DEPARTMENT OF INFORMATICS

TECHNICAL UNIVERSITY OF MUNICH

Master's Thesis in Robotics, Cognition, Intelligence

Detecting Vague Requirements with Machine Learning

Leo Hanisch

DEPARTMENT OF INFORMATICS

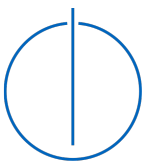
TECHNICAL UNIVERSITY OF MUNICH

Master's Thesis in Robotics, Cognition, Intelligence

Detecting Vague Requirements with Machine Learning

Detektion von vagen Anforderungen mit maschinellem Lernen

Author:	Leo Hanisch
Supervisor:	Prof. Dr. Dr. h.c. Manfred Broy
Advisor:	Dr. rer. nat. Sebastian Eder
Submission Date:	November 15, 2020



I confirm that this master's thesis in robotics, cognition, intelligence is my own work and I have documented all sources and material used.

Munich, November 15, 2020 Leo Hanisch

Acknowledgments

Abstract

Contents

1 Test

1.1 Section

Citation test [Dev+18].

1.1.1 Subsection

See Table 1.1, Figure 1.1, Figure 1.2, Figure 1.3.

Table 1.1: An example for a simple table.

A	B	C	D
1	2	1	2
2	3	2	3

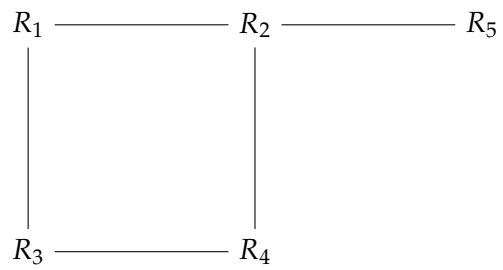


Figure 1.1: An example for a simple drawing.

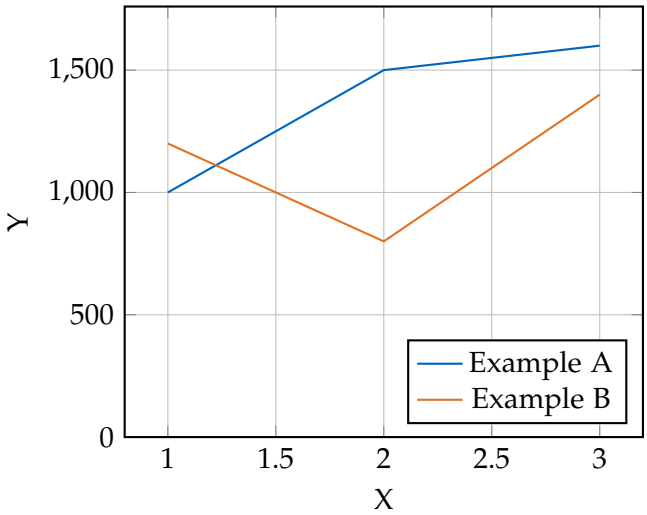


Figure 1.2: An example for a simple plot.

```
SELECT * FROM tbl WHERE tbl.str = "str"
```

Figure 1.3: An example for a source code listing.

2 Introduction

A software product is only as good as its development process [HDK93]. This process involves specifying and understanding requirements correctly as an integral part. Research has shown that this is prone to faults which can cost additional time and money [Fer+16] and may lead to severe project delay [Fem+14]. It is therefore desirable to avoid those drawbacks by recognizing misleading requirements at an early stage and, in the best case, by specifying them clearly before the next development step. To resolve those one must rephrase the requirements in an unambiguous way. However, this is non trivial since often domain knowledge is required to uncover and resolve the issues [Fem+17]. For example in a requirement containing *vague pronouns* like "The software must include a service *which* must be accessible via a user interface." it is unclear whether *which* relates to *The software* or *service*. Another example for a requirement defect are *loopholes*. A requirement stating that the software should be tested *as far as possible* leaves the reader room for interpretation and thus is ambiguous. The previously presented defects called *vague pronouns* and *loopholes* are examples for so called *Requirement Smells* defined by Femmer et al. [Fem+17]. If a requirement smell is fulfilled it indicates that a requirement is of insufficient quality. Consequently, requirement smells can be used to argue about the a requirement's quality

Before one can reason about a requirement's quality, one must check whether requirement smells apply or not. An approach to accomplish this are manual review. According to Salger [Sal13], reviews have several drawbacks. The review must be carried out by the relevant stakeholders and they must fully understand each requirement. Consequently, the reviewer needs domain knowledge in order to perform the reviews which makes the review more difficult to execute. Furthermore, the result of a review depends on the reviewer him/herself [BR83]. Moreover, the reviewer can be distracted by the earlier mentioned requirement smells. Therefore, Femmer et al. [Fem+17] conclude that reviews are costly and time consuming.

Knowing reviews are costly regarding time and money, tooling which supports the requirements engineering would be beneficial for the quality assurance process. Such tooling could support the reviewer by automatically indicating requirement smells and therefore speed up reviews. Further, in reality not only dedicated requirement smells are of interest, but more general speaking ambiguous or vague requirements. Such an assistant tool should be capable to process natural language since requirements are

mostly formulated in natural language [MFN04] and indicate whether a requirement is vague or not.

Machine Learning (ML) achieved in recent history remarkable results for natural language processing (NLP) tasks [Kha+16]. An example to mention is Google's neural network (NN) model called BERT which showcased how NNs can improve performance in transfer learning tremendously [Dev+18]. This recent success shows that NNs have great potential in transfer learning and consequently in detecting vague requirements.

The aim of this thesis is to further explore the potential of modern NNs in the context of detection of vague requirements. I want to contribute the following points.

List of Figures

List of Tables

Bibliography

[BR83] M. V. Z. R. Y. R. G. H. J. D. G. Basili and V. R. "The Software Indus- try: A State of the Art Survey." In: *Foundations of Empirical Software Engineering: The Legacy ofVictor R.Basili* (1983), p. 383.

[Dev+18] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers forLanguage Understanding." In: *CoRR* abs/1810.04805 (2018).

[Fem+14] H. Femmer, D. Méndez Fernández, E. Jürgens, M. Klose, I. Zimmer, and J. Zimmer. "Rapid Requirements Checks with Requirements Smells: Two Case Studies." In: 2014. DOI: 10.1145/2593812.2593817.

[Fem+17] H. Femmer, D. Méndez Fernández, S. Wagner, and S. Eder. "Rapid quality assurance with Requirements Smells." In: *Journal of Systems and Software* 123 (2017), pp. 190–213. ISSN: 01641212. DOI: 10.1016/j.jss.2016.02.047.

[Fer+16] D. M. Fernández, S. Wagner, M. Kalinowski, M. Felderer, P. Mafra, A. Vetrò, T. Conte, M. T. Christiansson, D. Greer, C. Lassenius, T. Männistö, M. Nayabi, M. Oivo, B. Penzenstadler, D. Pfahl, R. Prikladnicki, G. Ruhe, A. Schekelmann, S. Sen, R. Spinola, A. Tuzcu, J. L. de la Vara, and R. Wieringa. "Naming the pain in requirements engineering." In: *Empirical Software Engineering* 22.5 (2016), pp. 2298–2338. ISSN: 1382-3256 1573-7616. DOI: 10.1007/s10664-016-9451-7.

[HDK93] P. Hsia, A. M. Davis, and D. C. Kung. "Status report: requirements engineering." In: *IEEE Software* 10.6 (1993), pp. 75–79. ISSN: 0740-7459. DOI: 10.1109/52.241974.

[Kha+16] W. Khan, A. Daud, J. A. Nasir, and T. Amjad. "A survey on the state-of-the-art machine learning models in the context of NLP." In: *Kuwait Journal of Science* 43.4 (2016). Ef8is Times Cited:14 Cited References Count:69, pp. 95–113. ISSN: 2307-4108.

[MFN04] L. Mich, M. Franch, and P. L. Novi Inverardi. "Market research for requirements analysis using linguistic tools." In: *Requirements Engineering* 9 (2004), pp. 40–56. DOI: 10.1007/s00766-003-0179-8.

[Sal13] F. Salger. "Requirements reviews revisited: Residual challenges and open research questions." In: *2013 21st IEEE International Requirements Engineering Conference (RE)*. 2013, pp. 250–255. DOI: 10.1109/RE.2013.6636725.