CARL VON OSSIETZKY universität OLDENBURG

Fakultät II – Informatik, Wirtschafts- und Rechtswissenschaften
Department für Informatik

Knowledge Transfer-based Recommendations to Enable Self-Service Business Intelligence

Dissertation

Submitted in fulfillment of the requirements for the degree of
Doktors der Ingenieurwissenschaften (Dr.-Ing.)

Submitted by
M.Sc. Safwan Sulaiman

Supervisors:
Prof. Dr.-Ing. Jorge Marx Gómez
Prof. Dr. Hans-Georg Kemper

Disputation Date: 08.08. 2019
Oldenburg, Germany

Acknowledgements

First and foremost, I would like to express my deep gratitude and appreciation to my first supervisor Prof. Dr.-Ing. Jorge Marx Gómez for his valuable guidance, encouragement, inspiration, and constructive criticism. He supported me during the whole journey of undertaking my PhD starting from the moment I planned to continue my study in Germany, and throughout the difficult times that I faced until reaching this moment where I am now.

I would also like to thank my second supervisor Prof. Dr. Hans-Georg Kemper for his effort and feedback. My heartfelt appreciation goes to Dr. Joachim Kurzhöfer for the long hours of invaluable discussions, especially regarding the practical part of my thesis. Many thanks go to the members of my examination committee, Prof. Dr. Andreas Winter and Dr.-Ing. Sven Rosinger.

I gratefully acknowledge the support provided by Damascus University through a scholarship for continuing my study in Germany. Without this support, it was not possible for me to reach the goal that I dreamed about.

My sincere thanks go to my colleagues at the VLBA department. I will not forget the enjoyable time I spent in the VLBA group, it has also been a pleasure for me to work with the students while supervising their research in masters' and bachelors' theses. I am deeply grateful to Neil Ross Paterson for his big effort in the proofreading of my thesis.

My deepest appreciation goes to my close friends, who supported, encouraged and offered me advice through the ups and downs of my research. Without you, the journey would have been more difficult.

The success was not possible without the blessings of my parents, who believed in me to achieve this degree. I owe my deepest gratitude to my father Sulaiman, my mother Wesal, and my brothers Wassem and Magd. I love you and I missed you a lot.

Last but not least, my heartfelt thanks go to my lovely wife Muna and my daughter Maria. Achieving this thesis was taking a lot of your time, especially when I spent the weekends and vacations at the university. You always inspired and motivated me, thank you very much for all your love.

Safwan Sulaiman
Oldenburg, August 2019

Abstract

Enterprises nowadays face multiple challenges while operating their businesses. These result from the continuous change in the business environment, which has become more complex and dynamic. Moreover, this environment is influenced by globalization, legal changes, volatile markets, and technical progress. In this complex business world, the integration of *Business Intelligence* (BI) systems in enterprises will help decision-makers and management by providing them with important information at the right time, enabling them to react quickly to dynamic changes in the market.

BI has been widely deployed in enterprises in order to help the employees in their decision-making process at all levels. Statistics from Gartner Group show that investment in the BI domain has recently been very high. On the other hand, different studies and market researches show that the pervasiveness and the percentage usage rate of BI are still very low. The reason behind this is the complexity of the use of BI systems. Enterprise users often lack the analytical and technical skills required to use these systems. Business users cannot get the needed information unless they ask for support from IT or power users. The resulting large number of business user requests leads to overload at the power users, who have become a bigger bottleneck than before.

To address the above-mentioned issues and to improve the usage rate of BI systems, a new concept for self-service BI is proposed. Within this system, the knowledge (know-how) of power users is extracted in form of analysis paths and stored in a knowledge repository. This knowledge then is delivered to business users in form of recommendations in real time while they use the BI system. This will enable business users to get answers about their business questions at the right time.

This research is intended to support business users in enterprises to get the information they need for their decision-making processes, without relying on the power users. First, a new architecture of self-service BI systems is conceptualized and developed. After that, a prototype is implemented as a proof of concept. Finally, in order to evaluate and validate the benefits of the new architecture for business users, different business scenarios are demonstrated using the prototype in several enterprises in form of workshops. Within these workshops, the results are discussed with different BI roles that include power users as well as business users.

Zusammenfassung

Heutzutage stehen Unternehmen vor vielfältigen Herausforderungen in der Geschäftsausführung. Diese resultieren aus dem ständigen Wandel des Geschäftsumfeldes, das komplexer und dynamischer geworden ist. Darüber hinaus wird dieses Umfeld durch Globalisierung, Gesetzesänderungen, schwankende Märkte und technischen Fortschritt beeinflusst. In dieser komplexen Geschäftswelt wird die Integration von *Business Intelligence* (BI) Systemen in Unternehmen Entscheidungsträgern und Management helfen, indem wichtige Informationen zum richtigen Zeitpunkt zur Verfügung stehen, um schnell auf dynamische Veränderungen im Markt zu reagieren.

BI wird breites weit verbreitet in Unternehmen eingesetzt, um die Mitarbeiter in ihrem Entscheidungsprozess auf allen Ebenen zu unterstützen. Statistiken der Gartner Group zeigen, dass in der letzten Zeit die Investitionen in den BI-Bereich sehr hoch waren. Andererseits zeigen verschiedene Studien und Marktforschungen, dass die Verbreitung und der prozentuale Nutzungsgrad von BI immer noch sehr gering sind. Ein Grund dafür ist die Komplexität der Nutzung von BI-Systemen. Unternehmensbenutzern fehlen oft die analytischen und technischen Fähigkeiten, die zur Verwendung dieser Systeme erforderlich sind. Business User können die benötigten Informationen nur erhalten, wenn sie Unterstützung von der IT-Abteilung oder Power User anfordern. Die daraus resultierende große Anzahl von Business User Anfragen führt zu einer Überlastung der Power User, die zu einem größeren Engpass werden.

Um die oben genannten Probleme zu behandeln und die Nutzungsrate von BI-Systemen zu verbessern, wird ein neues Konzept für das Self-Service-BI vorgeschlagen. In diesem System wird das Wissen der Power User in Form von Analysepfaden extrahiert und in einem Knowledge Repository gespeichert. Dieses Wissen wird dann den Business Usern in Form von Empfehlungen während der Nutzung des BI-Systems in Echtzeit vermittelt. Dadurch können Business User zum richtigen Zeitpunkt Antworten auf ihre Geschäftsfragen erhalten.

Diese Forschung soll Business User in Unternehmen dabei unterstützen, die Informationen zu erhalten, die sie für ihre Entscheidungsprozesse benötigen, ohne die Power User einzubinden. Zunächst wird eine neue Architektur für Self-Service BI-Systeme konzipiert und entwickelt. Danach wurde ein Prototyp als Proof of Concept

implementiert. Um die Vorteile der neuen Architektur für Business Usern zu bewerten und zu validieren, werden verschiedene Geschäftsszenarien anhand des Prototyps in mehreren Unternehmen in Form von Workshops demonstriert. Im Rahmen dieser Workshops werden die Ergebnisse mit verschiedenen BI-Rollen diskutiert, die sowohl Power User als auch Business User umfassen.

Table of Contents

Abstract I
Zusammenfassung II
Table of Contents IV
List of Abbreviations and Acronyms VIII
List of Figures X
List of Tables XII
1 Introduction 1
1.1 Motivation 4
1.2 Problem Definition 6
1.3 Main Research Goal and Objectives 8
1.4 Thesis Structure 10
2 Theoretical Foundation and Related Work 12
2.1 Main Approaches of Business Intelligence 12
2.1.1 Business Intelligence Definitions 12
2.1.2 Architecture of Business Intelligence System 15
2.1.3 Business Intelligence Tools 17
2.1.4 Dimensional Model 19
2.1.5 Typical OLAP Operations 22
2.1.6 Self Service Business Intelligence 24
2.1.7 Users of Business Intelligence Systems 26
2.2 Data Mining 27
2.2.1 Data Mining Techniques 27
2.2.2 Data Mining Process 29
2.2.2.1 Data Collection 29
2.2.2.2 Feature Extraction and Data Cleaning 30
2.2.2.3 Analytical Processing and Algorithms 30
2.2.3 Frequent Pattern Mining 30
2.2.3.1 Sequential Pattern Mining 32
2.3 Knowledge Management 34
2.3.1 Tacit Knowledge vs Explicit Knowledge 34
2.3.2 The Modes of Knowledge Conversion 35
2.3.3 Knowledge Transfer Model 36
2.4 Recommender Systems 37
2.4.1 Goals of Recommender Systems 37
2.4.2 Types of Recommender Systems 38
2.5 Related Work 40
2.5.1 Knowledge-based Business Intelligence for Business User Information Self Service 40

2.5.2 Distribution of Business Intelligence Knowledge 40
2.5.3 Recommender System's and TEL's Related Work 41
2.6 Summary 43
3 Research Methodology 44
3.1 Introduction 44
3.2 Design Science Research in Information Systems 45
3.3 Guidelines for Design Science in Information Systems Research 48
3.4 Memorandum on Design-Oriented Information Systems Research 50
3.5 Summary 52
4 System Requirements and Conception of the New SSBI Architecture 53
4.1 Basic Idea and Knowledge Flow 53
4.2 System Requirements 55
4.2.1 General Requirements of the Main Concept 55
4.2.2 Requirements of the Extraction of PUs' Knowledge 55
4.2.3 Requirements for Applying or Transferring of Knowledge to BUs 56
4.3 Analysis of Available BI Systems and the System Choice for this Research 56
4.3.1 Selection Model of BI Systems 57
4.3.2 Evaluation of BI Systems 59
4.4 Conception of the New SSBI Architecture 62
4.4.1 Tracing Module 64
4.4.2 Analysis Module 65
4.4.3 Knowledge Repository Database 66
4.4.4 Recommender System 66
4.5 Knowledge Transfer Model Workflows 67
4.5.1 Power User Knowledge Extraction 67
4.5.2 Analysis Path Recommendation for Business Users 68
4.6 Summary 69
5 Knowledge Transfer based Self-Service Business Intelligence 70
5.1 Tracing Module 70
5.1.1 Types of User Interaction Data 71
5.1.2 Data Collection in Human-Web Interactions Domain 71
5.1.2.1 Server-Side Data Collection 72
5.1.2.2 Client-Side Data Collection 73
5.1.3 Analysis Phase 73
5.1.3.1 Analysis of Logging Mechanisms 74
5.1.3.2 Analysis of User Interfaces 75
5.1.4 Conception of the Tracing Module 80
5.1.4.1 User Interactions Catalog 81
5.1.4.2 Requirement Definition 82
5.1.4.3 Tracing Module Architecture 83
5.2 Analysis Module and Sequential Pattern Algorithm 87

5.2.1 Problem Formulation of Analysis Module 88
5.2.2 Phases of the Analysis Module 91
5.2.3 Factors Catalog of Analysis Module 93
5.2.3.1 Preparation of the Input Data (Phase 1) 94
5.2.3.2 Configuration of the Algorithm (Phase 2) 95
5.2.3.3 Algorithm Execution (Phase 3) 97
5.2.3.4 Post-Preparation of the Resulting Patterns (Phase 4) 98
5.2.3.5 Storage of Analysis Paths (Phase 5) 99
5.2.4 Requirements of the Analysis Module 100
5.2.5 Design and Conception of the Analysis Module 102
5.2.5.1 Model-View-Controller Design Pattern 102
5.2.5.2 Analysis Module Architecture 103
5.3 Knowledge Repository 105
5.3.1 Knowledge Repository Representation Schema 105
5.3.2 Domain Knowledge 107
5.4 Recommender System 107
5.4.1 Requirement Definition 108
5.4.2 Recommender System Architecture 109
5.4.2.1 Presentation Layer 110
5.4.2.2 Application Layer 111
5.4.2.3 Data Layer 115
5.5 Summary 116
6 Prototypical Implementation and Evaluation 117
6.1 Implementation Aspects of the Prototype 117
6.1.1 Tracing Module (BI-Tracer) 117
6.1.1.1 Implementation and Technologies 118
6.1.1.2 Mapping of Interactions Catalog 119
6.1.1.3 Prototype of BI Tracer 119
6.1.2 Analysis Module 120
6.1.2.1 Component Interactions of the Analysis Module 121
6.1.2.2 Parsing the Trace Files 122
6.1.2.3 Process Flow of the Analysis Module Prototype 123
6.1.3 Knowledge Repository 124
6.1.3.1 Storing the Analysis Paths with their Metadata in the Knowledge Repository 124
6.1.4 Recommender System 126
6.1.5 Datasets Used for Building Business Scenarios 128
6.2 Evaluation 130
6.2.1 Functional Evaluation 131
6.2.2 Use Case Evaluation 135
6.2.2.1 Industry Partners for the Evaluation of this Research 136
6.2.2.2 Feedback and Discussion of the Expert Interviews 140
6.2.2.3 Implications for the Partner Companies 142

6.2.2.4 SSBI Application Guidance 145
6.2.3 Argumentative Evaluation 153
6.2.4 Utilization of Knowledge-based Recommendations in Smart Cities Applications 153
6.2.4.1 Problem Definition 154
6.2.4.2 Smart Cities 156
6.2.4.3 Architecture for Enabling and Adopting of Smart City Applications 157
6.2.4.4 Discussion 160
6.3 Summary 161
7 Conclusion and Outlook 162
7.1 Summary of the Research 162
7.2 Outlook 163
References 165
Publications 181
Appendix 182
A. Excerpt of User Interactions Catalog 182
B. Questionnaire-based Expert Interview 184

List of Abbreviations and Acronyms

AP	Analysis Path
BARC	Business Application Research Center
BI	Business Intelligence
BU	Business User
CASE	Computer-aided Software Engineering
CIM	Corporate Information Management
CIO	Chief Information Officer
CRM	Customer Relation Management
CSS	Cascading Style Sheets
CSV	Comma-Separated Values
DBMS	Database Management Systems
DIY BI	DO IT Yourself Business Intelligence
DNA	Deoxyribonucleic Acid
DSR	Design Science Research
DSRM	Design Science Research Methodology
DSS	Decision Support System
EIS	Executive Information System
ER	Entity Relationship
ERP	Enterprise Resource Planning
ETL	Extracted, Transformed and Loaded
FSF	Free Software Foundation
GPL	General Public License
HCI	Human-Computer Interactions
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
ICT	Information and Communications Technologies
IDE	Integrated Development Environment
IP	Internet Protocol
IS	Information Systems
IT	Information Technology

J2EE	Java 2 Platform Enterprise Edition
JDBC	Java Database Connectivity
JSON	JavaScript Object Notation
KDD	Knowledge Discovery in Databases
KM	Knowledge Management
KPI	Key Performance Indicators
KTM	Knowledge Transfer Model
KTR	Knowledge Transfer-based Recommendations
MDX	Multi-Dimensional Expressions
MIS	Minimum Item Support
MVC	Model-View-Controller
OSS	Open Source Software
PC	Personal Computer
PEOU	Perceived Ease-Of-Use
POS	Point of Sale
PU	Power User
PUF	Perceived Usefulness
SCM	Supply Chain Management
SDB	Sequence Database
SECI	Socialization, Externalization, Combination, and Internalization
SPM	Sequential Pattern Mining
SPMF	Sequential Pattern Mining Framework
SQL	Structured Query Language
SSBI	Self-Service Business Intelligence
SSO	Single Sign-On
TAM	Technology Acceptance Model
TDWI	The Data Warehousing Institute
TEL	Technology Enhanced Learning
UI	User Interface
WWW	World Wide Web
XML	Extensible Markup Language

List of Figures

Figure 1. Market size of BI and analytics ($ billions) 2
Figure 2. Usage rate of BI systems 5
Figure 3. Median percentage of BI users in companies 6
Figure 4. Classes of BI applications 7
Figure 5. Knowledge transfer model 9
Figure 6. Thesis structure 11
Figure 7. Historical development of Decision Support Systems 13
Figure 8. Traditional architecture of BI systems 16
Figure 9. Retail sales scheme 21
Figure 10. Query results and cross-tabular report 21
Figure 11. Examples of typical OLAP operations on multidimensional data. 23
Figure 12. BI trend monitor 25
Figure 13. Four objectives of Do-It-Yourself or Self-Service BI. 26
Figure 14. The Data Mining process 29
Figure 15. Market basket analysis 31
Figure 16. SECI process 36
Figure 17. Information Systems Research Framework 46
Figure 18. DSRM process model 48
Figure 19. Abstract model of knowledge transfer 53
Figure 20. Knowledge transfer model between power and business users 54
Figure 21. Automatic knowledge transfer-based architecture for SSBI 63
Figure 22. Automatic knowledge transfer model workflows 68
Figure 23. Pentaho OLAP analysis 77
Figure 24. JasperReports analysis area 78
Figure 25. SpagoBI analysis area 79
Figure 26. Tracing module architecture 84
Figure 27. Process of capturing power user's interactions 86
Figure 28. Process of capturing business user's interactions 87
Figure 29. Excerpt of the trace files 89
Figure 30. Phases of the analysis module 92
Figure 31. The MVC architectural patterns 103
Figure 32. Analysis module architecture 104

Figure 33. Knowledge repository scheme 106
Figure 34. Conceptual Architecture of the Recommendation-based Self-Service BI ... 110
Figure 35. Recommendations matching and sorting algorithm 113
Figure 36. BI Tracer prototype 120
Figure 37. Components interactions of the analysis module 121
Figure 38. Knowledge repository scheme of the prototype 125
Figure 39. Component interactions of the recommender system 126
Figure 40. Building the manifest files of the recommender system for Google Chrome extension 127
Figure 41. Star scheme of ACME dataset 129
Figure 42. Schema of Chinook dataset 129
Figure 43. Testing of observer interface 133
Figure 44. Power user OLAP analysis 146
Figure 45. Resulting trace file of power user 147
Figure 46. Pentaho home area of business user 149
Figure 47. Recommendation to expand the dimension “year” 150
Figure 48. Displaying of the evaluator 151
Figure 49. Recommendation to change the chart settings 152
Figure 50. Number of smartphone users in Germany from 2015 to 2022 (in millions) 155
Figure 51. Knowledge-based recommendations architecture to enable self-service Smart City applications 158

List of Tables

Table 1. Technologies expected to help businesses to differentiate themselves from their competitors 3
Table 2. Sequence database 33
Table 3. Extracted sequential patterns 33
Table 4. Function matrix of BI tools 60
Table 5. Results of the comparative analysis among the three open source BI systems 62
Table 6. Attributes of the interaction 81
Table 7. Excerpt of the user interactions catalog 88
Table 8. Sequence database of trace files 90
Table 9. Possible sequential patterns 91
Table 10. Three parameter fields in a morphological format 94
Table 11. Preparation of the input data 95
Table 12. Configuration of the algorithm 96
Table 13. Execution of the algorithm 98
Table 14. Post-preparation of the resulting patterns 99
Table 15. Storage of analysis paths 100
Table 16. Selected technologies for tracing module 118
Table 17. Excerpt of interactions catalog - mapping 119
Table 18. Excerpt of trace file (Part1) 122
Table 19. Excerpt of trace file (Part2) 122
Table 20: Excerpt of trace file (Part3) 122
Table 21. Selected technologies for recommender system 128
Table 22. Selected evaluation methods 130
Table 23. Evaluation results of the mining algorithms 134
Table 24. Excerpt of the interactions catalog 183