

Sebastian Wittor

Automatische Erkennung und Messung von IT-Sicherheitsaufwänden

Bachelorarbeit

BEI GRIN MACHT SICH IHR WISSEN BEZAHLT



- Wir veröffentlichen Ihre Hausarbeit, Bachelor- und Masterarbeit
- Ihr eigenes eBook und Buch - weltweit in allen wichtigen Shops
- Verdienen Sie an jedem Verkauf

Jetzt bei www.GRIN.com hochladen
und kostenlos publizieren



Bibliografische Information der Deutschen Nationalbibliothek:

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de/> abrufbar.

Dieses Werk sowie alle darin enthaltenen einzelnen Beiträge und Abbildungen sind urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsschutz zugelassen ist, bedarf der vorherigen Zustimmung des Verlanges. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen, Auswertungen durch Datenbanken und für die Einspeicherung und Verarbeitung in elektronische Systeme. Alle Rechte, auch die des auszugsweisen Nachdrucks, der fotomechanischen Wiedergabe (einschließlich Mikrokopie) sowie der Auswertung durch Datenbanken oder ähnliche Einrichtungen, vorbehalten.

Impressum:

Copyright © 2016 GRIN Verlag
ISBN: 9783346152671

Dieses Buch bei GRIN:

<https://www.grin.com/document/538839>

Sebastian Wittor

Automatische Erkennung und Messung von IT-Sicherheitsaufwänden

GRIN - Your knowledge has value

Der GRIN Verlag publiziert seit 1998 wissenschaftliche Arbeiten von Studenten, Hochschullehrern und anderen Akademikern als eBook und gedrucktes Buch. Die Verlagswebsite www.grin.com ist die ideale Plattform zur Veröffentlichung von Hausarbeiten, Abschlussarbeiten, wissenschaftlichen Aufsätzen, Dissertationen und Fachbüchern.

Besuchen Sie uns im Internet:

<http://www.grin.com/>

<http://www.facebook.com/grincom>

http://www.twitter.com/grin_com

Automatische Erkennung und Messung von IT-Sicherheitsaufwänden

Bachelorthesis

Sebastian Wittor

Wirtschaftsinformatik

Sebastian Wittor

Studiengang: Wirtschaftsinformatik

Bachelorthesis: Automatische Erkennung und Messung von IT-Sicherheitsaufwänden

Eingereicht: 15.04.2016

Fachgebiet Wirtschaftsinformatik | Electronic Markets

Fachbereich Rechts- und Wirtschaftswissenschaften

Technische Universität Darmstadt

Abstrakt

IT security can be a decisive reason for the use of software. In this thesis, we developed a way for an automated classification of security issues in an issue tracking system to measure the impact of it security in open source project development. Based on issue tracking messages from SourceForge, we use patternbased refinement as a way to get a selection of bug tracker messages for a manual classification of security issues. We identify inhomogeneity as a main issue for machine learning. Bug tracker messages have different language and content quality and are additionally impured with spam and error messages and code segments. Thus, we used filter methods to clean them up. Based on another machine learning, we were able to classify these code segments and error messages. We optimized the results of our machine learning to classify security issue by stacking machine learning results to get an accuracy of 99.99%. The validation identifies two main problems with our machine learning method, new type of bug tracking messages and messages with just a few sentences. An statistical study based on the recognized security issues examined the number of relative security issues in a previous development phase as main influence of the number of security issues in the subsequently development phase. The influence of the software type and success of the software phase couldn't be confirmed

Inhaltsverzeichnis

Abbildungsverzeichnis	III
Tabellenverzeichnis	IV
Abkürzungsverzeichnis	V
1. Einleitung	1
1.1. Motivation	1
1.2. Ziele und Nutzen	1
1.3. Struktur der Arbeit	2
2. Grundlagen	3
2.1. Grundlagen zur IT-Sicherheit	3
2.1.1. Definition der IT-Sicherheit	3
2.1.2. Schwachstellen in der IT-Sicherheit	4
2.1.3. Schwachstellendatenbanken	5
2.2. Grundlagen zu Open Source Software und SourceForge	5
2.2.1. Definition der Open Source Software	6
2.2.2. SourceForge als Open Source Plattform	6
2.2.3. Bugtracker	7
2.2.4. Artefakt	8
2.3. Grundlagen des maschinellen Lernens	8
2.3.1. Einführung	8
2.3.2. Die Datenpräparation als Vorstufe des maschinellen Lernens	10
2.3.3. Algorithmen zur Textklassifizierung	12
2.3.4. Ensembles und Stacking	16
2.3.5. Evaluierung der Ergebnisse	17
3. Aktuelle Forschung	21
3.1. Übersicht der Herangehensweisen der aktuellen Forschung	21
3.2. Analytische Methoden zur Erkennung von IT-Schwachstellen	22
3.2.1. Mathematische Modelle	22
3.2.2. Automatische Methoden	23
3.3. Empirische Untersuchung von Aufwänden in der IT-Sicherheit	25
3.4. Diskussion der analytischen und empirischen Untersuchungen zur IT-Sicherheit	27

4. Patternbasierende Selektion und manuelle Klassifizierung von Artefakten	29
4.1. Datensammlung und Verarbeitung	29
4.2. Patternbasierende Selektion von Artefakten	30
4.3. Manuelle Klassifizierung von Sicherheitsschwachstellen	32
4.4. Analyse und Ergebnisse der manuellen Klassifizierung	34
4.5. Erfahrungen der manuellen Klassifizierung	37
4.6. Diskussion der Ergebnisse	38
5. Maschinelles Lernen	40
5.1. Problemstellung und Lösungsansätze	40
5.2. Datenpräparation	43
5.2.1. Erkennen von Codeartefakten	43
5.2.2. Generierung von Tokens	44
5.2.3. Metadaten Tokens	47
5.3. Durchführung und Ergebnisse des maschinellen Lernens	48
5.3.1. Wahl des optimalen k für k -NN	49
5.3.2. Veränderung der Accuracy durch Metadaten im maschinellen Lernen	52
5.3.3. Evaluierung der Tokenqualität	53
5.3.4. Evaluierung der Qualität der Algorithmen	54
5.4. Ensembles und Stacking	58
5.5. Validierung der Ergebnisse	60
5.5.1. Validierung auf den restlichen Artefakten der Datenbank	60
5.5.2. Validierung auf neuen Artefakten	62
5.6. Diskussion der Ergebnisse	63
6. Exemplarische statistische Untersuchung von ökonomischen Zusammenhängen	65
6.1. Problemstellung der statistischen Untersuchung	65
6.2. Forschungshypothesen und Modellerstellung	66
6.3. Statistische Auswertung des Modells	67
6.4. Einflussfaktoren von Security Bugs, Security Feature Request und Security Discussions	72
6.5. Diskussion der statistischen Ergebnisse	74
7. Zusammenfassung, Fazit und Ausblick	75
A. Anhang	VI
A.1. Regular Expressions des patternbased Refinement	VI
A.2. Ergebnisse des maschinellen Lernens mit und ohne Metadaten	IX
A.3. Regressionsergebnisse für Security Bugs, Feature Requests und Security Discussions	X

Abbildungsverzeichnis

1.	Beispiel einer Bugtracker-Nachricht in SourceForge	7
2.	Techniken des maschinellen Lernens	9
3.	Schritte zur Datenpräparation eines Dokuments	10
4.	Beispiel eines Decision Tree und seinen Knotentypen	13
5.	Grafische Darstellung der Funktionsweise des k -NN Algorithmus	15
6.	Beispiel der Ergebnisse einer zweidimensionalen Support Vektor Maschine	16
7.	Stacking und Ensemble Learning	17
8.	Verhältnis von positiven und negativen Klassifizierungen zueinander	18
9.	Beispiel eines ROC Graphen	19
10.	Evaluierung der MBL Ergebnisse	20
11.	Vorgehensweisen zur Messung von Aufwänden in der IT-Sicherheit	21
12.	Überblick über das Framework zur Datensammlung	30
13.	Funktionsweise der patternbasierenden Selektion	31
14.	Microsoft Access Formular der manuellen Klassifikation	32
15.	Tokenisierung der Codeartefakten	44
16.	RoC Ergebnisse für k -NN	50
17.	Ergebnisse des maschinellen Lernens mit und ohne Metadaten	52
18.	RoC Ergebnisse der Tokens in Bezug auf die verwendeten Algorithmen	55
19.	Metriken des MBL der Algorithmen	56
20.	RoC Ergebnisse der Algorithmen	57
21.	Mengendarstellung der MBL Ergebnisse	58
22.	Ergebnisse des Ensemblelernens	59
23.	Decision Tree des Ensembles	60
24.	Modell zur Untersuchung von SI Einflüssen	67
25.	Skizzenhafte Entwicklung der Anzahl an relativen SI in Bezug zum Softwarelebenszyklus	74