

Analyse und Vergleich von Konzepten zur automatisierten Informationsbewertung im Information Lifecycle Management

Mauricio Matthesius

Business Intelligence
X-CASE GmbH
Albert-Einstein-Straße 3
98693 Ilmenau
Mauricio.Matthesius@x-case.de

Dirk Stelzer

Fachgebiet Informations- und Wissens-
management
Technische Universität Ilmenau
Postfach 100565
98684 Ilmenau
Dirk.Stelzer@tu-ilmenau.de

Abstract: In diesem Beitrag werden die wesentlichen Funktionen von ILM anhand eines Vorgehensmodells kurz beschrieben. Es werden Anforderungen an die automatisierte Informationsbewertung definiert und verschiedene Konzepte zur automatisierten Bewertung von Informationen vorgestellt. Diese Konzepte werden verglichen und hinsichtlich ihrer Stärken und Schwächen bewertet. Abschließend wird der aktuelle Stand der automatisierten Informationsbewertung zusammengefasst und Hinweise für deren Weiterentwicklung gegeben.

1 Einleitung

Information Lifecycle Management (ILM) zielt darauf ab, Informationen¹ zu bewerten und kostengünstig zu verwalten [MHP05], [TS07]. Informationen, die einen höheren Wert² für ein Unternehmen aufweisen, müssen von Informationen mit geringerem Wert getrennt [Sh06], [MW99] und den Unternehmensanforderungen entsprechend auf dem jeweils sinnvollsten Speichermedium bereitgestellt und verwaltet werden [Ab05], [Bh05], [GR93], [MHP05], [Mo06]. Die wesentlichen Gründe für die Entstehung von ILM sind das rasche Datenwachstum [KKS07], [Za04] und die zahlreichen gesetzlichen Anforderungen [JA06], [KK05], [Sh06]. Das Datenwachstum führt zu hohen Speicher- und Administrationskosten [Ab06], [Al01], [TS07]. Die Kosten für die Administration der Datenträger bzw. Datenspeicher, wie beispielsweise die Datensicherung und Spiegelung der Daten sowie die sichere Aufbewahrung sind um das vier- bis achtfache

¹ Tatsächlich werden im Rahmen des ILM Daten bzw. Dateien analysiert und bewertet. Da in den meisten Publikationen der Betrachtungsgegenstand jedoch als Information bezeichnet wird, verwenden auch wir diesen Begriff.

² Möglichkeiten zur Bestimmung des Wertes einer Information werden im weiteren Verlauf dieses Beitrags erörtert.

höher als die Beschaffungskosten der Datenträger [Gr03]. Die Unternehmensdaten müssen aufgrund der gesetzlichen Anforderungen für längere Zeiträume sicher aufbewahrt werden, was zu einer zusätzlichen Erhöhung der Kosten führt. Anhand des in Abbildung 1 dargestellten Vorgehensmodells werden die wesentlichen Funktionen des ILM kurz beschrieben.

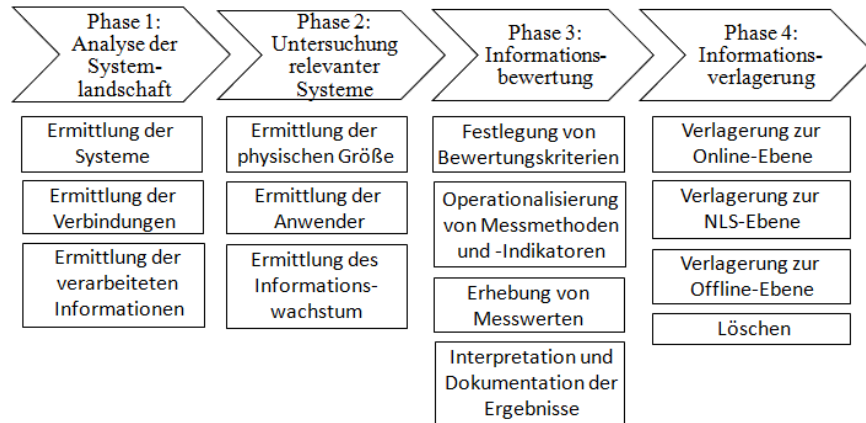


Abbildung 1: Vorgehensmodell des ILM

In Phase eins, „Analyse der Systemlandschaft“, werden in der Teilphase „Ermittlung der Systeme“ zunächst die in einem Unternehmen vorhandenen Systeme ermittelt und untersucht. Typische Beispiele sind Data Warehouses, operative Datenbanken oder Content Management Systeme. Dabei sind insbesondere die Verbindungen³ zwischen den Systemen sowie die durch die Systeme verarbeiteten Informationen von Interesse, die in den Teilphasen „Ermittlung der Verbindungen“ und „Ermittlung der verarbeiteten Informationen“ identifiziert werden. Die physische Größe, die Anwender der Systeme und das Informationswachstum der Systeme werden in Phase 2, „Untersuchung relevanter Systeme“, ermittelt. Größe und Wachstum der Systeme sind wichtige Indikatoren dafür, welche Systeme im Hinblick auf eine Kostenreduktion zuerst untersucht werden sollten. Des Weiteren werden die Anwender der jeweiligen Systeme ermittelt, da diese für die dritte Phase, die „Informationsbewertung“ von Bedeutung sind. In der Teilphase „Festlegung von Bewertungskriterien“ werden für die Informationsbewertung relevante Kriterien festgelegt, wie beispielsweise die Zugriffshäufigkeit auf die Informationen sowie das Anwender- und Administratorenwissen über die Systeme und Informationen. Vorgehensweise sowie Methoden zur Messung der relevanten Kriterien werden in der Teilphase „Operationalisierung von Messmethoden und -Indikatoren“ bestimmt. Anschließend erfolgt die „Erhebung der Messwerte“ und die „Interpretation und Dokumentation der Ergebnisse“. In Phase vier, „Verlagerung der Informationen“, werden die Informationen auf die dafür vorgesehenen Speichermedien verlagert. Haben die Informationen einen sehr hohen Wert für das Unternehmen, so verbleiben die Daten

³ Hierbei ist von besonderem Interesse, welche Systeme Informationen aus anderen Systemen aufnehmen, verarbeiten und erneut an andere Systeme weitergeben. So wird deutlich, ob bestimmte Informationen auf unterschiedlichen Systemen redundant gehalten werden und ob eine mehrfache Archivierung dieser Informationen innerhalb der folgenden Phasen des Vorgehensmodells notwendig und sinnvoll ist.

in dem jeweiligen System oder werden zurück in das System verlagert („Verlagerung zur Online-Ebene“). Informationen, auf die in regelmäßigen Abständen zugegriffen werden muss, für die aber eine Speicherung im operativen System wirtschaftlich nicht sinnvoll ist, werden auf so genannten Nearline Storage (NLS)-Systemen abgelegt („Verlagerung zur NLS-Ebene“) [LLZ04], [MHP05]. Die Informationen werden auf Archivierungssysteme verlagert, wenn auf diese für einen längeren Zeitraum nicht zugegriffen werden muss [KSH96] und die Aufbewahrung aber, zum Beispiel aufgrund rechtlicher Anforderungen, notwendig ist („Verlagerung zur Offline-Ebene“). Nicht mehr benötigte Informationen werden gelöscht („Löschen“).

Als schwierig erweist sich die Bewertung von Informationen [Ab06], [Bh05], [Ch05], [Me04], [TS07], [YB02]. Hierfür werden zahlreiche Angaben benötigt, wie beispielsweise die Anzahl der Lese- und Schreibzugriffe auf diese Objekte [Ch05], [El03]. Zusätzlich sind die gesetzlichen Anforderungen zur Aufbewahrung der Informationen [Da04], [Sh06] sowie die Gewohnheiten der Anwender [Me04], [Sha06] hinsichtlich der Informationsnutzung von Relevanz. Die Erhebung, kontinuierliche Aktualisierung und Auswertung dieser Informationen ohne automatisierte Unterstützung ist praktisch kaum möglich und wirtschaftlich nicht sinnvoll. Aus diesem Grund muss die Informationsbewertung zumindest teilweise automatisiert erfolgen. In der Literatur finden sich verschiedene Konzepte zur Gestaltung der automatisierten Informationsbewertung im ILM [Ch05], [Bh05], [CGY07], [Me04], [Sha06], [Ve05], [Za04]. Diese Konzepte stellen wir in einem der folgenden Abschnitte vor. Bisher ist unseres Wissens keine vergleichende Übersicht über die verschiedenen Konzepte zur automatisierten Informationsbewertung publiziert worden. Wir entwickeln Anforderungen an Konzepte zur automatisierten Informationsbewertung und bewerten deren Stärken und Schwächen. Zum Abschluss unseres Beitrags fassen wir den aktuellen Stand der automatisierten Informationsbewertung zusammen und geben Hinweise für deren Weiterentwicklung.

2 Anforderungen an Konzepte zur automatisierten Informationsbewertung

Im Zeitverlauf werden Informationen unterschiedlich oft verwendet bzw. weisen einen variierenden Wert für ein Unternehmen auf [Bh05], [Ch05], [Sha06]. In der Literatur werden verschiedene Kriterien für die Informationsbewertung verwendet. Denkbar ist die Bewertung der Informationen anhand der monetären Erträge, die aus der Nutzung der Informationen entstehen [G193], [MW99]. Eine solche Bewertung erweist sich jedoch als sehr aufwändig, da sich aufgrund der Informationskomplexität ein direkter Bezug zu bestimmten Umsatzgrößen nur schwer herstellen lässt. Zudem gibt es innerhalb jedes Unternehmens Informationen, welche zwar häufig genutzt werden, aber keine direkten monetären Erträge bewirken bzw. sich monetär kaum oder nur schwer bewerten lassen [Ch05], [Sha06]. Es ist auch eine Bewertung von Informationen anhand der entstandenen Kosten möglich, die bei der Sammlung, Akquisition und Erstellung der Informationen anfallen [MW99]. Je höher die entstandenen Kosten sind, desto höher ist der Wert, der den Informationen zugewiesen wird. Allerdings besteht die Möglichkeit, dass die Erzeugung der Informationen nur geringe Kosten verursacht hat, die

Informationen aber dennoch einen hohen Wert für das Unternehmen haben [OSW01]. Die oben genannten Kriterien eignen sich unserer Meinung nach für eine automatisierte Bewertung von Informationen nur schlecht, da sie sich nur schwer und mit hohem Aufwand bestimmen lassen und ihre Aussagekraft für das ILM gering ist. Wir bevorzugen die Zugriffshäufigkeit als Kriterium für die automatisierte Bewertung von Informationen im Rahmen des ILM. Dieses Kriterium lässt sich ohne großen Aufwand ermitteln und messen [Bh05], [Ch05], [CGY07], [Do04], [GS90], [Sha06], [Sk94]. Außerdem ist seine Aussagekraft im Hinblick auf eine angemessene Informationsverlagerung⁴ höher als die der anderen oben genannten Kriterien, da beispielsweise ermittelt werden kann, zu welchen Zeitpunkten tatsächlich auf eine Information zugegriffen wurde. Im Folgenden formulieren wir Anforderungen an Konzepte zur automatisierten Informationsbewertung.

Eine erste Anforderung ist die Berücksichtigung der Zugriffshäufigkeit auf die Informationen im Zeitverlauf. Ein Zugriff erfolgt, wenn eine Information gelesen oder verändert wird [Sha06]. Die Zugriffshäufigkeit variiert im Zeitverlauf [Bh05], [Ch05], [Sha06]. Die häufigsten Zugriffe erfolgen erfahrungsgemäß kurz nach dem Entstehen der Information.⁵ Danach wird in der Regel nur noch selten auf die Information zugegriffen. Etwa sechzig bis achtzig Prozent der Informationen innerhalb eines Unternehmens bleiben ungenutzt. Auf sie wird nicht mehr zugegriffen [BBK92], [Ch05], [In06], [Sha06], [Sh06], [Za04]. Die Zugriffshäufigkeit ist ein wichtiger Indikator für die Nutzung der Information.

Eine weitere Anforderung an die automatisierte Informationsbewertung ist die Klassifizierung der Informationen. Je nach ermittelter Zugriffshäufigkeit werden die Informationen verschiedenen Klassen zugeordnet [Ch05], [Do04], [Sha06]. Eine Klasse bezeichnet die Zugriffshäufigkeit einer Information sowie das geeignete Speichermedium [MHP05], auf welches sie nach der Informationsbewertung verlagert wird. Informationen, auf die häufig zugegriffen wird, werden auf kostenintensiven Speichern mit schnellen Zugriffszeiten [Pe05] abgelegt. Informationen, auf die in regelmäßigen Abständen zugegriffen wird, für die aber eine Ablage auf den kostenintensiven Speichermedien des operativen Systems nicht lohnend ist, werden auf Nearline Storage-Systemen abgelegt. Informationen, auf die nicht zugegriffen wird, deren Aufbewahrung aber aufgrund rechtlicher Anforderungen notwendig ist, können auf kostengünstigen Speichermedien langfristig aufbewahrt werden.⁶ Nicht mehr benötigte Informationen werden gelöscht.

⁴ Vgl. hierzu Abschnitt 1, Vorgehensmodell des ILM, Phase 4

⁵ Hierbei handelt es sich um einen Zeitraum von sechzig bis neunzig Tagen nach der Entstehung der Information [Ch05], [Za04].

⁶ Hierfür können Magnetbänder verwendet werden [Ja98]. Die Speicherung auf Magnetbändern gilt als kostengünstigste Alternative [HS96].

Die Berücksichtigung rechtlicher Anforderungen stellt eine weitere Anforderung an die automatisierte Informationsbewertung dar [MB07], [Sh06].⁷ Muss eine Information aufgrund rechtlicher Anforderungen aufbewahrt werden, so darf sie einerseits nicht gelöscht werden, obwohl beispielsweise für einen längeren Zeitraum nicht auf sie zugegriffen wurde. Andererseits existieren Informationen, die nach einem gesetzlich vorgeschriebenen Zeitraum gelöscht werden müssen, obwohl oft darauf zugegriffen wurde.

Die Automatisierung der Informationsbewertung unter Berücksichtigung des Anwender- und Administratorenwissens über die Systeme und Informationen ist eine weitere wichtige Anforderung. Die Automatisierung der Informationsbewertung sorgt für eine starke Vereinfachung, da der manuelle Aufwand für die Bewertung der Informationen durch Anwender oder Systemadministratoren erheblich reduziert werden kann [Ab06], [Ch05], [CW00], [KC03]. Hierfür müssen Funktionen implementiert werden, welche die Informationen anhand der Zugriffshäufigkeit bewerten und anschließend den zugehörigen Klassen zuordnen [Ch05], [Me04], [Sha06], [Ve05]. Die Bewertungsfunktionen müssen in der Lage sein, basierend auf der Historie der Zugriffshäufigkeit, die zukünftige Zugriffshäufigkeit zu prognostizieren bzw. zu berechnen [Ab06], [Do04], [El03], [Me04], [Pu94], [Sha06].⁸ Damit wird die automatisierte Informationsbewertung vereinfacht, da Zugriffshäufigkeiten und Klassenzugehörigkeiten nicht ständig neu berechnet werden müssen und eine Vorausplanung bezüglich eines definierten Zeitraums erfolgen kann. Dies wirkt sich positiv auf die Performance der Systeme und damit auf die Kosteneinsparungspotentiale aus. Trotz der Automatisierung sollte das Wissen der Anwender und Administratoren über die Systeme und Informationen berücksichtigt werden [Bh05], [Do04], [KR93], [Pa05]. Insbesondere bei der Definition der Bewertungsfunktionen und bei der Festlegung, ab welcher Zugriffshäufigkeit eine Information einer bestimmten Klasse angehören soll, ist dies von Relevanz.⁹

⁷ Ein Großteil der rechtlichen Anforderungen ist im Anschluss an die Informationsbewertung, und zwar bei der Informationsverlagerung auf das entsprechende Speichermedium, zu beachten. Relevante Rechtsvorschriften für die Aufbewahrung von Informationen sind im deutschsprachigen Raum das Bundesdatenschutzgesetz (Vgl. §§ 1, 6, 35 BDSG), das Handelsgesetzbuch (§§ 239, 257, 261 HGB), die Abgabenordnung (§§ 146, 147, 239 AO) sowie die aus diesen Rechtsvorschriften abgeleiteten Vorschriften der Grundsätze ordnungsmäßiger DV-gestützter Buchführungssysteme (GoBS; S. 1-17) und die Grundsätze zum Datenzugriff und zur Prüfbarkeit digitaler Unterlagen (GDPdU; S. 1-7). Ausländische Rechtsvorschriften sind beispielsweise der „Sarbanes Oxley Act“ (USA), der „Data Protection Act“ (Großbritannien) oder das „Electronic Ledger Storage Law“ (Japan).

⁸ Dazu können verschiedene Prognoseverfahren, wie zum Beispiel die lineare Regressionsanalyse [Sc05] oder Markovprozesse [Ha05] verwendet werden.

⁹ Anwender und Administratoren können wichtige Hinweise zur Verwendung der Informationen geben, beispielsweise auf Basis der zukünftigen Projekt- bzw. Auftragslage oder der Verwendung von Informationen innerhalb verschiedener Abteilungen des Unternehmens.

Wesentliches Ziel von ILM ist die Einsparung von Kosten. Die automatisierte Informationsbewertung sollte zu einer Reduktion der Administrationskosten für Datenspeicher [Me04], [Sha06] führen. Weiterhin muss eine Reduktion der Aufbewahrungskosten aufgrund der Speicherung der Informationen auf dem jeweils optimalen Speichermedium erfolgen [Sa99], [Sha06]. Darüber hinaus muss die Systemperformance berücksichtigt werden.¹⁰

3 Darstellung und Vergleich von Konzepten zur automatisierten Bewertung von Informationen im Rahmen des ILM

In diesem Abschnitt werden verschiedene Konzepte zur automatisierten Informationsbewertung vorgestellt und deren Stärken und Schwächen untersucht. Weiterhin wird überprüft, inwieweit die Konzepte die von uns definierten Anforderungen an die automatisierte Bewertung von Informationen¹¹ erfüllen.

Chen [Ch05] verwendet zur automatisierten Bewertung von Informationen die Zugriffshäufigkeit als Bewertungskriterium. Als wesentliche Anforderungen an ihr Konzept nennt Chen die Verwendung messbarer Indikatoren zur Bewertung der Informationen sowie die Verringerung der Notwendigkeit von Anwenderinteraktionen. Basierend auf der Zugriffshistorie einer Information wird ihr Wert für einen Zeitpunkt t berechnet. Die Zugriffshäufigkeiten der Vergangenheit werden dabei in verschiedene Zyklen aufgeteilt. Anhand dieser Zyklen wird mit Hilfe der Bewertungsfunktion der Wert einer Information berechnet, wobei aktuelle Zyklen stärker gewichtet werden als ältere. Je häufiger in der Vergangenheit auf eine Information zugegriffen wurde und je öfter diese Zugriffe kurz vor dem Bewertungszeitraum t erfolgten, desto höher der Wert der Information. Die Informationen werden anschließend entsprechend ihrer Zugriffshäufigkeit klassifiziert.¹² Eine Stärke des Konzepts von Chen ist die Verwendung der Zugriffshäufigkeit als Bewertungskriterium, da sich dieses ohne großen Aufwand ermitteln und messen lässt. Auch die höhere Gewichtung aktuellerer Zugriffszyklen als ältere erweist sich als sinnvoll.¹³ Eine Schwäche des Konzepts ist die Vernachlässigung des Anwender- und Administratorenwissens über die Systeme und Informationen. Die Informationsbewertung basiert ausschließlich auf technischen Kriterien.

Shah et al. bewerten und klassifizieren Informationen ebenfalls anhand der Zugriffshäufigkeit [Sha06]. Wichtige Anforderungen sind dabei die Berücksichtigung des Anwender- und Administratorenwissens, der Systemperformance und die Reduktion der Kosten für Datenspeicher. Anwender und Administratoren wirken aktiv an der Gestaltung der Bewertungsfunktionen und der anschließenden Verlagerung der Informationen mit. Die

¹⁰ Die Systemperformance wird positiv beeinflusst, indem nicht jede einzelne Information eines Systems untersucht und bewertet wird. Stattdessen kann ein prozentualer Anteil von Informationen, beispielsweise innerhalb eines Dateiodners, bewertet werden [MB07], [Sha06]. Der Wert dieser Informationen kann den restlichen Informationen zugewiesen werden.

¹¹ Vgl. hierzu Abschnitt 2

¹² Die Klassifizierung erfolgt dabei auf einer Werteskala von 1 (unwichtig) bis 10 (sehr wichtig).

¹³ Somit wird beispielsweise vermieden, dass für Informationen, die zwar für ältere Zyklen hohe Zugriffe aufweisen, jedoch aktuell kaum eine Rolle spielen, überbewertet werden.

Systemperformance wird bei der Bewertung der Informationen berücksichtigt, da nicht jede einzelne Information durch die Bewertungsfunktion untersucht und bewertet wird.¹⁴ Nach Aussage der Autoren können durch die optimale Verlagerung der Informationen im Anschluss an die Informationsbewertung Kosteneinsparungen von fünfzig bis fünfundsechzig Prozent erzielt werden.

Bhagwan et al. verfolgen ein ähnliches Konzept zur Bewertung von Informationen [Bh05]. Durch Verwendung des Kriteriums Zugriffshäufigkeit werden Informationen bewertet und klassifiziert. Im Vordergrund stehen die Performance der Systeme sowie die Kostenreduktion, die durch eine optimale Verlagerung der Informationen erreicht werden kann. Das Anwender- und Administratorenwissen spielt innerhalb dieses Konzeptes eine untergeordnete Rolle.

Verma et al. [Ve05] und Mesnier et al. [Me04] verwenden innerhalb ihrer Konzepte zur automatisierten Bewertung von Informationen anstelle des Bewertungskriteriums Zugriffshäufigkeit den Typ der Information bzw. den Dateityp. Informationen gleichen Typs, wie beispielsweise *.txt, besitzen nach Meinung der Autoren dieselben Eigenschaften und können einer gemeinsamen Klasse zugeordnet werden.¹⁵ Mesnier et al. unterscheiden zusätzlich, ob auf eine Information ein schreibender oder ein lesender Zugriff erfolgt. Auf Informationen vom Typ *.log erfolgt in der Regel ein schreibender Zugriff, weshalb diese Information entsprechend bewertet und danach auf Speichermedien verlagert wird, die für den Schreibzugriff optimiert sind. Allerdings vernachlässigen beide Konzepte die Zugriffshäufigkeit auf die Informationen. Somit könnten Informationen auf kostenintensive Speichermedien verlagert werden, obwohl seit einem längeren Zeitraum keine Zugriffe darauf erfolgten und eine Speicherung auf kostengünstigen Speichermedien sinnvoller wäre.

Eine starke Konzentration auf das Anwender- und Administratorenwissen bezüglich der Systeme und Informationen erfolgt innerhalb der Konzepte von Zadok et al. [Za04] sowie Chandra, Gehani und Yu [CGY07]. Jeder Anwender und Administrator erzeugt eigene Bewertungsfunktionen bzw. bewertet seine Informationen selbst. Dabei wird festgelegt, für welchen Zeitraum eine Information einen hohen Wert besitzt und ab wann eine Information auf kostengünstigere Speichermedien verlagert werden kann oder gelöscht werden soll. Dies führt zu Kosteneinsparungen, da Informationen entsprechend ihres Wertes verlagert bzw. behandelt werden. Allerdings setzen beide Konzepte eine angemessene Vorgehensweise der Anwender und Administratoren voraus. Denkbar ist aber, dass aus subjektiver Sicht der Anwender und Administratoren alle Informationen wichtig sind, obwohl nur selten darauf zugegriffen wird [Ch05]. Die Verwendung technischer und objektiver Bewertungskriterien wird bei diesen Konzepten vernachlässigt. Dies führt auch dazu, dass für die Informationsbewertung ein hoher Aufwand entsteht, da mögliche Kosteneinsparungen durch eine automatisierte Unterstützung bei der Bewertung nicht genutzt werden.

¹⁴ Vgl. Abschnitt 2, Anforderung Einsparung von Kosten

¹⁵ Bewertungsfunktionen überprüfen anschließend, ob zum Beispiel eine *.txt-Datei größer als ein Megabyte ist und ob die Auslastung des Speichermediums größer als achtzig Prozent ist. Trifft beides zu, wird die *.txt-Datei auf ein kostengünstigeres Speichermedium verlagert.

In Abbildung 2 werden die Konzepte gegenübergestellt und es wird angedeutet, inwieweit sie die in Abschnitt 2 erläuterten Anforderungen erfüllen.

Konzept	Chen	Shah et al.	Bhagwan et al.	Verma et al.	Mesnier et al.	Zadok et al.	Chandra, Gehani, Yu
Anforderungen							
Zugriffshäufigkeit							
Verwendung Zugriffshäufigkeit als Bewertungskriterium	X	X	X				
Klassifizierung							
Einteilung Informationen in Klassen	X	X	X	X	X	X	X
Rechtliche Anforderungen							
Berücksichtigung rechtlicher Anforderungen							
Automatisierung							
Bewertung der Informationen	X	X	X	X	X	X	X
Berücksichtigung Anwender- und Administratorenwissen		X		X	X	X	X
Prognose der Zugriffshäufigkeit							
Kosten							
Berücksichtigung Systemperformance		X	X	X			
Kostenreduktion	X	X	X	X	X	X	X

Abbildung 2: Vergleich der Konzepte

Alle vorgestellten Konzepte verwenden Bewertungsfunktionen für die automatisierte Informationsbewertung.¹⁶ Nach der Bewertung werden die Informationen klassifiziert¹⁷, um diese anschließend auf das jeweils optimale Speichermedium zu verlagern. Eine Reduktion der Kosten für die Administration der Datenspeicher und die Aufbewahrung der Informationen ist ebenfalls Ziel dieser Konzepte.¹⁸ Die Mehrzahl der Konzepte lässt das Anwender- und Administratorenwissen in die Bewertung der Informationen einfließen.¹⁹ Das Bewertungskriterium Zugriffshäufigkeit sowie die Systemperformance wird lediglich von drei der insgesamt sieben betrachteten Konzepte berücksichtigt.²⁰ Keines der vorgestellten Konzepte verwendet die Prognose der Zugriffshäufigkeit oder berücksichtigt die rechtlichen Anforderungen bzw. lässt diese in die Informationsbewertung einfließen.²¹

¹⁶ Vgl. Abbildung 2, Anforderung Automatisierung (Bewertung der Informationen)

¹⁷ Vgl. Abbildung 2, Anforderung Klassifizierung (Einteilung Informationen in Klassen)

¹⁸ Vgl. Abbildung 2, Anforderung Kosten (Kostenreduktion)

¹⁹ Vgl. Abbildung 2, Anforderung Automatisierung (Berücksichtigung Anwender- und Administratorenwissen)

²⁰ Vgl. Abbildung 2, Anforderung Zugriffshäufigkeit (Verwendung Zugriffshäufigkeit als Bewertungskriterium); Abbildung 2, Anforderung Kosten (Berücksichtigung Systemperformance)

²¹ Vgl. Abbildung 2, Anforderung Automatisierung (Prognose der Zugriffshäufigkeit); Abbildung 2 Rechtliche Anforderungen (Berücksichtigung rechtlicher Anforderungen)

4 Vorschläge zur Weiterentwicklung der automatisierten Informationsbewertung

Für die automatisierte Bewertung von Informationen müssen Bewertungsfunktionen definiert werden. Als Bewertungskriterium eignet sich die Zugriffshäufigkeit auf die Informationen. Bei der Definition der Bewertungsfunktionen und der Festlegung, ab welcher Zugriffshäufigkeit verschiedene Informationen welcher Informationsklasse zugeordnet werden, müssen Anwender und Administratoren der Systeme einbezogen werden [Bh05], [Do04], [KR93], [Pa05]. Dies ist wichtig, da die Bewertung von Informationen unternehmensspezifisch ist [Sh06], [GM05], [MMH96] und die Anwender und Administratoren die Besonderheiten der Systeme und Informationen am Besten kennen²² [KR93]. Gleichfalls sollten Experten hinzugezogen werden, welche die rechtlichen Anforderungen kennen. Die rechtlichen Anforderungen sind sehr komplex und werden vom Gesetzgeber häufig verändert bzw. angepasst [JA06], [KK05]. Keines der in Abschnitt 2 vorgestellten Konzepte zur automatisierten Informationsbewertung berücksichtigt rechtliche Anforderungen. Dieses Gebiet ist ein offenes Forschungsthema [Ab06], [JA06], [MB07]. Einen ersten Ansatz hierfür liefern Mont und Beato [MB07]. Sie definieren Funktionen, die ein Ziel, ein Ereignis und eine Aktion beinhalten. Das Ziel beschreibt das Speichermedium, auf das die Informationen verlagert werden. Das Ereignis enthält die Vorgehensweise, wie die Informationen aufgrund rechtlicher Anforderungen behandelt werden sollen, beispielsweise ab wann vertrauliche Informationen gelöscht werden müssen. Die Aktion beschreibt, ob die Informationen verlagert, kopiert oder gelöscht werden müssen. Eine automatisierte Aufnahme und Anpassung rechtlicher Anforderungen ist dennoch nicht möglich. Diese Aufgabe obliegt den Anwendern und Administratoren. Dieser Ansatz trägt zumindest teilweise zu einer Berücksichtigung der rechtlichen Anforderungen bei. Deshalb empfiehlt sich die Aufnahme dieses Ansatzes in ein zukünftig zu entwickelndes Konzept zur automatisierten Bewertung von Informationen, welches zugleich die Anforderungen aus Abschnitt 2 erfüllt.

5 Zusammenfassung und Ausblick

Einerseits müssen wichtige Informationen für betriebliche Aufgaben schnell verfügbar sein. Andererseits zwingt steigender Kostendruck viele Unternehmen dazu, Informationen auf möglichst kostengünstigen Speichern mit langen Zugriffszeiten vorzuhalten. ILM hilft zu entscheiden, welche Informationen auf teuren und schnellen Speichermedien verbleiben müssen und welche auf langsame aber kostengünstige Medien verlagert werden können. ILM unterstützt die Analyse und Bewertung der betrieblichen Informationen, die Ermittlung der jeweils angemessenen Speichermedien sowie die eventuelle Verlagerung der Informationen auf diese Speichermedien. Praktische Erfahrungen [JA06], [Sh06] zeigen, dass die Informationsbewertung eine besondere Herausforderung für viele Unternehmen darstellt. Allerdings dürfen die Kosten für die Durchführung des ILM, insbesondere für die Informationsbewertung, die durch das ILM erreichbaren Einsparungen nicht zunichte machen.

²² Vgl. Abschnitt 2

Unsere Analysen haben ergeben, dass insbesondere die automatisierte Unterstützung der Informationsbewertung erst wenig entwickelt ist. Unser Beitrag ist ein Schritt zur Verbesserung und Weiterentwicklung der automatisierten Informationsbewertung im Rahmen des ILM. Die in Abschnitt 3 vorgestellten Konzepte zur automatisierten Informationsbewertung konzentrieren sich auf die Bewertung von Dokumenten und Dateien. Ein großer Teil der strukturierten Informationen eines Unternehmens befindet sich innerhalb von Datenbanken, beispielsweise in einem Data Warehouse-System [Sh06].²³ Für die Bewertung von Informationen in Datenbanken und Data Warehouse-Systemen gibt es unseres Wissens nach bisher keine angemessenen Konzepte, weshalb dieser Bereich ein interessanter und viel versprechender Gegenstand zukünftiger Forschung und Entwicklung sein wird [Ab06], [BSW06], [Sh06].

Literaturverzeichnis

- [Ab05] Abd-El-Malek, M. et al.: Ursa Minor: versatile cluster-based storage. In (Gibson G. Hrsg.): Proceedings of the 4th USENIX Conference on File and Storage Technologies, San Francisco, 2005; S. 59-72.
- [Ab06] Abd-El-Malek, M. et al.: Early experiences on the journey towards self-* storage. In (Lomet, B. Hrsg.): Data Engineering Bulletin, Ausgabe 29, Nr. 4, Atlanta, 2006; S. 55-62.
- [Al01] Allen, N.: Don't waste your storage dollars: what you need to know. Research Note, Gartner Group Inc., Stamford 2001.
- [BBK92] Bennett, J.; Bauer, M.; Kinchlea, D.: Characteristics of files in NFS environments. In: ACM SIGSMALL/PC Notes, Ausgabe 18, Nr. 3/4, 1992; S. 18-25.
- [Bh05] Bhagwan R. et al.: Time-varying Management of Data Storage. In (Candea G.; Oppenheimer, D. Hrsg.): First Workshop on Hot Topics in Systems Dependability, Yokohama, 2005.
- [BSW06] Ballou, D.; Smith, I.; Wang, Y.: Sample-Based Quality Estimation of Query Results in Relational Database Environments. In: IEEE Transactions on Knowledge and Data Engineering, Ausgabe 18, Nr. 5, 2006; S. 639-650.
- [CGY07] Chandra, S.; Gehani, A.; Yu, X.: Automated Storage Reclamation Using Temporal Importance Annotations. In: 27th International Conference on Distributed Computing Systems, Toronto, 2007; S. 12.
- [Ch05] Chen Y.: Information valuation for Information Lifecycle Management. In (Parashar, M. et al. Hrsg.): Proceedings of the 2nd International Conference on Autonomic Computing, Seattle, Washington, 2005; S. 135-146.
- [CW00] Chaudhuri, S.; Weikum, G.: Rethinking database system architecture: towards a self-tuning RISC-style database system. In: International Conference on Very Large Databases, Morgan Kaufmann-Verlag, 2000; S. 1-10.
- [Da04] Sabine Dauen: Aufbewahrungspflichten. Haufe-Verlag, München, 2004; S. 3-122.
- [Do04] Douglass, F. et al.: Position: Short Object Lifetimes Require a Delete-Optimized Storage System. In: Proceedings of the 11th workshop on ACM SIGOPS European workshop: beyond the PC, ACM Press, New York, 2004.
- [El03] Ellard, D. et al.: Attribute-Based File Prediction of File Properties. Harvard, 2003; S. 1-14.

²³ Des Weiteren ist speziell im Data Warehouse-Bereich mit einem starken Informationswachstum zu rechnen, da die Bedeutung von Data Warehouses für Unternehmen stark zunimmt [Sh06], [WW07].

- [GI93] Glazer R.: Measuring the value of information: The information-intensive organization. In: IBM System Journal Ausgabe 32, Nr. 1, 1993; S. 99-110.
- [GM05] Gillet, F.; Mendel, T.: Organic IT: IT-Kosten senken, Unternehmensabläufe beschleunigen. In: (Kuhlin, B.; Thielmann, H. Hrsg.): Real-time Enterprise in der Praxis: Fakten und Praxis. Berlin, Heidelberg, New York, 2005; S. 483-502.
- [Gr03] Gray, J.: A conversation with Jim Gray. In (ACM Press Hrsg.): Queue Storage, Ausgabe 1, Nr. 4, New York, 2003; S. 8-17.
- [GR93] Gray, J.; Reuter, A.: Transaction Processing: Concepts and Techniques. Morgan Kaufmann, San Francisco, 1993; S. 54.
- [GS90] Bezalel, G.; Sheng, O.: Dynamic file migration in distributed computer systems.
- [Ha05] Hansen, K.: Prognose mithilfe von Markovprozessen. In: (Mertens, P.; Rässler, S. Hrsg.): Prognoserechnung, Physica-Verlag, 6. Auflage, Heidelberg, 2005; S. 279-303.
- [HS96] Hillyer, B.; Silberschatz, A.: Random I/O scheduling in online tertiary storage systems. In: ACM SIGMOD International Conference on Management of Data, 1996; S. 195-204.
- [In06] Inmon, W.: Information Lifecycle Management for Data Warehousing: Matching Technology to Reality. Quebec et al., 2006; S. 1-20.
- [JA06] Johnson, C.; Agrawal, R.: Intersections of Law and Technology in Balancing Privacy Rights with Free Information Flow. In (Hafner, C.; O'Rourke, M. Hrsg.): 4th International Conference on Law and Technology, Cambridge, 2006.
- [Ja98] Janko, W.: Informationswirtschaft 1. Grundlagen der Informatik für die Informationswirtschaft. Springer-Verlag, Berlin, 1998.
- [KC03] Kephart, J.; Chess, D.: The vision of autonomic computing. In: IEEE Computer, Ausgabe 36, Nummer 1, 2003; S. 41-50.
- [KK05] Kaarst-Brown, M.L.; Kelly, S.: IT-Governance and Sarbanes Oxley: The latest sales pitch or real challenges for the IT Function? In: Proceedings of the 38th Hawaii International Conference on System Sciences, Big Islands, 2005.
- [KKS07] Kanakamedala, K.; Kaplan, J.; Srinivasaraghavan, R.: A smarter approach to data storage. In: The McKinsey Quarterly, 2007; S. 1-4.
- [KR93] Keeney, R.; Raiffa, H.: Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Cambridge University-Press, 1993.
- [KSH96] Küspert, K.; Schaarschmidt, R.; Herbst, A.: Archiv Transaktionen: Ein Ansatz für asynchrones, transaktionsgesichertes Archivieren großer Datenmengen in Datenbanksystemen. In: ITG-Fachbericht zur 4. ITG/GI/GMA-Fachtagung Softwaretechnik in Automation und Kommunikation - Rechnergestützte Teamarbeit (STAK). München, 1996; S. 195-211.
- [LLZ04] Liu, B.; Li, J.; Zhang, Y.: Optimal Data Dispatching Methods in Near-Line Tertiary Storage System. In: (Li, Q.; Wang, G.; Feng L. Hrsg.): Proceedings of the 5th International Conference: Advances in Web-Age Information Management. Dalian, China, 2004, S. 690-695.
- [MB07] Mont, M.; Beato, F.: On Parametric Obligation Policies: Enabling Privacy-aware Information Lifecycle Management in Enterprises. In (IEEE Computer Society): Proceedings of the 8th IEEE International Workshop on Policies for Distributed Systems and Networks, Washington, 2007; S. 51-55.
- [Me04] Mesnier, M. et al.: File classification in self-* storage systems. In (Ibrahim, M. Hrsg.): Proceedings of the 1st International Conference on Autonomic Computing (ICAC-04). New York 2004; S. 44-51.
- [MHP05] Maier, R.; Hädrich, T.; Peinl, R.: Enterprise Knowledge Infrastructures. Springer-Verlag, Berlin, Heidelberg, New York 2005; S. 109, 116, 254-257.
- [MMH96] Marcial, F.; Matthes, J.; Hartmann, R.: Digitale Archivierung mit EDM-Systemen. In: CAD-CAM Report Nr. 5, 1996; S. 74-81.

- [Mo06] Mont, M.: On Privacy-aware Information Lifecycle Management in Enterprises: Setting the Context. In (Paulus, S.; Pohlmann, N.; Reimer, H. Hrsg.): ISSE 2006 - Securing Electronic Business Processes. Vieweg-Verlag, Wiesbaden 2006; S. 405-415.
- [MW99] Moody, D.; Walsh, P.: Measuring the value of information: An asset valuation approach. In (Heje, J. P. et al. Hrsg.): Proceedings of the 7th European Conference on Information Systems. Copenhagen, Frederiksberg, 1999; S. 496-512.
- [OSW01] Oppenheim, C.; Stenson, J.; Wilson, R.: The attributes of information as an asset. In: New Library World, Ausgabe 102, Nr. 11/12, 2001; S. 458-464.
- [Pa05] Patrascu, R. et al.: New approaches to optimization and utility elicitation in autonomic computing. Technical Report, Toronto, 2005.
- [Pe05] Petrocelli, T.: Data Protection and Information Lifecycle Management. Prentice Hall International, New York et al., 2005; S. 23.
- [Pu94] Puterman, M.: Markov Decision Processes: Discrete Stochastic Dynamic Programming. Wiley-Verlag, 1994.
- [Sa99] Santry, D. et al.: Deciding When to Forget in the Elephant File System. In: Proceedings of the 17th ACM Symposium on Operating Systems Principles, 1999; S. 110-123.
- [Sha06] Shah, G. et al.: ACE: Classification for Information Lifecycle Management. Almaden et al., 2006; S. 1-15.
- [Sh06] Short, J.: Information Lifecycle Management: An Analysis of End User Perspectives. San Diego, 2006; S. 2-37.
- [Sc05] Schneeberger, H.: Punkt-, Intervallprognose und Test auf Strukturbruch mithilfe der Regressionsanalyse. In: (Mertens, P.; Rässler, S. Hrsg.): Prognoserechnung, Physica-Verlag, 6. Auflage, Heidelberg, 2005; S. 91-104.
- [Sk94] Skyrme, D.: Ten Ways to Add Value to Your Business. In (Aslib Association for Information Management Hrsg.): Managing Information, Ausgabe 1, Nr.3, London, 1994; S. 20-25.
- [TS07] Tome, G.; Sollbach W.: Grundlagen und Modelle des Information Lifecycle Management. Springer-Verlag, Berlin, Heidelberg, New York, 2007; S. 22-30, 152, 171-172.
- [Ve05] Verma, A. et al.: An Architecture for Lifecycle Management in Very Large File Systems. In: 13th NASA Goddard, 22nd IEEE Conference on Mass Storage Systems and Technologies (MSST 2005), Monterey, 2005; S. 160-168.
- [WW07] Wixom, B.; Watson, H.: Introduction to the Minitrack on Data Warehousing and Business Intelligence. In: Proceedings of the 40th Hawaii International Conference on System Sciences, 2007; S. 214.
- [YB02] Yates-Mercer, P.; Bawden, D.: Managing the paradox: the valuation of knowledge and knowledge management. In: Journal of Information Science, Ausgabe 28, Nr.1, London, 2002; S. 19-29.
- [Za04] Zadok, E. et al.: Reducing Storage Management Costs via Informed User-Based Policies. In: 12th NASA Goddard, 21st IEEE Conference on Mass Storage Systems and Technologies (MSST 2004), Maryland, 2004; S. 101-105.