



Otto-Friedrich-Universität Bamberg
Lehrstuhl für Praktische Informatik



Bachelor

im Studiengang Wirtschaftsinformatik
der Fakultät Wirtschaftsinformatik und Angewandte Informatik
der Otto-Friedrich-Universität Bamberg

Zum Thema:

Automatisierung der Datenerfassung für Wissensdatenbanken im technischen Kontext

Vorgelegt von:
Vasilyev Petr

Themensteller:
Prof. Dr. Guido Wirtz

Abgabedatum:
31.03.2017

Inhaltsverzeichnis

1	Einführung	1
1.1	Motivation	1
1.2	Zielsetzung	1
1.3	Aufbau der Arbeit	2
2	Grundlagen von Expertensystemen	3
2.1	Begriffsdefinition	3
2.2	Architektur eines Expertensystems	5
2.3	Wissensbasis	7
2.4	Wissensakquisitionskomponente	9
3	Automatisierung der Datenerfassung	12
3.1	Die Wissensträgerschnittstelle	12
3.2	Maschineller Wissenserwerb	13
3.3	Schnittstelle für Validierung und Speicherung	13
4	Implemetierung	14
4.1	Ansatz 1	14
4.2	Ansatz 2	14
4.3	Ansatz 3	14
5	Evaluation	14
5.1	Test 1	14
5.2	Test 2	14
5.3	Test 3	14
6	Fazit	14
7	Ausblick	14
	Literaturverzeichnis	15

Abbildungsverzeichnis

1	Begriffsabgrenzung, [Mat00, S.30]	4
2	Expertensystems nach Haun, [Mat00, S.126]	5
3	Phasen der Expertensystementwicklung, [GD94, S.138]	7
4	Expertensystems nach Beierle und Kern-Isberner, [CG14, S.18]	8
5	Wissensakquisitionskomponente	11

Abkürzungsverzeichnis

IaaS Infrastructure-as-a-Service

PaaS Platform-as-a-Service

SaaS Software-as-a-Service

DSS Decision Support System

1 Einführung

1.1 Motivation

Das Themengebiet von Cloud Computing, das aus den Servicemodellen Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) und Software-as-a-Service (SaaS) besteht [PT11], gehört zu den am meisten versprechenden Themen der letzten Jahre. Insbesondere PaaS ist vom großen Interesse für die zukunftsorientierte Anwendungsentwicklung. Der Hauptvorteil für die Entwickler besteht darin, dass die Serverkonfiguration vom PaaS-Anbieter durchgeführt wird. Dies ermöglicht die Effizienzsteigerung der Anwendungsentwicklung, da die Programmierer mehr Zeit für den eigentlichen Entwicklungsprozess haben [SG14, S.218]. Die Reduzierung der Entwicklungskosten gehört laut George Lawton ebenso zu den Vorteilen von PaaS [Geo08, S.14].

Momentan befindet sich der PaaS-Markt in einer Wachstumsphase, die durch eine hohe Anzahl von PaaS-Anbietern gekennzeichnet ist. In [SG14] werden bereits 68 PaaS-Anbieter erwähnt. Darüber hinaus bieten verschiedene PaaS-Dienstleister ein unterschiedliches Angebot an Laufzeitumgebungen, Kapazitäten und Systemkonfigurationen. Um einen Überblick über das vorhandene PaaS-Angebot zu schaffen, wurde ein Expertensystem namens *PaaSfinder*¹ entwickelt. Dabei handelt es sich um eine Web-Anwendung, die dem Nutzer die Suche nach der PaaS-Plattform erleichtern soll. Zu den wichtigsten Funktionen von *PaaSfinder* zählen Filtersuche mittels Merkmalen der PaaS-Plattform und 1-zu-1 Vergleich zwischen den zwei PaaS-Anbietern. *PaaSfinder* ist ein Open Source Projekt und steht frei für die Nutzung oder Verbesserungsvorschlägen zur Verfügung.

Da sich der PaaS-Bereich sehr schnell entwickelt, besteht für *PaaSfinder* das Problem der Aktualität der vorliegenden Daten. Ein PaaS-Anbieter kann beispielsweise einen neuen Service zur Plattform hinzufügen oder Informationen im Bezug auf bestehende Dienstleistungen ändern. Aus diesem Grund muss die Wissensbasis von *PaaSfinder* regelmäßig aktualisiert und gepflegt werden, um konsistent, vollständig und aktuell zu bleiben. Momentan werden die Daten für *PaaSfinder* manuell erfasst und aktualisiert. Dies ist vor allem zeitaufwändig und fehleranfällig. Außerdem kann ein wichtiges Update leicht übersehen werden.

1.2 Zielsetzung

Das Ziel dieser Bachelorarbeit ist die Entwicklung eines Konzeptes zur Automatisierung der Datenerfassung für Wissensdatenbanken. Dabei ist die Sicherstellung von Aktualität, Konsistenz und Fehlerfreiheit einer Wissensdatenbank von Bedeutung. Die praktische Umsetzung soll am Beispiel von *PaaSfinder* erfolgen.

Sébastien Gebus betont, dass ein System mit generischen Wissenserfassungsmethoden für einen konkreten Anwendungsfall kaum einsetzbar ist [SK09, S.100]. Aus diesem Grund beschränkt sich die vorliegende Arbeit auf die Auswahl der Methoden, die für *PaaSfinder* und dazugehörige Wissensbasis am meisten geeignet sind.

Ein weiteres Problem ist der Automatisierungsgrad der Wissenserfassungsmethoden. In mehreren Studien hat sich herausgestellt, dass ein halb-automatisierter Ansatz für die Wissenserfassung am meisten geeignet ist [HDNR97], [SK09], [Ghe92]. So sind Hiroko

¹<https://paasfinder.org>

Fujihara et al. zum Schluss gekommen, dass Wissenserfassungsprozess primär zu dem Aufgabenbereich eines Wissensingenieurs gehört. Die Methoden, die auf automatisierte Weise Daten erfassen, sollen dabei nur als Unterstützung dienen, um die Produktivität und Effizienz des Fachexperten zu erhöhen [HDNR97, S.219]. Daher liegt der Fokus dieser Arbeit auf einen halb-automatisierten Ansatz, der automatisierte Datenerfassung und manuelle Kontrolle durch einen Experten kombinieren soll.

1.3 Aufbau der Arbeit

Die vorliegende Arbeit ist folgendermaßen aufgeteilt. In Kapitel 2 wird der Begriff eines Expertensystems und das grundlegende Konzept der Datenerfassung erläutert. Anschließend wird definiert, welche Informationen für die Wissensbasis von *PaaSfinder* notwendig sind. In Kapitel 3 handelt es sich um die Auswahl und Beschreibung geeigneter Wissenserfassungsmethoden im Hinblick auf *PaaSfinder*. Die konkrete Umsetzung der ausgewählten Methoden wird in Kapitel 4 beschrieben. In Kapitel 5 werden die implementierten Methoden evaluiert. Dabei wird die Nützlichkeit der Methoden bestimmt, in dem die gewonnenen Daten auf die Qualität ausgewertet werden. In Kapitel 6 handelt es sich um die Zusammenfassung des Endergebnis der durchgeführten Arbeit. Kapitel 7 gibt eine mögliche Richtung für zukünftige Forschung im betrachteten Bereich.

2 Grundlagen von Expertensystemen

2.1 Begriffsdefinition

Ursprünglich waren Expertensysteme Anwendungsprogramme, die logische Schlussfolgerungen aus einer Wissensbasis ziehen konnten. Außerdem konnten sie überprüfen, ob eine Aussage aus einer vorhandenen Wissensbasis abgeleitet werden kann [Pet10, S.75]. Daher handelt es sich in der früheren Literatur meist um Anwendungen, die ihr Wissen in Form von logischen Ausdrücken darstellen und in der Lage waren, neue Erkenntnisse vom bestehenden Wissen abzuleiten [Ghe92]. Im Laufe der Zeit hat sich das Konzept eines Expertensystems auf andere Anwendungsbereiche verbreitet. Aus diesem Grund gibt es mehrere Definitionen, die im Allgemeinen ähnlich sind und im Spezifischen Merkmale des zugehörigen Anwendungsbereichs beinhalten.

Allgemein lässt sich sagen, dass ein Expertensystem ein Computersystem (Hardware und Software) ist, das in einem bestimmten Bereich Wissen und Schlussfolgerungsfähigkeit eines menschlichen Experten nachbildet [CG14, S.12]. Aus Sicht der Wirtschaftsinformatik zielen Expertensysteme drauf ab, das Expertenwissen menschlicher Fachleute in der Wissensbasis eines Computers abzuspeichern und für eine Vielzahl von Problemlösungen zu nutzen [PFW⁺12, S.59]. Im Weiteren gehen Beierle und Kern-Isberner auf die Eigenschaften ein, die ein Expertensystem aufweisen soll [CG14, S.12]. Im Rahmen dieser Arbeit sind folgende Eigenschaften besonders relevant:

- Anwendung des Wissens eines oder mehrerer Experten, um Probleme in einem bestimmten Anwendungsbereich zu lösen.
- Leicht lesbare Wissensdarstellung.
- Möglichst anschauliche und intuitive Benutzerschnittstelle.
- Leichte Wartbarkeit und Erweiterbarkeit des Wissens im Expertensystem.
- Unterstützung beim Wissenstransfer vom Experten zum System.

Es ist auch wichtig anzumerken, dass die Begriffe “Künstliche Intelligenz“, “Wissensbasiertes System“ und “Expertensystem“ in einer engen Beziehung zueinander stehen. Haun gibt eine systematische Abgrenzung dieser Begriffe, die sich folgendermaßen beschreiben lässt [Mat00, S.30]:

- *Künstliche Intelligenz* stellt den Oberbegriff dar und bildet den theoretischen Rahmen für die Entwicklung von Wissensbasierten Systemen und Expertensystemen.
- *Wissensbasierte Systeme* sind eine Teilmenge der Anwendungen der künstlichen Intelligenz. Sie wenden die Wissensverarbeitung auf ein konkretes Aufgabengebiet an und verwalten allgemeines Wissen explizit und getrennt von dem Rest des Systems.
- *Expertensysteme*, die ein Teilbereich der Wissensbasierten Systeme sind, stellen eine Spezialisierung von Wissensbasierten Systemen dar. Sie verwalten spezifisches Expertenwissen, das von einem Experten stammt und auf praxisbezogene Probleme angewandt wird.

Graphisch lässt sich die vorliegende Abgrenzung in der Abbildung 1 darstellen:



Abbildung 1: Begriffsabgrenzung, [Mat00, S.30]

Nach dieser Abgrenzung lässt sich feststellen, dass der Unterschied zwischen einem Wissensbasierten System und einem Expertensystem darin besteht, dass das Wissen im Endeffekt von einem Experten stammt. Allerdings ist dieses Kriterium zu einfach und nicht besonders aussagekräftig. Beierle und Kern-Isberner weisen drauf hin, dass nach diesem Kriterium viele existierenden Wissensbasierten Systeme als Expertensysteme bezeichnet werden könnten [CG14][S.11]. Aufgrund dessen stellen die Autoren die Eigenschaften dar, die ein Experte in der Regel besitzt. Im Hinblick auf das Thema der vorliegenden Arbeit sind folgende Merkmale vom Interesse:

- Experten sind selten und teuer.
- Experten sind nicht immer verfügbar.
- Leistungsfähigkeit der Experten ist nicht konstant, sondern kann nach Tagesverlauf schwanken.
- Expertenwissen kann oft nicht als solches weitergegeben werden.
- Expertenwissen kann verloren gehen.

Ein gutes Beispiel hinsichtlich der Gefahr, dass Expertenwissen verloren gehen kann, wird in [SK09][S.94] vorgestellt. Gebus nimmt den Bezug auf die sogenannte Baby-Boomgeneration, die bald in die Rente geht. Infolgedessen verlassen gleichzeitig viele Experten Unternehmen. Damit geht ebenso das Expertenwissen verloren.

Zusammenfassend lässt sich sagen, dass die Entwicklung eines Expertensystems ein hohes Potenzial besitzt. Allerdings kann ein Expertensystem nicht als Ersatz eines menschlichen Experten betrachtet werden. Vielmehr geht es um eine Erfassung, Darstellung und Pflege des Expertenwissens in einem Expertensystem, um die Arbeitsprozesse effizienter zu gestalten und sowohl erfahrene als auch neue Anwender in einem bestimmten Wissensbereich bei der Aufgabenabwicklung zu unterstützen.

2.2 Architektur eines Expertensystems

Beierle und Kern-Isberner betonen, dass die Trennung zwischen der Darstellung des Wissens (Wissensbasis) und der Wissensverarbeitung (Wissensverarbeitungskomponente) der wichtigste Aspekt eines Wissensbasierten Systems ist. [CG14, S.11]. Die Wissensbasis kann man sich als eine Art Datenstruktur vorstellen, in der das benötigte Wissen gespeichert wird. Die Wissensverarbeitungskomponente umfasst eine Menge von anwendungsunabhängigen Algorithmen, die mithilfe der Wissensbasis eine Lösung für ein gegebenes Problem erarbeiten. Somit stehen die Wissensbasis und die Wissensverarbeitungskomponente in einer engen Beziehung zueinander.

Allgemein umfasst ein Expertensystem folgende Bestandteile [Pet10, S.75]:

- *Wissensbasis*, die Expertenwissen in Form von Fakten in einer bestimmten Sprache speichert sowie Regeln zur Wissensorganisation beinhaltet.
- *Inferenzmaschine*, die unter Berücksichtigung des zugrunde liegenden Wissensbedarfs die Wissensbasis durchsucht bis das System einen Problemlösungsvorschlag erarbeitet hat oder herausfindet, dass keiner existiert.
- *Dialogkomponente*, die eine Schnittstelle zwischen dem Nutzer und dem System darstellt.
- *Erklärungskomponente*, die dem Benutzer erläutert, warum und auf welche Weise eine bestimmte Lösung gefunden bzw. nicht gefunden wurde [Mat00, S.126].
- *Wissensakquisitionskomponente*, die den Entwickler des Expertensystems bei der Erweiterung, Änderung und Wartung der Wissensbasis unterstützt.

Laut Tecuci stellen Wissensbasis und Inferenzmaschine grundlegende Bestandteile eines Expertensystems dar und bilden damit den Kern des Expertensystems [Ghe92, S.1444]. Dialogkomponente, Erklärungskomponente und Wissensakquisitionskomponente gehören zur sogenannten Schale und sind für die Kommunikation zwischen dem Systemverwalter und dem Nutzer zuständig (siehe Abbildung 2).

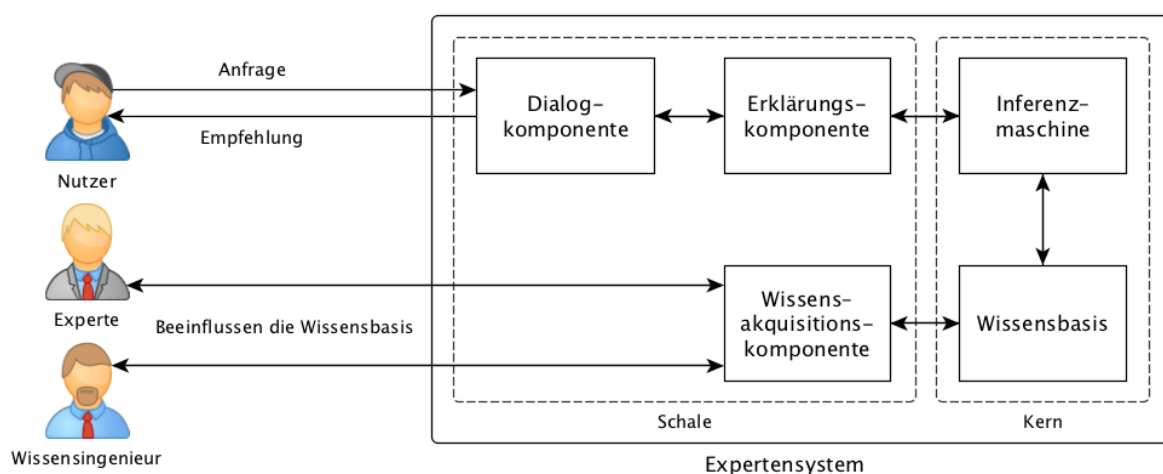


Abbildung 2: Expertensystems nach Haun, [Mat00, S.126]

Im Hinblick auf die Interaktion gibt es drei Gruppen, die mit dem Expertensystem interagieren:

- *Nutzer*, der das Expertensystem zum Lösen eines Problems benutzt und mit der Dialogkomponente kommuniziert. Der Wissensingenieur und der Experte können ebenso als Nutzer auftreten [JFI93, S.758].
- *Wissensingenieur*, der sich mit dem Aufbau und Wartung der Wissensbasis beschäftigt. Unter anderem ist Wissensmodellierung ein wichtiger Aufgabenbereich eines Wissensingenieurs [JFI93, S.742].
- *Experte*, der über spezifisches Erfahrungswissen verfügt, das für das Expertensystem relevant ist.

Der Ablauf der Kommunikation zwischen dem Nutzer und dem Expertensystem sieht folgendermaßen aus:

- Der Nutzer schickt eine Anfrage an die Dialogkomponente des Expertensystems.
- Die Dialogkomponente übermittelt die Anfrage an die Inferenzmaschine.
- Die Inferenzmaschine erarbeitet eine Lösung für das gegebene Problem mittels der Wissensbasis und gibt das Ergebnis an die Dialogkomponente zurück.
- Anschließend teilt die Dialogkomponente dem Nutzer die Lösung des Problems mit. Falls keine Lösung zum Problem existiert, wird eine entsprechende Fehlermeldung angezeigt.

Auf der anderen Seite können die Inhalte der Wissensbasis von einem Wissensingenieur mithilfe der Wissensakquisitionskomponente beeinflusst werden. Der Wissenserwerb durch den Wissensingenieur ist die verbreitetste Vorgehensweise, neue Daten für ein wissensbasiertes System zu erschließen. Meistens handelt es sich um ein Interview zwischen dem Wissensingenieur und dem Experten [Pet10, S.76], [HDNR97, S.210]. Neben dem Interview kann der Wissensingenieur eine Recherche der verfügbaren Wissensquellen wie Text, technische Zeichnungen oder Web-Ressourcen durchführen. Anschließend werden die Daten vom Wissensingenieur formalisiert und in die Wissensbasis gespeichert.

Die Wissensbasis kann in einigen Fällen von einem fachlichen Experten beeinflusst werden. Dafür ist eine geeignete Expertenschnittstelle innerhalb der Wissensakquisitionskomponente notwendig, die den Experten ermöglicht, ihr Erfahrungswissen selbst zu formalisieren und gegebenenfalls zu warten [JFI93, S.743]. Die Überprüfung des Dateninputs ist ebenfalls die Aufgabe der Wissensakquisitionskomponente. Dies kann mittels Durchführung automatisierten Tests bei jeder Änderungsanfrage erfolgen, um die Konsistenz der Wissensbasis zu gewährleisten [JFI93, S.743].

Um ein geeignetes Konzept der automatisierten Datenerfassung zu entwickeln, ist ein grundlegendes Verständnis von der Struktur und Funktionsweise der Wissensbasis sowie der Wissensakquisitionskomponente erforderlich. Im weiteren Verlauf der Arbeit werden die Erkenntnisse über die Wissensbasis und die Wissensakquisitionskomponente erläutert, die in der Forschung von Expertensystemen entstanden sind.

2.3 Wissensbasis

Neben der Inferenzmaschine stellt die Wissensbasis den zentralen Teil eines Expertensystems, der die Daten des gesamten Systems beinhaltet [JFI93, S.754]. Im Folgenden werden der allgemeine Prozess der Wissensbasisentwicklung, der Inhalt der Wissensbasis und die Möglichkeiten der Wissensrepräsentation thematisiert. Gheorghe Tecuci beschreibt folgende Phasen bei der Entwicklung der Wissensbasis [Ghe92, S.1444]:

1. Systematische Erfassung vom Expertenwissen
2. Verfeinerung der Wissensbasis
3. Reorganisation der Wissensbasis

In der ersten Phase werden das Vokabular und die geeignete Wissensrepräsentation festgelegt. Gebus und Leiviskä betonen, dass die Wissensrepräsentation entscheidenden Einfluss auf die Generierung und spätere Handhabung der Wissensbasis hat [SK09, S.95]. Die initialen Daten werden meistens im Rahmen eines Interviews zwischen dem Wissensingenieur und dem Experten erfasst [Ghe92, S.1444]. Das Ergebnis der ersten Phase ist eine initiale Wissensbasis, die unvollständig und teilweise widersprüchlich ist. In der zweiten Phase wird die initiale Wissensbasis mithilfe der geeigneten Datenerfassungsmethoden solange erweitert und verbessert, bis sie vollständig und korrekt genug ist, um ein gegebenes Problem richtig zu lösen. In der dritten Phase wird die vollständige und korrekte Wissensbasis reorganisiert, um die Effizienz der Lösungsberechnung zu steigern [Ghe92, S.1445]. Zusammenfassend lassen sich die Phasen in der Abbildung 3 darstellen:

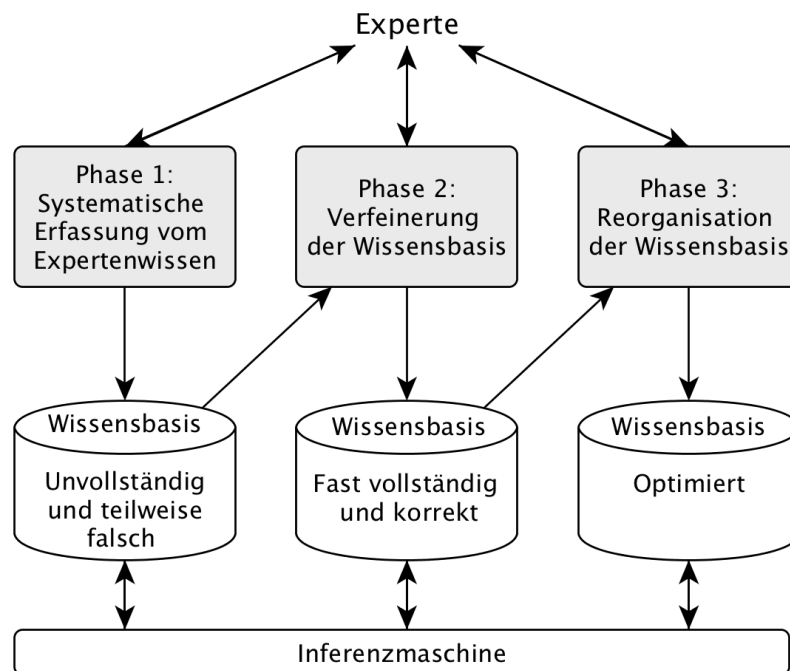


Abbildung 3: Phasen der Expertensystementwicklung, [GD94, S.138]

In Bezug auf den Inhalt der Wissensbasis unterscheiden Beierle und Kern-Isberner folgende Wissensarten [CG14, S.5]:

- *Fachspezifisches Wissen*. Dabei handelt es sich um das spezifischste Wissen, das sich nur auf den gerade betrachteten Problemfall bezieht. Das sind z.B. Fakten, die von Beobachtungen oder Untersuchungsergebnissen stammen.
- *Regelhaftes Wissen*, das den eigentlichen Kern der Wissensbasis darstellt. Dieses Wissen kann noch genauer differenziert werden:
 - *Bereichsbezogenes Wissen*, das sich auf den gesamten Problembereich beziehen. Das kann sowohl theoretisches Fachwissen als auch Erfahrungswissen sein. Anders gesagt handelt es sich um generisches Wissen.
 - *Allgemeinwissen*, das z.B. um generelle Problemlösungsheuristiken, Optimierungsregeln oder auch allgemeines Wissen über Objekte und Beziehungen in der realen Welt beinhaltet.

Unter Berücksichtigung der Differenzierung der Wissensarten innerhalb der Wissensbasis beschreiben die Autoren in [CG14, S.18] auf eigene Weise die Architektur des Expertensystems, die in der Abbildung 4 dargestellt wird.

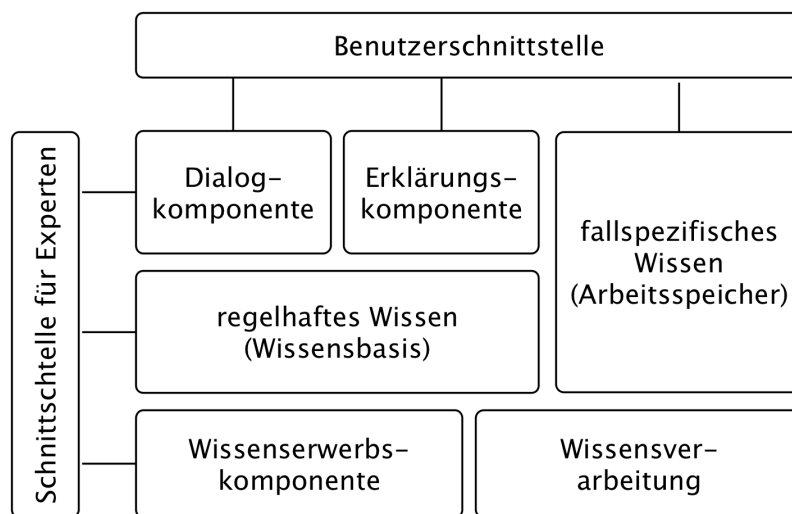


Abbildung 4: Expertensystems nach Beierle und Kern-Isberner, [CG14, S.18]

Laut Beierle und Kern-Isberner können verschiedene Wissensarten in einem wissensbasierten System je nach dem Anwendungsbereich unterschiedlich umfangreich auftreten. Ein hochspezialisiertes System kann beispielsweise über sehr wenig oder gar kein Allgemeinwissen verfügen. Auf der anderen Seite kann ein wissensbasiertes System den Schwerpunkt auf das gewöhnliche Alltagswissen legen [CG14, S.5-6].

Ein weiterer Aspekt beim Aufbau der Wissensbasis ist die Wissensrepräsentation. Die grundlegende Aufgabe der Wissensrepräsentation ist die Formularisierung von Wissen, um eine maschinelle Verarbeitung erst zu ermöglichen [Mat00, S.22]. Sinz und Ferstl unterscheiden folgende Formen der Wissensrepräsentation [OE13, S.366]:

- *Regelorientierte Darstellung*, in der das Wissen als (WENN, DANN)-Beziehungen beschrieben wird. Diese Darstellungsform wird beispielsweise bei Prolog-Regeln eingesetzt.
- *Objektorientierte Darstellung*, die das Konzept der Objekttypen übernimmt und mit deklarativen Operatorbeschreibungen verbindet.
- *Constraints Darstellung*, die Modellbeschreibungen aus dem Operations Research benutzt. Dabei handelt es sich um Lösungsräume durch Nebenbedingungen und Zielvorgaben.

Hinsichtlich der Wissensrepräsentation stellen Ferstl und Sinz imperative und deklarative Paradigmen gegenüber [OE13, S.366]. Ein Programm, das dem imperativen Paradigma folgt, besteht aus einer Folge von Befehlen, die nacheinander ausgeführt werden [OE13, S.341]. Bei einem deklarativen Programm handelt es sich um eine Beschreibung der Aufgabenaußensicht. Ein deklaratives Programm hat keine festgelegten Lösungsverfahren je Aufgaben. Stattdessen wird eine Lösung zum Zeitpunkt der Aufgabendurchführung mittels Inferenzmaschine abgeleitet [OE13, S.361].

Allgemein beziehen sich die Autoren darauf, dass an ein wissensbasiertes System nur geringe Anforderungen bezüglich Vollständigkeit, Widerspruchsfreiheit und Eindeutigkeit gestellt werden können. Aus diesem Grund ist das deklarative Paradigma für die Wissensrepräsentation besser geeignet. Folgende Gründe nennen die Autoren für die deklarative Umsetzung der Wissensbasis [OE13, S.366]:

- *Wissensdarstellung*: Da ein Mensch das Erfahrungswissen durch assoziative Beziehungsmuster aufbaut, ist die deklarative Wissensdarstellung eher geeignet.
- *Wissensauswertung*: Änderungen von Erfahrungswissen werden normalerweise in deklarativen Form erfasst.
- *Wissensverfügbarkeit*: Die Codewartung von einem imperativen Programm ist fehleranfällig, kosten- und zeitintensiv, da das Erfahrungswissen häufig geändert und aktualisiert werden muss.

Der objektorientierte Ansatz ist eine weitere Möglichkeit, das Wissen zu beschreiben. Ein Beispiel für die objektorientierte Implementierung wird in [KM90] vorgestellt. Die Wissensbasis wird dabei als eine Sammlung von Klassen, Objekten und Methoden definiert [KM90, S.40]. Der große Vorteil solcher Umsetzung besteht in der Modularität des Wissens. Das Wissen in unabhängige Module aufgeteilt wird. Da die einzelnen Module unabhängig voneinander sind, können sie getrennt getestet und modifiziert werden, ohne den Rest der Wissensbasis zu beeinträchtigen. Dies ermöglicht hohe Flexibilität bei der Wissensbasiserweiterung [KM90, S.43].

Neben der Implementierung der Wissensbasis ist eine geeignete Umsetzung der Wissensakquisitionskomponente erforderlich, um die Wissensbasis aktuell, möglichst fehlerfrei und konsistent zu halten. Im Folgenden wird die Wissensakquisitionskomponente in Hinsicht auf den allgemeinen Aufbau und Funktionen thematisiert.

2.4 Wissensakquisitionskomponente

Bei der Wissensakquisitionskomponente handelt es sich um einen Bestandteil eines Expertensystems, der den Wissensingenieur oder einen Experten beim Aufbau und späterer Erweiterung der Wissensbasis unterstützt [GTW90, S.18]. Allgemein umfasst die Wissensakquisitionskomponente zwei grundsätzliche Aufgaben, nämlich den Wissenserwerb und die Prüfung des Dateninputs auf Konsistenz, Vollständigkeit und Einschränkungen des Expertensystems [JFI93, S.759].

Unter dem Wissenserwerb wird eine Übertragung sowie Eingliederung von Wissen über Problemlösungsverfahren in ein Computerprogramm verstanden [GTW90, S.178]. Es werden folgende Grundarten des Wissenserwerbs unterschieden [JFI93, S.742]:

- *Indirekter Wissenserwerb:* Ein Wissensingenieur führt ein Interview mit einem Experten, oder allgemein mit einem Wissensträger. Die Analyse der Dokumente, die für das System relevant sind, gehört ebenso zur Aufgabe des Wissensingenieurs.
- *Direkter Wissenserwerb:* Ein Wissensträger gibt sein Wissen selbst mittels einer Schnittstelle ins Expertensystem ein.
- *Automatischer Wissenserwerb:* Die Wissensbasis wird entweder mithilfe der automatisierten Datenererschließung aus verfügbarer Dokumente oder Methoden des maschinellen Lernens erweitert.

Die Methoden des indirekten Wissenserwerbs lassen sich grundsätzlich in unstrukturierte und strukturierte Verfahren unterteilen. Das unstrukturierte Interview ist die am häufigsten verwendete Methode [GTW90, S.76]. Dabei stellt der Wissensingenieur dem Experten problembezogene Zwischenfragen, um ein möglichst vollständiges Bild des zur Problemlösung erforderlichen Wissens zu bekommen. Die Hauptschwierigkeit bei der Wissenserhebung durch ein Interview ist die Formulierung der Fragen. Wenn die Fragen zu spezifisch sind, können wichtige Informationen ausgelassen werden [SK09, S.95]. Eine strukturierte Vorgehensweise der Wissenserhebung ist die Protokollanalyse, wobei der Experte beim Lösen eines Problems aufgezeichnet wird. Um den Lösungsweg nachvollziehbar zu machen, kann der Experte die Aufgabe gezielt langsamer durchführen oder die Aufzeichnung mit Kommentaren versehen. Anschließend werden die erzielten Ergebnisse vom Wissensingenieur formalisiert und ins Expertensystem eingetragen.

Beim direkten Wissenserwerb soll eine geeignete Schnittstelle im Rahmen der Wissensakquisitionskomponente zur Verfügung gestellt werden, die es dem Wissensträger ermöglicht, sein Wissen ins System einzugeben. Die Schnittstelle soll eine dem Wissensträger bekannte Wissensrepräsentation verwenden und benutzerfreundlich bei der Dateneingabe sein [JFI93, S.743]. Ein Beispiel der Benutzerfreundlichkeit ist gleichzeitige Validierung der Benutzereingaben sowie eine Rückmeldung bei unzulässigen Aktionen. Der direkte Wissenserwerb hat den Vorteil, dass die Wissensbasis ohne den Wissensingenieur erweitert werden kann. Allerdings betonen die Autoren in [JFI93, S.765], dass das nur in gut verstandenen und strukturierten Anwendungsbereichen möglich sei.

Zu dem automatisierten Wissenserwerb gibt es am wenigsten Erkenntnisse, die allgemein anwendbar sind. Meistens handelt es sich um Lösungen, die nur innerhalb eines spezifischen Anwendungsbereichs funktionieren. Nichtsdestotrotz lässt sich sagen, dass das Wissen sich entweder aus vorhandenen Daten (maschinelles Lernen) oder aus verfügbaren Dokumenten (automatische Dokumentenanalyse) generieren lässt [GTW90, S.78]. Die Implementierung hängt jedoch vom Anwendungsgebiet ab. Ein Beispiel für den Fall großer

Datenmenge und Anwendung des maschinellen Lernens ist ein Diagnosesystem, das für mögliche Diagnose Fälle mit Symptomen enthält. Für die Textanalyse können beispielsweise Bedienungsanleitungen analysiert werden, wobei diese Vorgehensweise gewisse Einschränkungen ausweist. Die Schwierigkeit dabei besteht darin, dass das Erfahrungswissen nicht in den Textdokumenten zu finden ist [GTW90, S.79].

In allen Fällen des Wissenserwerbs werden die Daten an die zentrale Schnittstelle weitergereicht. Diese Schnittstelle beschäftigt sich mit der Zwischenspeicherung und der Prüfung der Datensätze auf Korrektheit, bevor die Daten endgültig in der Wissensbasis gespeichert werden. Zum Teil kann der Wertebereich direkt im einzelnen Modul des Wissenserwerbs eingeschränkt werden. Beispielsweise kann ein Eingabefeld in der Wissensträgerschnittstelle nur positive Zahlen zulassen. Für den restlichen Teil werden automatische Tests durchgeführt, die sicherstellen, dass neue Daten keine Inkonsistenzen in Bezug auf die Einschränkungen der Wissensbasis erzeugen [JFI93, S.765].

Zusammenfassend lässt sich die Wissensakquisitionskomponente schematisch in der Abbildung 5 wie folgt darstellen:

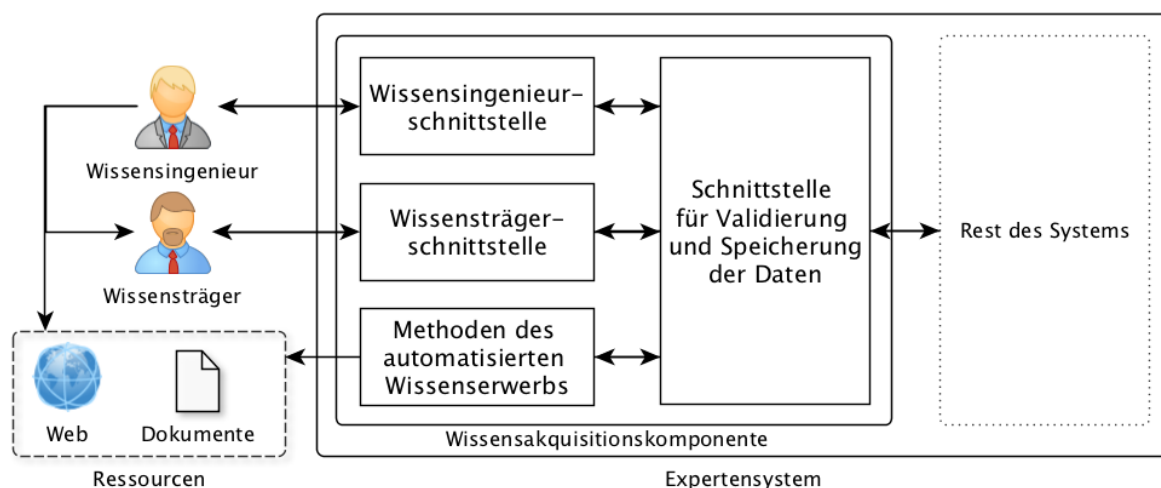


Abbildung 5: Wissensakquisitionskomponente

In der Abbildung 5 sieht man deutlich, dass die Wissensakquisitionskomponente modular aufgebaut ist. Allgemein kann dieses Modell in jedem Expertensystem eingesetzt werden, wobei die konkrete Umsetzung in Bezug auf den Anwendungsbereich spezifiziert wird. Dabei lässt sich der Automatisierungsgrad der Datenerfassung leicht anpassen, indem der Schwerpunkt auf gewünschte Bestandteile der Wissensakquisitionskomponente gelegt wird. Beispielsweise kann sich ein Expertensystem mit umfangreichen Falldaten auf das maschinelle Lernen konzentrieren. Ein System, das auf der Datenerfassung mithilfe der Wissensträgerschnittstelle basiert, wird in [SK09, S.97] vorgestellt.

Im weiteren Verlauf der Arbeit werden die Wissensträgerschnittstelle, die Komponente mit Methoden des automatisierten Wissenserwerbs und die Schnittstelle für Validierung und Speicherung der Daten thematisiert, da sie ein Potenzial für die Automatisierung der Datenerfassung aufweisen.

3 Automatisierung der Datenerfassung

Mit der Wissensakquisitionskomponente wurde bereits angedeutet, dass die Datenerfassung für Expertensysteme meistens nur teilweise automatisierbar ist. Die Autoren in [GD94] weisen ebenso darauf hin, dass manuelle und maschinelle Wissenserschließung jeweils eigene Stärke haben, die gegenseitig nicht ersetzt werden können [GD94, S.137]. Aus diesem Grund ist bei der Datenerfassung ein hybrides Modell sinnvoll, das die Vorteile manueller und maschineller Verfahren kombiniert. Aufgrund der Fragestellung dieser Arbeit wird es im weiteren auf automatisierte und halb-automatisierte Bestandteile der Wissensakquisitionskomponente beschränkt.

3.1 Die Wissensträgerschnittstelle

Bei der Wissensträgerschnittstelle handelt es sich um eine semi-automatisierte Methode zur Datenerfassung, die einen direkten Wissenstransfer zwischen einem Wissensträger und dem Expertensystem ermöglicht. Im folgenden Abschnitt wird das allgemeine Konzept der Wissensträgerschnittstelle erläutert. Darauf folgend wird ein relevantes Beispiel aus der verwandten Forschung vorgestellt und analysiert. Anschließend werden die grundlegenden Anforderungen zusammengefasst, die bei der Entwicklung der Wissensträgerschnittstelle berücksichtigt werden sollen.

Die Idee der Wissensträgerschnittstelle besteht darin, dass ein Wissensträger sein Wissen selbst und jederzeit ins System eingeben kann, ohne die Hilfe des Wissensingenieurs zu benötigen. Das Verfahren wird als semi-automatisiert bezeichnet, da die Eingabe manuell vom Wissensträger ausgelöst und automatisiert vom System bearbeitet wird.

Eine verwandte Arbeit bezüglich der Anwendung der Wissensträgerschnittstelle wird in [SK09] vorgestellt. In der Arbeit handelt es sich um ein wissensbasiertes Decision Support System (DSS) für ein Unternehmen im Bereich der Elektrotechnik. Gebus und Leiviskä definieren ein DSS als ein Informationssystem, das die Führungskräfte bei den Entscheidungen von nicht-strukturierten Problemen unterstützen soll [SK09, S.94]. Das Ziel des Systems besteht in der Datenerfassung (z.B. Geräteausfall) aus der Produktionsableitung, um die Produktionsprozesse zu optimieren. Da es sich grundsätzlich um ein wissensbasiertes System handelt, kann die Studie als ein relevantes Beispiel im Rahmen dieser Arbeit betrachtet werden.

Stichpunkte für weitere Analyse von [SK09]

- DSS war schon im Unternehmen, allerdings nur als reine Datenbank
- Prozesse werden von einem Anlagenbediener kontrolliert ausgehend von seiner Erfahrung und Kenntnisse
- Entscheidungen bei Problemen im Produktionsprozess werden von einzelnen Personen getroffen, die über Expertenwissen verfügen. Das Ziel: Wissen zu erfassen.
- Wissensrepräsentation

3.2 Maschinelles Wissenserwerb

Eine Auswahl an generellen Methoden für den maschinellen Wissenserwerb

3.3 Schnittstelle für Validierung und Speicherung

Zweck, Konzept

4 Implementierung

4.1 Ansatz 1

4.2 Ansatz 2

4.3 Ansatz 3

5 Evaluation

5.1 Test 1

5.2 Test 2

5.3 Test 3

6 Fazit

7 Ausblick

Literatur

- [CG14] CHRISTOPH BEIERLE und GABRIELE KERN-ISBERNER: *Methoden wissens-basierter Systeme: Grundlagen, Algorithmen, Anwendungen*. Vieweg, 2014.
- [GD94] GHEORGHE D. TECUCI und DAVID DUFF: *A framework for knowledge base refinement through multistrategy learning and knowledge acquisition*. Knowledge Acquisition, Seiten 137 – 162, 1994.
- [Geo08] GEORGE LAWTON: *Developing Software Online with Platform-as-a-Service Technology*. Computer, 2008.
- [Ghe92] GHEORGHE D. TECUCI: *Automating Knowledge Acquisition as Extending, Updating, and Improving a Knowledge Base*. Transactions on Systems, Man, and Cybernetics, Seiten 1444–1460, 1992.
- [GTW90] GEORG GOTTLOB, THOMAS FRÜHWIRTH und WERNER HORN: *Expertensysteme*. Springer-Verlag Wien New York, 1990.
- [HDNR97] H. FUJIHARA, D. B. SIMMONS, N. C. ELLIS und R. E. SHANNONS: *Knowledge Conceptualization Tool*. Transactions on Knowledge and Data Engineering, 1997.
- [JFI93] JOSEF MEYER-FUJARA, FRANK PUPPE und IPKE WACHSMUTH: *Expertensysteme und Wissensmodellierung*, Kapitel 7, Seiten 714–766. Addison-Wesley, 1993.
- [KM90] K. S. LEUNG und M. H. WONG: *An Expert-System Shell Using Structured Knowledge: An Object-Oriented Approach*. Computer, 23(3):38–47, 1990.
- [Mat00] MATTHIAS HAUN: *Wissensbasierte Systeme*. Expert Verlag, 2000.
- [OE13] OTTO K FERSTL und ELMAR J. SINZ: *Grundlagen der Wirtschaftsinformatik*. Oldenbourg Verlag, 2013.
- [Pet10] PETER ZÖLLER-GREER: *Künstliche Intelligenz: Grundlagen und Anwendungen*. Composita Verlag, 2010.
- [PFW⁺12] PETER MERTENS, FREIMUT BODENDORF, WOLFGANG KÖNIG, ARNOLD PICOT, MATTHIAS SCHUMANN und THOMAS HESS: *Grundzüge der Wirtschaftsinformatik*. Springer-Verlag, 2012.
- [PT11] PETER MELL und TIMOTHY GRANCE: *The NIST Definition of Cloud Computing*. NIST Special Publication 800 - 145, 2011.
- [SG14] STEFAN KOLB und GUIDO WIRTZ: *Towards Application Portability in Platform as a Service*. In: *8th International Symposium on Service Oriented System Engineering (SOSE)*, Seiten 218–229, 2014.
- [SK09] SÉBASTIEN GEBUS und KAUKO LEIVISKÄ: *Knowledge Acquisition for Decision Support Systems on an Electronic Assembly Line*. Expert Systems with Applications, Seiten 93–101, 2009.

Ich erkläre hiermit gemäß §17 Abs. 2 APO, dass ich die vorstehende Bachelorarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Bamberg, den 31.03.2017

Petr Vasilyev