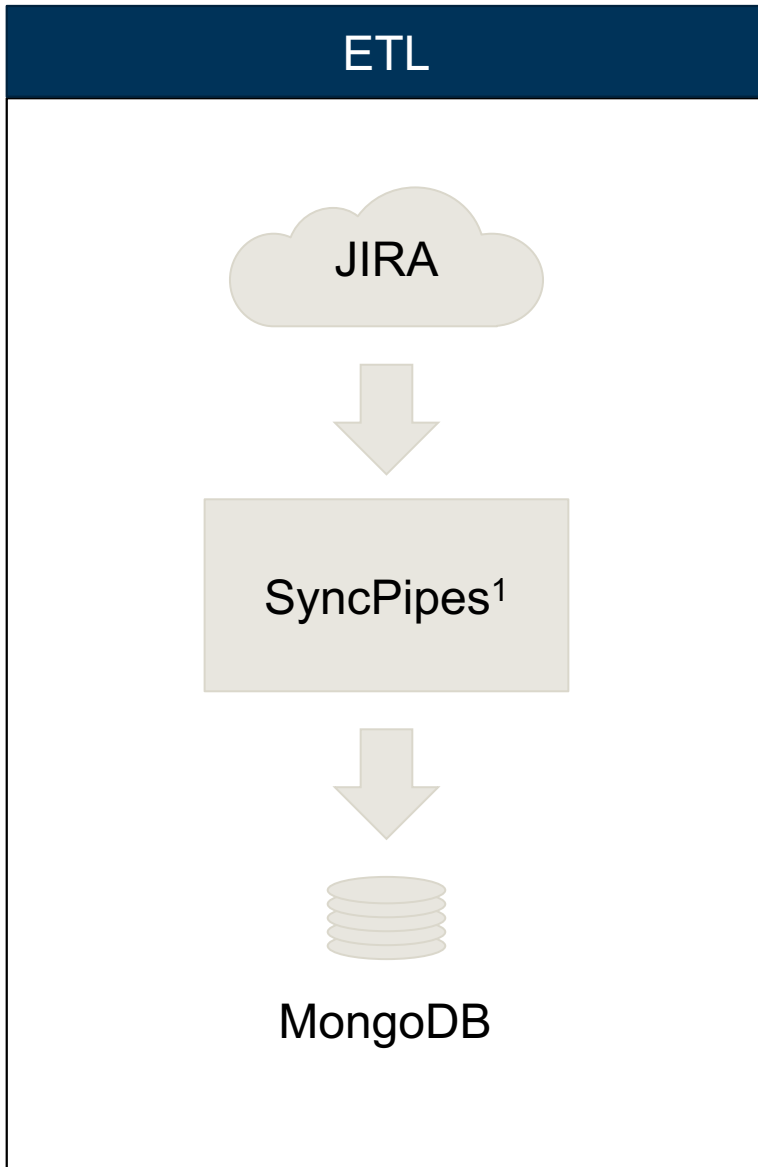
Approach

Techniques

- Information Retrieval [1]
- Stack Trace [5]
- Textual and categorical similarity [3] [4]
- Topic model [6]

History

- NLP [2]
- Categorical Features [7]
- Automatic Filter (rule based) [1]
- Top-k most similar [7]



1) SyncPipes (<https://wwwmatthes.in.tum.de/pages/2gh0u9d1afap/SyncPipes>)

Current Status

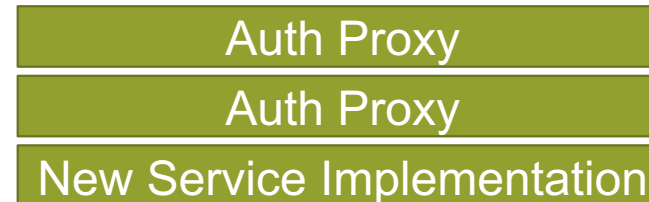
SyncPipes¹

Facts

- Easy kick start into ETL
- Simple pipeline configurations
- ETL ist hard!

Hurdles

- Passwords stored as is without hashing
- No client certificate support
- Incremental updates missing (Sync)



1) SyncPipes (<https://wwwmatthes.in.tum.de/pages/2gh0u9d1afap/SyncPipes>)

Current Status

Defects

- Defect management process over tool restrictions
- Each bug considered to be unique in the first place (applications, contexts, releases, etc.)
- Defect reports are cluttered with non-natural language (noise) [1]

Existing duplicate detection tools

- Paid plugins
- Researchers' implementations
- No built-in solution

Current Status

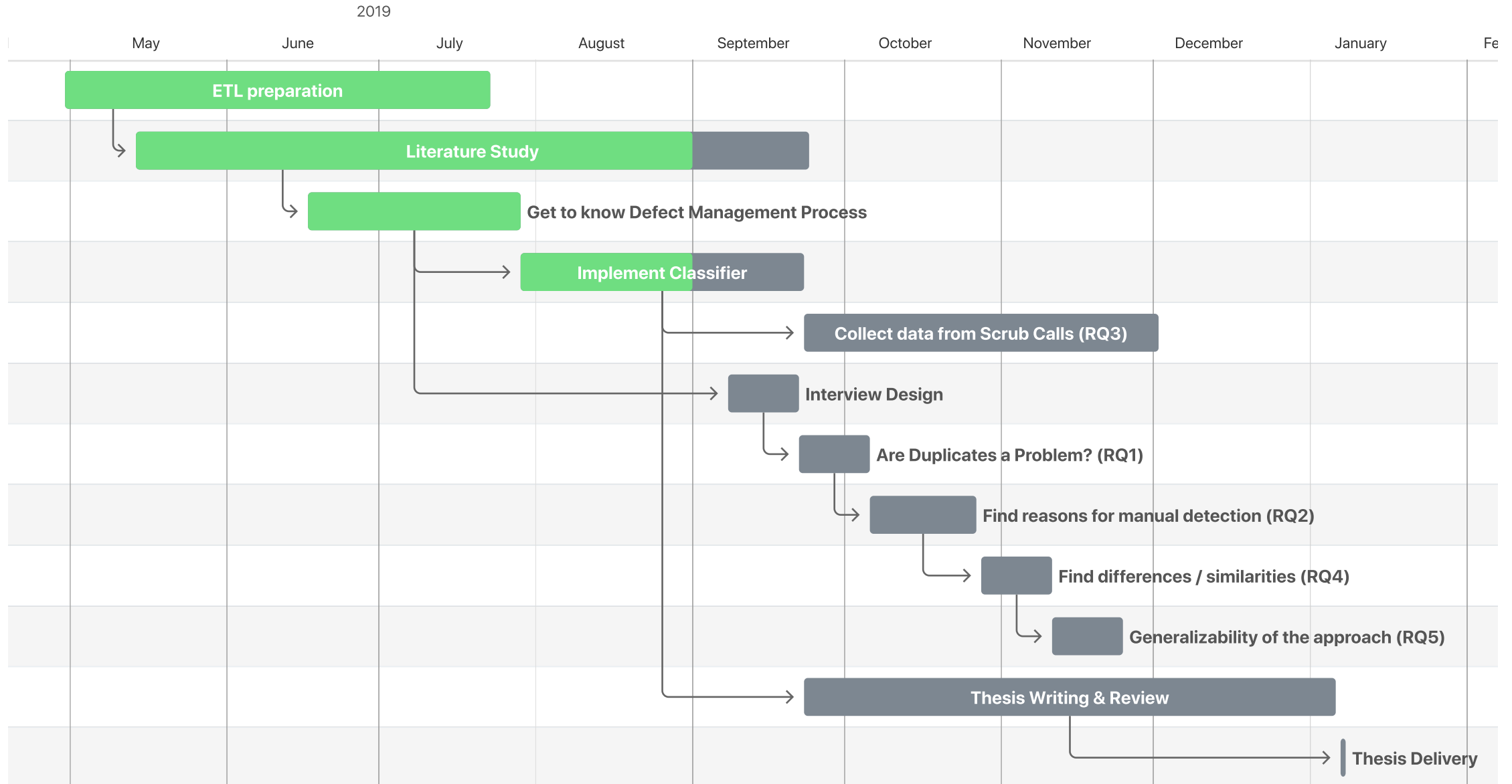
Jira Project

- Multiple sources of input (distributed teams)
- 1000+ Updates daily (includes new issues)
- 50+ new issues daily
- Only rely on duplicate links instead of resolved as “Duplicate” (clear connection between two tickets)

Scrub

- Daily
- Pre-Review by one Defect Manager (daily)
- 4+ Scrub participants (2+ person hours daily)

Timeline





Maximilian Flis

Technische Universität München
Faculty of Informatics
Chair of Software Engineering for Business
Information Systems

Boltzmannstraße 3
85748 Garching bei München

Tel +49.89.289. 17132
Fax +49.89.289.17136

maximilian.flis@tum.de
www.matthes.in.tum.de



Literature

- [1]: Aminoroaya, Z., Neysiani, B. S., & Shahraki, M. H. N. (2018). Detecting Duplicate Bug Reports Techniques. *Research Journal of Applied Sciences*, 13, 522–531.
- [2]: Lazar, A., Ritchey, S., & Sharif, B. (2014, May). Improving the accuracy of duplicate bug report detection using textual similarity measures. In *Proceedings of the 11th Working Conference on Mining Software Repositories* (pp. 308-311). ACM.
- [3]: Jalbert, N., & Weimer, W. (2008, June). Automated duplicate detection for bug tracking systems. In *2008 IEEE International Conference on Dependable Systems and Networks With FTCS and DCC (DSN)* (pp. 52-61). IEEE.
- [4]: Runeson, P., Alexandersson, M., & Nyholm, O. (2007, May). Detection of duplicate defect reports using natural language processing. In *Proceedings of the 29th international conference on Software Engineering* (pp. 499-510). IEEE Computer Society.

Literature

- [5]: Banerjee, S., Cukic, B., & Adjeroh, D. (2012, October). Automated duplicate bug report classification using subsequence matching. In *2012 IEEE 14th International Symposium on High-Assurance Systems Engineering* (pp. 74-81). IEEE.
- [6]: Nguyen, A. T., Nguyen, T. T., Nguyen, T. N., Lo, D., & Sun, C. (2012, September). Duplicate bug report detection with a combination of information retrieval and topic modeling. In *Proceedings of the 27th IEEE/ACM International Conference on Automated Software Engineering* (pp. 70-79). ACM.
- [7]: Sun, C., Lo, D., Khoo, S. C., & Jiang, J. (2011, November). Towards more accurate retrieval of duplicate bug reports. In *Proceedings of the 2011 26th IEEE/ACM International Conference on Automated Software Engineering* (pp. 253-262). IEEE Computer Society.
- [8]: Efstathiou, V., Chatzilenas, C., & Spinellis, D. (2018, May). Word embeddings for the software engineering domain. In *Proceedings of the 15th International Conference on Mining Software Repositories* (pp. 38-41). ACM.

Graphics / Images

- (1): [Alexandre](#), FR: Dead Smiley. Creative Commons CCBY. <https://thenounproject.com/term/dead-smiley/134580/>