

Information – Organisation – Produktion

Hrsg.: Hans Corsten, Michael Reiß, Claus Steinle,  
Stephan Zelewski

Stefan Beißel

# **Ontologiegestütztes Case-Based Reasoning**



**RESEARCH**

Stefan Beiel

## **Ontologiegesttztes Case-Based Reasoning**

# **GABLER RESEARCH**

Information – Organisation – Produktion

Herausgegeben von Professor Dr. Hans Corsten,  
Professor Dr. Michael Reiß, Professor Dr. Claus Steinle  
und Professor Dr. Stephan Zelewski

Die Schriftenreihe präsentiert Konzepte, Modelle und Methoden zu drei zentralen Domänen der Unternehmensführung. Information, Organisation und Produktion werden als Bausteine eines integriert angelegten Managementsystems verstanden. Der Erforschung dieses Bereiches dienen sowohl theoretische als auch anwendungsorientierte Beiträge.

Stefan Beißel

# **Ontologiestütztes Case-Based Reasoning**

Entwicklung und Beurteilung  
semantischer Ähnlichkeitsindikatoren  
für die Wiederverwendung  
natürlichsprachlich repräsentierten  
Projektwissens

Mit einem Geleitwort von Univ.-Prof. Dr. Stephan Zelewski



**RESEARCH**

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

Dissertation Universität Duisburg-Essen, 2011

1. Auflage 2011

Alle Rechte vorbehalten

© Gabler Verlag | Springer Fachmedien Wiesbaden GmbH 2011

Lektorat: Stefanie Brich | Nicole Schweitzer

Gabler Verlag ist eine Marke von Springer Fachmedien.

Springer Fachmedien ist Teil der Fachverlagsgruppe Springer Science+Business Media.

[www.gabler.de](http://www.gabler.de)



Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlags unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Umschlaggestaltung: KünkelLopka Medienentwicklung, Heidelberg

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Printed in Germany

ISBN 978-3-8349-3064-4

## Geleitwort

Das Werk von Herrn Dr. Beißel befasst sich mit einem herausfordernden Problemkomplex, der an der Nahtstelle zwischen Betriebswirtschaftslehre, Wirtschaftsinformatik und Informatik – letztgenannte einschließlich der Erforschung Künstlicher Intelligenz (KI) – angesiedelt ist.

Aus betriebswirtschaftlicher Sicht behandelt der Autor ein zentrales Realproblem des Projektmanagements. Es betrifft die Schwierigkeit, zur Vorbereitung (z. B. im Rahmen der Teilnahme an Projektausschreibungen), zur Durchführung und auch zur Kontrolle von neuen Projekten dasjenige Wissen zu nutzen, das über bereits durchgeführte, ältere Projekte in durchaus beachtlichem Umfang zur Verfügung steht. Als Barrieren gegenüber der Wiederverwendung von „an sich“ verfügbarem Projektwissen erweisen sich vor allem zwei Aspekte. Einerseits erreicht das Wissen, das über bereits durchgeführte Projekte sowohl im eigenen Unternehmen als auch in branchenweit oder öffentlich zugänglichen Publikationen grundsätzlich vorhanden ist, oftmals einen derart großem Umfang, dass es von den zuständigen Mitarbeitern eines Unternehmens in „vertretbaren“ Zeiträumen von einigen Stunden oder Tagen bei Weitem nicht vollständig ausgewertet werden kann. Andererseits scheitert der Einsatz der Automatischen Informationsverarbeitung in der Regel daran, dass das einschlägige Wissen über bereits durchgeführte Projekte zum großen Teil in natürlichsprachlicher Form repräsentiert wird, wie z. B. in „Projektberichten“, „Business Cases“ und Protokollen von „Debriefings“.

Aus den vorgenannten Gründen stellt es ein betriebswirtschaftliches Desiderat dar, umfangreiches und größtenteils natürlichsprachlich repräsentiertes Wissen über alte Projekte computergestützt auswerten und auf neue Projekte anwenden zu können. Die konventionellen Techniken, die seitens der Wirtschaftsinformatik und der Informatik zur Bewältigung solcher Wissensverarbeitungsprobleme angeboten werden, reichen jedoch für die oben skizzierte Aufgabe des Managements von Projektwissen nicht aus. Zu solchen konventionellen Techniken zählen beispielsweise Text- und Dokumentenverarbeitungstechniken sowie Information-Retrieval- und Business-Intelligence-Techniken. Sie leiden alle darunter, dass sie zwar große Informationsmengen zu verarbeiten vermögen, dies jedoch lediglich auf syntaktisch-statistischer Basis. Sie sind aber nicht – zumindest derzeit noch nicht – in der Lage, komplex strukturiertes und natürlichsprachlich repräsentiertes Wissen, wie es etwa im Falle von Wissen über Projekte vorliegt, auf der semantischen Ebene mit inhaltlichem Verständnis für das jeweils repräsentierte Wissen zu verarbeiten.

Aus der Erforschung Künstlicher Intelligenz sind zwei Techniken bekannt, die grundsätzlich dazu geeignet sind, zwei unterschiedliche Aspekte des zuvor skizzierten Re-alproblems erfolgreich zu behandeln. Einerseits werden seit längerer Zeit Ontologien erforscht, die Computer in die Lage versetzen, natürlichsprachlich repräsentiertes Wissen mit inhaltlichem Verständnis auf der semantischen Ebene zu verarbeiten. Andererseits wird die Technik des Case-based Reasonings seit Längerem dafür eingesetzt, um Wissen über alte Fälle – wie z. B. Projekte – aufgrund von ähnlichkeitsbasierten Analogieschlüssen auf neue Fälle zu übertragen. Diese Technik bleibt aber in der Regel auf die Verarbeitung numerischen Wissens beschränkt, weil die essenziellen Ähnlichkeitsmetriken numerische Fallattribute voraussetzen scheinen. Aufgrund dieser Verschiedenartigkeit von Ontologien einerseits und Case-based Reasoning andererseits sind die beiden Techniken der Ontologien und des Case-based Reasonings bislang noch kaum miteinander kombiniert worden. Darüber hinaus haben sie auch einzeln, bis auf einige wenige „randständige“ Anwendungen außerhalb der „etablierten“ Forschungsstränge, noch keinen substantziellen Eingang in die betriebswirtschaftliche Forschung gefunden.

Der Autor hat sich der großen Herausforderung gestellt, die beiden Techniken der Ontologien und des Case-based Reasonings im Rahmen des computergestützten Managements von projektbezogenem Wissen erstmals miteinander zu kombinieren. Damit hat er an der „vordersten Forschungsfront“ einen bemerkenswerten Beitrag zum betriebswirtschaftlichen Erkenntnisfortschritt geleistet. Darüber hinaus hat er mit großer Liebe zum Detail ein Konzept für ein systematisches computergestütztes Wissensmanagement als Teil des betrieblichen Projektmanagements entwickelt. Es ist dabei behilflich, den innovativen Verbund aus Ontologien und Case-based Reasoning in die Geschäftsroutinen des betrieblichen Alltags zu integrieren und unterstreicht auf diese Weise auch die hohe Praxisrelevanz der hier vorgelegten Forschungsergebnisse.

Insbesondere die „intelligente“ Lösung des Problems, die Ähnlichkeit zwischen Projekten im Hinblick auf Projektwissen zu ermitteln, das zum größeren Teil nur in natürlichsprachlich repräsentierter Form vorliegt, beeindruckt durch ihren innovativen Charakter. Dazu gehört vor allem der Ansatz des Autors, implizites Ähnlichkeitswissen, das in einer natürlichsprachlichen Projektmanagement-Ontologie verborgen ist, mittels einer inhaltsbezogenen und somit semantischen „Vermessung“ von Ähnlichkeitsabständen in einer Ontologie und darauf aufbauendem Projektwissen zu explizieren. Darüber hinaus beweist der Autor eine beeindruckende Transferkompetenz, indem er seine konzeptionellen Erkenntnisse auf die prototypische Implementierung eines ontologiegestützten Case-based-

Reasoning-Systems erfolgreich übertragen hat. Diese gut gelungene Implementierung eines Software-Prototyps nötigt dem Verfasser dieses Geleitworts großen Respekt ab. Der Autor sollte in Erwägung ziehen, seinen Prototyp zu einer kommerziell verwertbaren, professionellen Software für die betriebliche Praxis weiterzuentwickeln oder weiterentwickeln zu lassen.

Der Autor hat mit seiner hier veröffentlichten Dissertation eine hervorragende interdisziplinäre Forschungsarbeit an der Nahtstelle zwischen Betriebswirtschaftslehre, Wirtschaftsinformatik, Informatik und KI-Forschung geleistet. Seine Untersuchungen beeindrucken nicht nur durch ihren innovativen Charakter, sondern auch die profunden Sach- und Methodenkenntnisse des Autors sowie seine bemerkenswerte Liebe zur Detailarbeit. Hierdurch geht der Autor weit über das hinaus, was von einer „üblichen“ betriebswirtschaftlichen Dissertation erwartet werden kann. Aus den vorgenannten Gründen ist den Ausführungen des Autors eine möglichst breite Resonanz sowohl im wissenschaftlichen Bereich als auch in der betrieblichen Praxis zu wünschen. Dies gilt nicht nur für betriebswirtschaftlich interessierte Leserinnen und Lesern aus dem engeren Anwendungsgebiet des Projektmanagements, sondern ebenso für Rezipienten aus Wirtschaftsinformatik, Informatik und KI-Forschung, die sich für betriebswirtschaftliche Anwendungen ihrer Forschungsergebnisse interessieren.

Stephan Zelewski

## **Vorwort**

Die vorliegende Dissertation wurde im Dezember 2010 vom Promotionsausschuss der Fakultät für Wirtschaftswissenschaften der Universität Duisburg-Essen angenommen.

Die Dissertation befasst sich mit der Entwicklung einer Technik, um die Wiederverwendung von natürlichsprachlich repräsentiertem Projektwissen zu unterstützen. Bei dieser Technik handelt es sich um das ontologiegestützte Case-Based Reasoning, das die Aufgabe der Identifikation von hinreichend ähnlichen Projektfällen und Selektion von mindestens einem ähnlichsten Projektfall computergestützt durchführt. Die Machbarkeit wird anhand eines Prototyps demonstriert, der ein Ontologie-Tool und ein Case-Based-Reasoning-Tool beinhaltet und für Projekte aus dem Bereich des IT-Managements angewendet wird. Die Dissertation setzt sich aus zwei Hauptbestandteilen zusammen. Im ersten Teil erfolgt eine Erläuterung der Grundlagen für die Problembearbeitung, speziell zu Projektmanagement, Ontologien und Case-Based Reasoning. Im zweiten Teil erfolgt die Planung, Entwicklung und Anwendung eines ontologiegestützten Case-Based-Reasoning-Systems.

An dieser Stelle möchte ich mich bei Professor Dr. Stephan Zelewski für seine umfassende Unterstützung bedanken. Seine anregenden und fachlich beeindruckenden Hinweise und Ideen waren mir stets eine große Unterstützung. Außerdem möchte ich mich bei Professor Dr. Heimo H. Adelsberger für die Erstellung des Zweitgutachtens bedanken.

Mein Dank gilt auch meiner Frau Andrea für Ihre Geduld und Ermutigung.

Stefan Beißel

## Inhaltsverzeichnis

<b>Abkürzungs- und Akronymverzeichnis.....</b>	<b>XV</b>
<b>Symbolverzeichnis.....</b>	<b>XXI</b>
<b>Abbildungsverzeichnis.....</b>	<b>XXIII</b>
<b>Tabellenverzeichnis.....</b>	<b>XXVII</b>
<b>1 Einführung in den Problembereich technikgestützten Managements von Projektwissen.....</b>	<b>1</b>
1.1 Betriebswirtschaftliches Desiderat .....	1
1.2 State-of-the-art der verfügbaren Techniken zur Erfüllung des betriebswirtschaftlichen Desiderats.....	4
1.3 Wissenschaftliche Problemstellung.....	6
1.4 Ziele der Untersuchung .....	12
1.5 Aufbau der Untersuchung .....	13
<b>2 Grundlagen für die Problembearbeitung .....</b>	<b>17</b>
2.1 Projekte und Projektmanagement als relevanter Realitätsbereich .....	17
2.2 Wissensmanagement .....	19
2.2.1 Natürlichsprachliche Repräsentation von Projektwissen .....	19
2.2.2 Definition von Wissensmanagement.....	21
2.3 Übersicht über die Techniken für die Problemlösung.....	22
2.3.1 Ontologiestützung .....	22
2.3.1.1 Definition von Ontologien .....	22
2.3.1.2 Repräsentationssprachen .....	24
2.3.1.3 Abgrenzung von alternativen Techniken .....	27
2.3.1.4 Vor- und Nachteile der Ontologie für die Untersuchung .....	30
2.3.2 Case-Based Reasoning .....	32
2.3.2.1 Definition von Case-Based Reasoning.....	32

2.3.2.2	Abgrenzung von alternativen Techniken .....	35
2.3.2.3	Vor- und Nachteile des Case-Based Reasonings für die Untersuchung .....	37
2.3.3	Eignung des ontologiegestützten Case-Based Reasonings .....	40
<b>3</b>	<b>Ontologiegestütztes Case-Based Reasoning für das Management von Projektwissen.....</b>	<b>41</b>
3.1	Definition eines Vorgehensmodells für die Erstellung des ontologiegestützten CBR-Systems .....	41
3.2	Anwendung des Vorgehensmodells .....	48
3.2.1	Konzept des ontologiegestützten CBR-Systems .....	48
3.2.1.1	Umfang des Konzepts .....	48
3.2.1.2	Anforderungsanalyse für die Softwareauswahl.....	48
3.2.1.2.1	Grundlagen der Softwareauswahl .....	48
3.2.1.2.2	Selektion einer Technik zur Softwareauswahl .....	49
3.2.1.2.3	Anforderungskatalog für ein Ontologie-Tool .....	55
3.2.1.2.4	Anforderungskatalog für ein CBR-Tool .....	69
3.2.1.3	Systemanalyse .....	81
3.2.2	Entwurf des ontologiegestützten CBR-Systems .....	84
3.2.2.1	Ziele des Entwurfs.....	84
3.2.2.2	Auswahl eines Ontologie-Tools .....	85
3.2.2.2.1	Marktanalyse zur Identifizierung von Alternativen .....	85
3.2.2.2.2	Bewertung der Alternativen .....	88
3.2.2.2.3	Selektion der geeignetsten Alternative.....	109
3.2.2.3	Auswahl eines CBR-Tools .....	113
3.2.2.3.1	Marktanalyse zur Identifizierung von Alternativen .....	113
3.2.2.3.2	Bewertung der Alternativen .....	114
3.2.2.3.3	Selektion der geeignetsten Alternative.....	130
3.2.2.4	Definition der Schnittstellen zwischen dem Ontologie-Tool und dem CBR-Tool.....	133
3.2.2.5	Rückkopplungen durch Schnittstellen-Erkenntnisse.....	133
3.2.2.6	Auswahlentscheidung hinsichtlich des Ontologie-Tools und des CBR- Tools.....	133
3.2.3	Implementierung des ontologiegestützten CBR-Systems .....	134
3.2.3.1	Ziele der Implementierung .....	134

3.2.3.2	Erstellung der Fallbasis .....	134
3.2.3.3	Erstellung der Ontologie .....	139
3.2.3.3.1	Ziele und Umfang.....	139
3.2.3.3.2	Methode zur Erstellung der Ontologie .....	140
3.2.3.3.3	Anwendung der Methode zur Erstellung der Ontologie .....	140
3.2.3.3.3.1	Bestimmung der Domäne .....	140
3.2.3.3.3.2	Nutzung bestehender Ontologien .....	141
3.2.3.3.3.3	Sammlung wichtiger Begriffe für die Ontologie.....	141
3.2.3.3.3.4	Definition der Klassen und der Klassenhierarchie.....	143
3.2.3.3.3.5	Definition der Attribute und Relationen.....	146
3.2.3.3.3.6	Definition der Wertebereiche und der Kardinalitäten von Attributen und Relationen.....	152
3.2.3.3.3.7	Erzeugung von Instanzen .....	155
3.2.3.4	Erstellung von Ähnlichkeitsmaßstäben .....	159
3.2.3.5	Erstellung eines Algorithmus zur Aggregation von lokalen Ähnlichkeitswerten.....	167
3.2.3.6	Erstellung eines Prototyps für ontologiegestütztes Case-Based Reasoning.....	174
3.2.3.7	Test des Prototyps für ontologiegestütztes Case-Based Reasoning .....	191
<b>4</b>	<b>Fazit.....</b>	<b>217</b>
<b>5</b>	<b>Ausblick .....</b>	<b>219</b>
	<b>Literaturverzeichnis .....</b>	<b>221</b>
	<b>Anhang .....</b>	<b>257</b>

**Abkürzungs- und Akronymverzeichnis**

AAAI	Association for the Advancement of Artificial Intelligence
AHP	Analytic Hierarchy Process
AIAI	Artificial Intelligence Applications Institute
AIX	Advanced Interactive eXecutive
AMD	Advanced Micro Devices, Incorporated
ANP	Analytic Network Process
API	Application Programming Interface
ASCII	American Standard Code for Information Interchange
ASWC	Asian Semantic Web Conference
AT&T	American Telephone and Telegraph Company
BMIR	Biomedical Informatics Research
BNCOD	British National Conference on Databases
BSD	Berkeley Software Distribution
BSI	Bundesamt für Sicherheit in der Informationstechnik
bzgl.	bezüglich
bzw.	beziehungsweise
C.I.	Consistency Index
C.R.	Consistency Ratio
CAISE	Conference on Advanced Information Systems Engineering
CASL	Common Algebraic Specification Language
CBR	Case-Based Reasoning
CBR-Tool	Case-Based-Reasoning-Tool
CBR-System	Case-Based-Reasoning-System
CC	Common Criteria for Information Technology Security Evaluation
CHISEL	Computer Human Interaction & Software Engineering Lab
CO-ODE	Collaborative Open Ontology Development Environment
CPU	Central Processing Unit
CSV	Comma-separated Values
DAML	DARPA Agent Markup Language
DARPA	Defense Advanced Research Projects Agency
DEA	Data Envelopment Analysis
DFKI	Deutsches Forschungszentrum für Künstliche Intelligenz

---

DIN	Deutsches Institut für Normung
DL	Description Logics
DNS	Domain Name System
DoD	Department of Defense
dt.	deutsch
DV	Datenverarbeitung
e.V.	eingetragener Verein
ECAI	European Conference on Artificial Intelligence
ECCBR	European Conference on Case-Based Reasoning
ECIR	European Conference on Information Retrieval
ER	Entity-Relationship
ERD	Entity-Relationship Diagramm
ESWC	European Semantic Web Conference
et al.	et alii
EWMF	European Web Mining Forum
f.	folgende
ff.	fort folgende
FLAIRS	Florida Artificial Intelligence Research Society
F-Logic	Frame Logic
FTP	File Transfer Protocol
FQAS	Flexible Query Answering Systems
FQDN	Full Qualified Domain Name
FZI	Forschungszentrum Informatik
ggf.	gegebenenfalls
GAIA	Group of Artificial Intelligence Applications
GIS	Geographic Information Systems
GPL	General Public License
GPM	Gesellschaft für Projektmanagement
GUI	Graphical User Interface
H.	Heft
Hrsg.	Herausgeber
HTML	Hypertext Markup Language
IBM	International Business Machines Corporation
ICADL	International Conference on Asian Digital Libraries

---

ICCBR	International Conference on Case-Based Reasoning
ICONIP	International Conference on Neural Information Processing
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IFM	Institute for Manufacturing
IJCAI	International Joint Conference on Artificial Intelligence
INRECA	Induction and Reasoning from Cases
IP	Internet Protocol
IR	Information Retrieval
ISBMDA	International Symposium on Biological and Medical Data Analysis
ISO	International Organization for Standardization
IST	Information Society Technologies
iSQI	International Software Quality Institute
IT	Informationstechnologie
IWT	Innovatie door Wetenschap en Technologie
jCOLIBRI	java Cases and Ontology Libraries Integration for Building Reasoning Infrastructure
Jg.	Jahrgang
JVM	Java Virtual Machine
KDO	Knowledge Discovery and Ontologies
KI	Künstliche Intelligenz
KIF	Knowledge Interchange Format
KMU	Klein- und Mittelunternehmen
KSL	Knowledge Systems, AI Laboratory
LP	Limited Partnership
max.	maximal
min.	mindestens
MSDOS	Microsoft Disk Operating System
MyCBR	My Case-Based Reasoning
NAT	Network Address Translation
ncRNA	Non-coding Ribonucleic Acid
NLP	Natural Language Processing
No.	Number
Nr.	Nummer
o.a.	oben angeführt

---

OCML	Operational Conceptual Modelling Language
OIL	Ontology Inference Layer
o.J.	ohne Jahresangabe
OKBC	Open Knowledge Base Connectivity
OntoViz	Ontology Visualization
OSX	Operating System X
o.V.	ohne Verfasser
OWL	Web Ontology Language
OXML	OntoStudio Extensible Markup Language
PAL	Protégé Axiom Language
PC	Personal Computer
PMI	Project Management Institute
PMBOK	Project Management Body of Knowledge
PowerPC	Performance optimization with enhanced RISC Performance Chip
RACER	Renamed Abox and Concept Expression Reasoner
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
RIF	Rule Interchange Format
RISC	Reduced Instruction Set Computer
RuleML	Rule Markup Language
S.	Seite
SAP R/3	Systeme, Anwendungen und Produkte Real-time data processing/3-tier
SAS	Statement on Auditing Standards
Sek.	Sekunden
s.o.	siehe oben
SPARQL	SPARQL Protocol and RDF Query Language
SPSS	Superior Performing Software System
SWO	Swoop Ontology Object File
SWRL	Semantic Web Rule Language
SQL	Structured Query Language
TCP/IP	Transmission Control Protocol/Internet Protocol
TGViz	Touch-Graph Visualization
TXT	Text
u.	und

---

u.a.	unter anderem
UK	United Kingdom
UML	Unified Modeling Language
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
US	United States
USA	United States of America
usw.	und so weiter
v. Chr.	vor Christus
vgl.	vergleiche
VMware	Virtual Machine Ware
Vol.	Volume
VPN	Virtual Private Network
W3C	World Wide Web Consortium
WG	Working Group
WSUS	Windows Server Update Services
XML	Extensible Markup Language
z.B.	zum Beispiel

**Symbolverzeichnis**

$a_{ij}$	Paarvergleichsurteil zwischen Vergleichsobjekt $i$ und $j$
$a_i^f$	Attributswert des Falls $f$
$a_i^q$	Attributswert des Anfrage-Falls $q$
$A_i^f$	Attribut des Falls $f$
$A_i^q$	Attribut des Anfrage-Falls $q$
$A$	Evaluationsmatrix (AHP)
$c^s$	Klasse $s$
$c_i^f$	Klasse, zu der die Instanz mit dem Attribut $A_i^f$ gehört
$c_i^q$	Klasse, zu der die Instanz mit dem Attribut $A_i^q$ gehört
$d$	Abstand zwischen numerischen Attributswerten
$e$	Schwellenwert für hinreichende Ähnlichkeit
$f$	Fall
$F$	Menge aus Fällen
$F_n$	Fall $n$
$G_z$	Gewicht des Beurteilungskriteriums $z$ (Nutzwertanalyse)
$i$	Vergleichsobjekt (AHP)
$I$	Einheitsmatrix
$j$	Vergleichsobjekt (AHP)
$k$	Hilfswert für die Ermittlung der Ähnlichkeit zwischen zwei Klassen
$\lambda$	Eigenwert
$\lambda_{\max}$	maximaler Eigenwert
$M$	Menge mit Ähnlichkeitswerten
$n$	natürliche Zahl
$N$	normierte Evaluationsmatrix (AHP)
$N_y$	Nutzwert einer Entscheidungsalternative $y$ (Nutzwertanalyse)
$p_i$	Priorität zum Vergleichsobjekt $i$ (AHP)
$p_i^{\text{id}}$	Priorität zum Vergleichsobjekt $i$ im Ideal Mode (AHP)

$P_j$	Gesamtpriorität einer Alternative $j$ (AHP)
$P_{yz}$	Punktwert, den eine Entscheidungsalternative $y$ beim Beurteilungskriterium $z$ erreicht (Nutzwertanalyse)
$q$	Anfrage-Fall
$r_h^f$	Relationswert des Falls $f$
$r_h^q$	Relationswert des Anfrage-Falls $q$
$rvi1$	relational verknüpfte Instanz 1
$rvi2$	relational verknüpfte Instanz 2
$s_i$	Zeilensumme zu Vergleichsobjekt $i$ (AHP)
$sim_a(a_i^q, a_i^f)$	Wert für die Ähnlichkeit zwischen den Attributwerten $a_i^q$ und $a_i^f$
$sim_c(c_h^q, c_h^f)$	Wert für die Ähnlichkeit zwischen den Klassen $c_h^q$ und $c_h^f$
$sim_g(q, f)$	Wert für die Ähnlichkeit zwischen einem Anfrage-Fall $q$ und einem Fall $f$
$sim_r(r_h^q, r_h^f)$	Wert für die Ähnlichkeit zwischen den Relationswerten $r_h^q$ und $r_h^f$
$t$	Transition (PETRI-Netz)
$tab(a_i^q, a_i^f)$	Ähnlichkeitswert im Tabellenfeld mit Zeilenindex $a_i^q$ und Spaltenindex $a_i^f$
$tax(a_i^q, a_i^f)$	Ähnlichkeitswert in der Taxonomie, der an dem gemeinsamen übergeordneten Knoten der Knoten $a_i^q$ und $a_i^f$ hinterlegt ist
$u_i$	Gewicht einer Ähnlichkeit
$v$	Eigenvektor
$v_i$	Bedeutungsurteil (AHP)
$w_i$	aggregiertes Bedeutungsurteil zu Vergleichsobjekt $i$ (AHP)
$x$	Ähnlichkeit
$y$	Entscheidungsalternative (Nutzwertanalyse)
$z$	Beurteilungskriterium (Nutzwertanalyse)
$\vec{0}$	Nullvektor
$\{\}$	leere Menge

**Abbildungsverzeichnis**

Abbildung 1:	Aufbau der Untersuchung in Form eines PETRI-Netzes .....	15
Abbildung 2:	Wissensgebiete im PMBOK-Guide .....	18
Abbildung 3:	CBR-Cycle.....	34
Abbildung 4:	Fallidentifikation und -selektion im ontologiegestützten Case-Based Reasoning .....	35
Abbildung 5:	Wasserfallmodell .....	42
Abbildung 6:	Spezifische Ausgestaltung des Wasserfallmodells .....	45
Abbildung 7:	Auswahlverfahren für die Tools im Wasserfallmodell.....	46
Abbildung 8:	Qualitätsmodell für externe und interne Softwarequalität.....	57
Abbildung 9:	Kriterienebenen für die Auswahl eines Ontologie-Tools .....	62
Abbildung 10:	Bedeutungsurteile in Expert Choice für die Ontologie-Tool-Auswahl .....	69
Abbildung 11:	Beispiel für Ähnlichkeitswerte in einer Taxonomie.....	75
Abbildung 12:	Kriterienebenen für die Auswahl eines CBR-Tools .....	76
Abbildung 13:	Bedeutungsurteile in Expert Choice für die CBR-Tool-Auswahl .....	81
Abbildung 14:	Ontologiegestütztes CBR-System .....	82
Abbildung 15:	Semantisches Datenmodell zum ontologiegestützten Case-Based Reasoning .....	83
Abbildung 16:	Hierarchie des Entscheidungsproblems .....	90
Abbildung 17:	Fehlerzustand in Apollo.....	97
Abbildung 18:	Graphical User Interface von Apollo.....	99
Abbildung 19:	Graphical User Interface von OntoStudio .....	99
Abbildung 20:	Graphical User Interface von OntoTrack .....	100
Abbildung 21:	Graphical User Interface von Protégé .....	100
Abbildung 22:	Graphical User Interface von Swoop.....	101
Abbildung 23:	Gesamtprioritäten in Expert Choice .....	111
Abbildung 24:	Gesamtprioritäten nach Entfernen der Alternative Protégé.....	112
Abbildung 25:	Hierarchie des Entscheidungsproblems .....	116
Abbildung 26:	Fehlerzustand in CASPIAN.....	121
Abbildung 27:	Fehlerzustand in CBR-Shell .....	121

Abbildung 28:	Graphical User Interface von CASPIAN.....	123
Abbildung 29:	Graphical User Interface von CBR-Shell .....	123
Abbildung 30:	Graphical User Interface von Induce-It .....	124
Abbildung 31:	Graphical User Interface von MyCBR als Protégé-Plug-In .....	124
Abbildung 32:	Graphical User Interface vom selbstständigen MyCBR.....	125
Abbildung 33:	Gesamtprioritäten in Expert Choice .....	132
Abbildung 34:	Darstellung der Klassenhierarchie.....	144
Abbildung 35:	Visualisierung der Klassenhierarchie in Protégé mit Jambalaya.....	146
Abbildung 36:	Definition von Attributen und Relationen in Protégé.....	149
Abbildung 37:	Attribute und Relationen der Ontologie .....	151
Abbildung 38:	Instanzen der Ontologie.....	158
Abbildung 39:	Beispiel zum Ähnlichkeitsmaßstab Taxonomie .....	161
Abbildung 40:	Bestimmung der Anzahl von Knoten im Beispiel 1 .....	163
Abbildung 41:	Bestimmung der gemeinsamen Attribute im Beispiel 1 .....	164
Abbildung 42:	Bestimmung der Anzahl von Knoten im Beispiel 2 .....	164
Abbildung 43:	Bestimmung der gemeinsamen Attribute im Beispiel 2.....	165
Abbildung 44:	Bestimmung der Anzahl von Knoten im Beispiel 3 .....	166
Abbildung 45:	Bestimmung der gemeinsamen Attribute im Beispiel 3 .....	166
Abbildung 46:	Ablauf des Algorithmus.....	173
Abbildung 47:	Taxonomie für das Attribut „Betriebssystemname“.....	178
Abbildung 48:	Taxonomie für das Attribut „Modell“ .....	182
Abbildung 49:	Entfernung von Instanzen zur gemeinsamen Oberklasse .....	186
Abbildung 50:	Ontologie mit hinterlegten Ähnlichkeitswerten für die Klassen .....	188
Abbildung 51:	Zuweisung von Ähnlichkeitswerten zu Klassen in Protégé .....	189
Abbildung 52:	Menüpunkt „Options...“ in Protégé .....	189
Abbildung 53:	Berücksichtigung unbekannter Attributswerte in Protégé.....	190
Abbildung 54:	Menüpunkt „Configure Special Values...“ in Protégé.....	190
Abbildung 55:	Umgang mit unbekanntem Attributs- und Relationswerten in Protégé.....	191
Abbildung 56:	Screenshot vom Retrieval-Ergebnis in MyCBR.....	195

---

Abbildung 57:	Positionen der Werte „Microsoft_Windows_XP“ und „Suse_Linux_10“ in der Taxonomie für das Attribut „Betriebssystemname“.....	199
Abbildung 58:	Positionen der Instanzen „Microsoft_Windows_XP“ und „Suse_Linux_10“ in der Ontologie (Auszug) .....	200
Abbildung 59:	Positionen der Instanzen „WSUS“ und „Webapplikation“ in der Ontologie (Auszug) .....	201
Abbildung 60:	Screenshot vom Retrieval-Ergebnis in MyCBR.....	205
Abbildung 61:	Positionen der Instanzen „Client_10“ und „Server_A2“ in der Ontologie (Auszug) .....	209
Abbildung 62:	Positionen der Werte „Red_Hat_Linux_9“ und „Microsoft_Windows_XP“ in der Taxonomie für das Attribut „Betriebssystemname“.....	210
Abbildung 63:	Positionen der Instanzen „Red_Hat_Linux_9“ und „Microsoft_Windows_XP“ in der Ontologie (Auszug) .....	211

**Tabellenverzeichnis**

Tabelle 1:	Relative Bedeutung zweier Beurteilungsobjekte für das übergeordnete Beurteilungsobjekt.....	61
Tabelle 2:	Ergebnisse der Paarvergleiche für die Kriterien der ersten Ebene.....	62
Tabelle 3:	Ergebnisse der Paarvergleiche für die Subkriterien der Funktionalität.....	64
Tabelle 4:	Ergebnisse der Paarvergleiche für die Subkriterien der Benutzbarkeit.....	65
Tabelle 5:	Inkonsistenzwerte der Evaluationsmatrizen.....	66
Tabelle 6:	Berechnung der Bedeutungsurteile für die Kriterien der ersten Ebene.....	68
Tabelle 7:	Beispiel für Ähnlichkeitswerte in einer Tabelle.....	74
Tabelle 8:	Ergebnisse der Paarvergleiche für die Kriterien der ersten Ebene.....	77
Tabelle 9:	Ergebnisse der Paarvergleiche für die Subkriterien der Funktionalität.....	78
Tabelle 10:	Ergebnisse der Paarvergleiche für die Subkriterien der Benutzbarkeit.....	79
Tabelle 11:	Ergebnisse der Paarvergleiche für die Subkriterien der Übertragbarkeit.....	79
Tabelle 12:	Inkonsistenzwerte der Evaluationsmatrizen.....	80
Tabelle 13:	Alternativen im AHP-Verfahren.....	88
Tabelle 14:	Bewertung der Tools hinsichtlich des Kriteriums Ontologierstellung.....	91
Tabelle 15:	Bewertung der Tools hinsichtlich des Kriteriums Konsistenzprüfung.....	93
Tabelle 16:	Interne Ontologiesprachen.....	95
Tabelle 17:	Bewertung der Tools hinsichtlich des Kriteriums Ontologiesprachen.....	95
Tabelle 18:	Bewertung der Tools hinsichtlich des Kriteriums Zuverlässigkeit.....	97
Tabelle 19:	Bewertung der Tools hinsichtlich des Kriteriums GUI.....	101
Tabelle 20:	Bewertung der Tools hinsichtlich des Kriteriums Dokumentation.....	103
Tabelle 21:	Ontologiesprachen beim Export.....	105
Tabelle 22:	Bewertung der Tools hinsichtlich des Kriteriums Übertragbarkeit.....	105

Tabelle 23:	Inkonsistenzwerte der Evaluationsmatrizen .....	107
Tabelle 24:	Berechnung der Bedeutungsurteile.....	108
Tabelle 25:	Einzelprioritäten in Expert Choice .....	110
Tabelle 26:	Änderung der Rangordnung beim Entfernen von Alternativen.....	112
Tabelle 27:	Alternativen im AHP-Verfahren .....	114
Tabelle 28:	Bewertung der Tools hinsichtlich des Kriteriums Algorithmen.....	117
Tabelle 29:	Bewertung der Tools hinsichtlich des Kriteriums Ähnlichkeitsmaßstab .....	119
Tabelle 30:	Bewertung der Tools hinsichtlich des Kriteriums Fallrepräsentation ...	120
Tabelle 31:	Bewertung der Tools hinsichtlich des Kriteriums Zuverlässigkeit .....	122
Tabelle 32:	Bewertung der Tools hinsichtlich des Kriteriums GUI.....	125
Tabelle 33:	Bewertung der Tools hinsichtlich des Kriteriums Dokumentation .....	127
Tabelle 34:	Bewertung der Tools hinsichtlich des Kriteriums Import von Fallwissen .....	128
Tabelle 35:	Bewertung der Tools hinsichtlich des Kriteriums Import von Ontologiewissen .....	129
Tabelle 36:	Inkonsistenzwerte der Evaluationsmatrizen .....	130
Tabelle 37:	Einzelprioritäten in Expert Choice .....	131
Tabelle 38:	Änderung der Rangordnung beim Entfernen von Alternativen.....	132
Tabelle 39:	Auszug der extrahierten Entitäten .....	143
Tabelle 40:	Attribute und Relationen der Ontologie .....	148
Tabelle 41:	Wertebereiche und Kardinalitäten der Attribute und Relationen .....	154
Tabelle 42:	Instanzen der Klasse „Projekt“ (Teil 1).....	155
Tabelle 43:	Instanzen der Klasse „Projekt“ (Teil 2).....	156
Tabelle 44:	Instanzen der Klasse „Betriebssystem“ .....	156
Tabelle 45:	Instanzen der Klasse „Client“.....	156
Tabelle 46:	Instanzen der Klasse „Server“ .....	157
Tabelle 47:	Instanzen der Klasse „Individualanwendung“ .....	157
Tabelle 48:	Instanzen der Klasse „Standardanwendung“ .....	157
Tabelle 49:	Beispiel zum Ähnlichkeitsmaßstab Tabelle .....	160
Tabelle 50:	Zuordnung der Ähnlichkeitsmaßstäbe.....	174
Tabelle 51:	Ähnlichkeitstabelle für das Attribut „Projekttyp“ .....	176

Tabelle 52:	Ähnlichkeitstabelle für das Attribut „Euro“ .....	176
Tabelle 53:	Ähnlichkeitstabelle für das Attribut „Personentage“ .....	177
Tabelle 54:	Ähnlichkeitstabelle für das Attribut „Familie“ .....	179
Tabelle 55:	Ähnlichkeitstabelle für das Attribut „Anwendungsname“ .....	180
Tabelle 56:	Ähnlichkeitstabelle für das Attribut „Anwendungsbereich“ .....	181
Tabelle 57:	Ähnlichkeitstabelle für das Attribut „Typ“ .....	183
Tabelle 58:	Gewichtung der lokalen Ähnlichkeiten .....	185
Tabelle 59:	Bestimmung von $k$ .....	187
Tabelle 60:	Bekannte Instanzen und Werte des Anfrage-Falls 1 .....	193
Tabelle 61:	Erweiterte Ähnlichkeitstabelle für das Attribut „Anwendungsname“ .....	194
Tabelle 62:	Retrieval-Ergebnis für Anfrage-Fall 1 gemäß MyCBR .....	195
Tabelle 63:	Gegenüberstellung der Werte von Anfrage-Fall 1 und vom Fall „Online-Banking“ .....	197
Tabelle 64:	Lokale Ähnlichkeitswerte für den Anfrage-Fall 1 und den Fall „Online-Banking“ .....	202
Tabelle 65:	Bekannte Instanzen und Werte des Anfrage-Falls 2 .....	203
Tabelle 66:	Erweiterte Ähnlichkeitstabelle für das Attribut „Anwendungsname“ .....	204
Tabelle 67:	Retrieval-Ergebnis für Anfrage-Fall 2 gemäß MyCBR .....	205
Tabelle 68:	Gegenüberstellung der Werte von Anfrage-Fall 1 und vom Fall „Online-Banking“ .....	207
Tabelle 69:	Lokale Ähnlichkeitswerte für den Anfrage-Fall 2 und den Fall „Optimierung_Clientstart“ .....	211
Tabelle 70:	Ergebnisse der Reliabilitätsprüfung zum Retrieval mit dem Anfrage-Fall 1 .....	214
Tabelle 71:	Ergebnisse der Reliabilitätsprüfung zum Retrieval mit dem Anfrage-Fall 2 .....	214

# 1 Einführung in den Problembereich technikgestützten Managements von Projektwissen

## 1.1 Betriebswirtschaftliches Desiderat

Die Bedeutung des Projektmanagements für unsere Gesellschaft hat in den letzten Jahren stark zugenommen.<sup>1</sup> Aktuelle Probleme im Projektmanagement, insbesondere das Scheitern von Projekten und eine mangelnde Wertsteigerung durch Projektarbeit, besitzen eine hohe Aufmerksamkeit. Laut einer Studie der TECHNISCHEN UNIVERSITÄT MÜNCHEN waren nur 43% der untersuchten Projekte erfolgreich.<sup>2</sup> Die allgemeine Abbruchquote von IT-Projekten wurde mit 20% gemessen.<sup>3</sup> In der aktuellsten Chaos-Studie der STANDISH GROUP<sup>4</sup> wurde für das Jahr 2008 sogar eine Abbruchquote in Höhe von 24% angegeben. Laut einer Studie von GRÖGER<sup>5</sup> trägt lediglich 13% der in der deutschen Wirtschaft geleiteten Projektarbeit zur Wertsteigerung bei. 87% dieser Projektarbeit ist Wertvernichtung. Folglich wurden innerhalb eines Jahres zwischen 112 und 194 Milliarden Euro im Rahmen von Projektarbeit verschwendet.

Neben dem Interesse an Projektmanagement nimmt auch das Interesse an Wissensmanagement seit einiger Zeit rapide zu.<sup>6</sup> Da es zwischen einem erfolgreichen Wissensmanagement und einem erfolgreichen Projektmanagement enge Beziehungen<sup>7</sup> gibt, liegt es nahe, nach neuen Verbesserungsansätzen zu suchen. Die engen Beziehungen zwischen einem erfolgreichen Wissensmanagement und einem erfolgreichen Projektmanagement füh-

---

1 Vgl. GPM (2007), S. 67.

2 Im Rahmen der Studie wurden Projekte aus der IT-Domäne über einen Zeitabschnitt von drei Jahren untersucht. Vgl. FRIEDMANN (2006), S. 1.

3 Vgl. RICHTER/BENDER/KLINGER/HERBOLZHEIMER (2008), S. 4.

4 Vgl. STANDISH GROUP (2009), S. 1.

5 Diese Angaben basieren auf einer Studie, in der über einen Zeitraum von fast vier Jahren 962 Führungskräfte zu ihren Erfahrungen mit dem Projektmanagement in deutschen Unternehmen und der öffentlichen Verwaltung befragt wurden. Vgl. GRÖGER (2004), S. 8 f.

6 „Die vielfachen wirtschaftspolitischen und soziologischen Debatten über den Übergang von der Industrie- zur Informations- oder Wissensgesellschaft mögen hierzu ebenso beigetragen haben wie die ökonomische Diskussion über die Erweiterung klassischer Faktorsystematiken um den Produktionsfaktor Wissen.“; ZELEWSKI (2005) S. 133.

7 Diese engen Beziehungen liegen darin begründet, dass Wissen ein primärer Einsatzfaktor für die Bearbeitung eines Projekts ist. Wenn das Wissensmanagement erfolgreich ist, steht Wissen tendenziell in umfangreicherem und qualitativ höherem Maß bereit. Wissen ist als Voraussetzung für eine effektive und effiziente Projektbearbeitung zu betrachten. Maßstäbe hierfür sind die Qualität des Projektergebnisses sowie Zeit und Kosten als Projekt-Inputs.

## 2 Einführung in den Problembereich technikgestützten Managements von Projektwissen

ren dazu, dass sowohl die Disziplin „Projektmanagement“ als auch die Disziplin „Wissensmanagement“ von grundlegenden Aufgabenbereichen tangiert werden. Das Spannungsfeld interdisziplinärer Aufgabenbereiche ist daher ein wichtiger Ansatzpunkt für neue Verbesserungen. Die Wiederverwendung von Projektwissen ist einer dieser interdisziplinären Aufgabenbereiche.

Bei der Vorbereitung und Durchführung neuer Projekte ist die Wiederverwendung von unternehmensspezifischem Wissen aus bereits durchgeführten Projekten sehr hilfreich, um z.B. Projekt-Erwartungen realistisch zu bewerten und passende Projekt-Anforderungen zu spezifizieren. Laut einer Studie der DEUTSCHEN GESELLSCHAFT FÜR PROJEKTMANAGEMENT<sup>8</sup> ist das Fehlen einer Wiederverwendung von diesem Wissen eine der größten Schwachstellen beim Projektmanagement. Wenn z.B. Wissen über Budgetabschätzungen und Zeitplänen aus bereits durchgeführten Projekten nicht wiederverwendet wird, können wiederholte Fehler schwer vermieden werden.

Das Wissen ist hauptsächlich in den Erinnerungen und Unterlagen der Akteure gespeichert. Im Rahmen von Dokumentationen und durch die Generierung von Lessons Learned werden Dokumente erstellt, die elektronisch oder papierbasiert im Unternehmen aufbewahrt werden. Daraus entsteht eine unüberschaubare Menge an Dokumenten, die sich mit traditionellen Techniken<sup>9</sup> nur noch mit großem Aufwand verwalten lässt. Das große Volumen ist eine Barriere bei der Wiederverwendung des Wissens.

Studien belegen, dass 80% der Informationen eines Unternehmens in Textdokumenten enthalten sind.<sup>10</sup> Daraus folgt, dass Informationen hauptsächlich in Form natürlichsprachlicher Dokumente vorliegen. Die natürlichsprachliche Repräsentation des Wissens ist eine weitere Barriere bei der Wiederverwendung des Wissens.

Konventionelle Wissensmanagementtechniken der betrieblichen Praxis reichen zur Überwindung der Barrieren nicht aus, da in der Praxis die relevanten Informationen meist

---

8 Vgl. ENGEL/TAMDJIDI/QUADEJACOB (2008), S. 5.

9 Unter traditionellen Techniken werden die manuelle Bearbeitung und die in der Praxis verbreitete teilautomatische Bearbeitung mit Hilfe von computergestützten Techniken verstanden, wie z.B. Abfragen von Datenbanken oder Volltextsuchen in Dateiverzeichnissen.

10 Vgl. TAN (1999), S. 1. Obwohl die 80%-Regel vielfach zitiert wird, ist sie durchaus kritisch zu betrachten. Je nach Studie gibt es Abweichungen im gemessenen Verhältnis von unstrukturierten Daten, wie Textdokumenten, zu strukturierten Daten. Vgl. GRIMES (2010). Im Rahmen verschiedener Studien wurde der Anteil der unstrukturierten Daten mit 53% bis 90% angegeben. Für das Verhältnis spielt außerdem die Arbeitsweise im betrachteten Unternehmen eine große Rolle. So wird ein Unternehmen, in dem hauptsächlich mit strukturierten Daten gearbeitet wird, einen höheren Anteil an strukturierten Daten besitzen als ein Unternehmen, in dem hauptsächlich mit unstrukturierten Daten gearbeitet wird.

manuell aus elektronischen Informationsspeichern<sup>11</sup> recherchiert oder mit erfahrenen Akteuren kommuniziert werden. Dadurch ergeben sich mehrere Probleme für ein effektives Auffinden von passenden Informationen.

Zum einen sind die Dokumente von unerfahrenen Akteuren schwer zu durchschauen, da anhand der Dokumente das relevante Wissen meist nur implizit ersichtlich und somit auch nicht sofort verfügbar ist. Auch lassen sich viele Informationen nicht ohne Hintergrundwissen aus anderen Texten erschließen. Erschwerend ist außerdem, dass verschiedene Akteure oft auch verschiedene Begriffe benutzen, um dieselben Informationen zu speichern. Auch sind die Beziehungen der Projekte untereinander nicht sofort erkennbar. Z.B. könnte ein abgeschlossenes Projekt nur auf Grund einer starken Synergienutzung mit anderen Projekten erfolgreich gewesen sein. Wenn auch die Dokumente von diesen anderen Projekten gesucht und verstanden werden müssen, führt dies dazu, dass für die Erkennung von Ähnlichkeiten zwischen bereits durchgeführten und aktuellen Projekten eine erhöhte Anzahl von Dokumenten mit natürlichsprachlichem Text strukturiert werden muss. Eine Strukturierung von natürlichsprachlichem Text ist grundsätzlich schwer.<sup>12</sup>

Zum anderen ist das Wissen über Ähnlichkeitsmerkmale zwischen Projekten in großem Umfang in den Erinnerungen erfahrener Akteure gespeichert. Solange dieses Ähnlichkeitswissen nicht extern gespeichert wird, kann eine langfristige Verfügbarkeit des Wissens nicht sichergestellt werden. Wissen kann für das Unternehmen verloren gehen, wenn betreffende Akteure ausscheiden. Wissen kann temporär unzugänglich sein, wenn Akteure nicht kooperieren wollen oder aus Zeitmangel nicht können.<sup>13</sup>

Das betriebswirtschaftliche Desiderat besteht aus einer im Vergleich zu manuellen Techniken kostengünstigeren, zeitlich weniger aufwändigen und qualitativ hochwertigeren Identifikation von hinreichend ähnlichen Fällen und Selektion von einem ähnlichsten Fall im Zuge der Wiederverwendung von Erfahrungswissen aus Projekten. Ein Fall repräsentiert hierbei ein Projekt.

Relevante Kosten sind der Ressourceneinsatz von Projektwissen, Sachgütern und Dienstleistungen. Das Ziel der Ressourceneinsparung wird dadurch erreicht, dass der Res-

---

11 Bei elektronischen Informationsspeichern handelt es sich beispielsweise um Dokumenten-Management-Systeme, die die elektronische Speicherung von Dokumenten ermöglichen. Die aus diesen Systemen abrufbaren Dokumente beinhalten Informationen, die in Form von natürlich- oder formal-sprachlichen Elementen repräsentiert werden.

12 Vgl. WOOD/ROSS-KERR (2010), S. 250 u. MOENS (2006), S. 4. Hier wird natürlichsprachlicher Text zur Kategorie der unstrukturierten Daten gezählt.

13 Vgl. PUPPE/STOYAN/STUDER (2004), S. 621.

sourceneinsatz von Projektwissen, Sachgütern und Dienstleistungen geringer ausfällt. Eine neue Technik soll einen geringeren Ressourceneinsatz verursachen als eine herkömmliche manuelle oder automatische Technik. Der Ressourceneinsatz besteht bei einer manuellen Technik primär aus Arbeitskraft.

Zeiteinsparungen werden durch Prozessbeschleunigungen erreicht. Die benötigte Zeit zur Durchführung des Prozesses zur Identifikation von hinreichend ähnlichen Fällen und Selektion eines ähnlichsten Falls kann durch eine Änderung des Prozessinhalts reduziert werden, der aus den erforderlichen Tätigkeiten der am Prozess beteiligten Akteure besteht. Nicht allein das Vorhandensein von Projektwissen, sondern vielmehr ein möglichst schneller und einfacher Zugriff darauf sind entscheidend für Wiederverwendung von Projektwissen.

Eine Verbesserung der Qualität bedeutet, dass die Merkmale der Fallidentifikation und -selektion dazu geeignet sind, festgelegte Erfordernisse zu erfüllen. Diese Erfordernisse sind die umfangreiche Einbindung von Hintergrundwissen und die Aufhebung der Begrenzung auf das eigene Hintergrundwissen einzelner Akteure. Denn es soll nicht nur das eigene Hintergrundwissen einzelner Akteure, sondern eine gemeinsame Wissensbasis genutzt werden können.

### **1.2 State-of-the-art der verfügbaren Techniken zur Erfüllung des betriebswirtschaftlichen Desiderats**

Zur Erfüllung des vorgenannten betriebswirtschaftlichen Desiderats werden seitens des State-of-the-arts der automatischen Informationsverarbeitung und des computergestützten Wissensmanagements vor allem die Techniken des Information Retrievals (IR) und Natural Language Processings (NLP) angeboten.<sup>14</sup>

Die Bedeutung des Begriffs Information Retrieval ist weit gefasst. Eine allgemeine Definition stammt von MANNING<sup>15</sup>, der Information Retrieval als das Finden von unstrukturiertem Material (meistens Dokumente, die Text beinhalten) innerhalb von großen Informationsmengen beschreibt, um einen Informationsbedarf zu befriedigen. Aus den folgenden Gründen ist Information Retrieval für den Untersuchungsgegenstand ungeeignet: Information Retrieval basiert hauptsächlich auf Statistiken über Dokumentensammlungen. Es wird sehr schwierig, innerhalb dieser Technik benutzerdefinierte Ähnlichkeitsmaßstäbe

---

14 Vgl. z.B. VOORHEES (1999), S. 32 f. u. SMEATON (1999), S. 99 ff.

15 Vgl. MANNING (2008), S. 1.

mit einem vernünftigen zeitlichen und kostenmäßigen Aufwand zu integrieren. Außerdem basiert Ähnlichkeit in den meisten Techniken des Information Retrievals hauptsächlich auf Zeichenfolgen. Es ist nicht geklärt, wie domänenspezifisches Wissen, z.B. domänenspezifische Synonyme und Abkürzungen genutzt werden können. Die Verwendung von zusätzlichem Hintergrundwissen, wie z.B. die Gewichtung von Ähnlichkeiten zwischen Attributswerten, ist beim Information Retrieval nicht möglich.<sup>16</sup>

Laut MANARIS<sup>17</sup> befasst sich Natural Language Processing mit den sprachlichen Aspekten der Kommunikation zwischen menschlichen Akteuren oder zwischen menschlichen und maschinellen Akteuren sowie der Integration dieser Aspekte in Techniken und Prozesse. Natural Language Processing umfasst Techniken, die den Inhalt von Texten verstehen sollen. Zu diesen Techniken gehören z.B. Spracherkennung, Schnittstellen für natürliche Sprache, Diskursmanagement und interaktive Maschinenübersetzung.<sup>18</sup> Diese Techniken beschränken sich nicht auf rein statistische Verfahren, wie z.B. die Ermittlung von Worthäufigkeiten bei Techniken des Information Retrievals, sondern versuchen die Semantik eines Satzes anhand der Identifikation von Verben, Substantiven usw. herauszufinden. Die Techniken des Natural Language Processings teilen einen Satz mit Hilfe von verschiedenen Algorithmen in seine Hauptbestandteile. Die Algorithmen führen Analysen auf unterschiedlichen Ebenen durch. Die Analysen können oberflächlich oder tiefgehend sein.<sup>19</sup> Oberflächliche Analysen berücksichtigen ausschließlich die Syntax des Satzes und nicht die Semantik. Sie ermitteln z.B. Verben, Substantive usw., z.B. mit Part-of-Speech-Tagging<sup>20</sup>. Tiefgehende Analysen gehen einen Schritt weiter und versuchen, die Semantik eines Satzes zu deuten. Durch die Offenlegung der Semantik eines Satzes kann der Kontext von Wörtern berücksichtigt werden. Ein negativer Aspekt der tiefgehenden Analysen ist, dass sie oft sehr rechenintensiv sind.<sup>21</sup> Außerdem arbeiten sie meist mit kleinen Wörterbüchern und sind nicht robust, wenn unbekannte Begriffe auftauchen.<sup>22</sup> Im Kontext der Untersuchung ist dies sehr wahrscheinlich, da domänenspezifische Begriffe aus der Domäne

---

16 Vgl. WATSON (1997), S. 45.

17 Vgl. MANARIS (1998), S. 6.

18 Vgl. MANARIS (1998), S. 3.

19 Vgl. ACHANANUPARP/HU/ZHOU/ZHANG (2008), S. 204.

20 Unter Part-of-Speech-Tagging ist die Zuordnung von Wörtern eines Textes zu Wortarten zu verstehen.

21 Vgl. MOSCHITTI/BASILI (2004), S. 182 u. LENZ/HÜBNER/KUNZE (1998), S. 246.

22 Vgl. LENZ/HÜBNER/KUNZE (1998), S. 246.