

BEDEUTUNG DES DESIGN FOR X FÜR DIE INFORMATIONSVERSORGUNG VON ENTWICKLERN INNERHALB DES ENTWICKLUNGSPROZESSES

Andreas Henrich, Karlheinz Morgenroth

Zusammenfassung

Einen entscheidenden Erfolgsfaktor in Entwicklungsprozessen bildet die Informationsversorgung der Mitarbeiter. So wirkt sich eine optimale Versorgung der Mitarbeiter mit relevanter Information nicht nur auf die Entwicklereffizienz positiv aus, vielmehr können durch eine Wiederverwendung von bewährten Konstruktionsprinzipien, Bauteilen und Werkstoffen auch erhebliche Qualitätssteigerungen in den erstellten Artefakten erzielt werden. Im vorliegenden Beitrag werden hierzu verschiedene Möglichkeiten aufgezeigt und diskutiert, wie die für ein Projekt relevanten DfX-Kriterien genutzt werden können, um die Suche nach potentiell relevanten Dokumenten – oder allgemeiner Artefakten – zielgerichteter durchzuführen.

Die in [20] präsentierte Methodik eines kontextbasierten Information Retrievals zur Versorgung der Mitarbeiter mit zu ihrer aktuellen Tätigkeit relevanter Information bildet hierzu eine denkbare Grundlage. Sie geht zunächst von einem einheitlichen Nutzermodell zur Beschreibung des Kontextes einer Person aus. Der in diesem Papier vorgestellte Ansatz sieht die Aufnahme der DfX-Kriterien und ihrer Gewichtung als weitere Aspekte für die Definition des Kontexts von Entwicklungsingenieuren vor. Dies betrifft sowohl die sich durchaus ändernden DfX-Vorgaben in den einzelnen Entwicklungsphasen eines Projektes als auch die in diesem Zusammenhang jeweils eingesetzten Methoden und Werkzeuge. Durch die explizite Erfassung der in einzelnen Arbeitsschritten vorliegenden DfX-Kriterien ermöglicht der Einsatz eines entsprechenden kontextbasierten Information Retrieval Systems zu Beginn der einzelnen Entwicklungsphasen eine Empfehlung von in analogen Situationen bewährten Dokumenten, Vorgehensweisen oder Methoden.

1 Motivation

Mit der zunehmenden Menge an verfügbarer Information wird die Identifizierung relevanter Information für die einzelne Person immer schwieriger. Für die einzelnen Mitarbeiter reicht die in einer Organisation technisch abrufbare Information von

- eher allgemeinen Inhalten über
- neue Themen und Erkenntnisse im eigenen oder in angrenzenden Fachgebieten, die zur Weiter- und Fortbildung dienen können, über
- Dokumente zur Anwendung von und zu Erfahrungen mit Vorgehensmodellen, Methoden und Werkzeugen, bis zu
- konkreten Teilergebnissen oder Komponenten, die direkt in einem aktuell bearbeiteten Projekt Verwendung finden können.

Die Anwendung bewährter Vorgehensweisen aber auch die Wiederverwendung von Komponenten bildet dabei einen Faktor, der zur Steigerung der Effizienz und Produktivität innerhalb eines Entwicklungsprozesses und zu einer höheren Qualität bei den entwickelten Produkten

führen kann. Dies bedeutet, dass der Identifikation von relevanten Inhalten, die zu einer Wiederverwendung herangezogen werden können oder in Entscheidungssituationen wertvolle Informationen liefern, ein erheblicher Stellenwert zukommt.

Trotz des Einsatzes von Suchmaschinen für Inter- und Intranet sind die gefundenen Ergebnisse für eine anfragende Person oft nicht zielführend und erschöpfend im Hinblick auf das bestehende Informationsbedürfnis. Gründe hierfür können in einem zu unscharf artikulierten Informationsbedürfnis oder schlichtweg in der Wahl falscher Suchbegriffe liegen. In der Konsequenz unterlassen Anwender häufig die Durchführung einer Suche wegen des damit verbundenen Aufwands und der ungewissen bzw. in der Vergangenheit unbefriedigenden Qualität der Ergebnisse (vgl. z. B. [32]). Die Frage ist daher, wie die Qualität der von einer Suchmaschine beim Einsatz im Rahmen eines Konstruktionsprozesses gelieferten Ergebnisse verbessert werden kann. Einen wichtigen Beitrag kann dabei die Einbeziehung des Kontextes der anfragenden Person liefern. Eine interessante Teilkomponente dieses Kontextes bilden wiederum die relevanten DfX-Kriterien und ihre Gewichtung für die aktuelle Aufgabe.

Wenn ein Entwicklungsingenieur zur Absicherung einer Entscheidung im Rahmen eines Entwicklungsprozesses nach relevanten Dokumenten sucht, so wird seine Einschätzung der Relevanz von gefundenen Dokumenten maßgeblich von den im Projekt dominierenden DfX-Kriterien bestimmt sein. Klassische Suchmaschinen für Internet und Intranet (z. B. Wissensmanagement- und PDM-Systeme) können dies jedoch nicht berücksichtigen. So würden bspw. bei den Suchbegriffen „Getriebe, Drehmoment, Material“ in Abhängigkeit davon, ob eine kostenorientierte Massenfertigung oder eine zuverlässigkeitsorientierte Einzelfertigung vorliegt, unterschiedliche Dokumente relevant sein. Dies würde bei der Suche mit herkömmlichen Suchfunktionen aber keine Berücksichtigung finden.

In Bild 1 ist ein Szenario dargestellt, in dem ein Entwickler innerhalb eines laufenden Entwicklungsprojektes mit der Arbeit an einem Gesamtentwurf beginnt. Die Grundlage für den Gesamtentwurf bilden Anforderungslisten, prinzipielle Lösungen sowie ein Vorentwurf. Die Abbildung deutet nun beispielhaft an, wie bei einer Prozessdurchführung, in der die jeweils relevanten DfX-Kriterien erfasst werden, eine Berücksichtigung dieser Kriterien bei einer Anfrage möglich ist. Beim Beginn der Arbeit am Gesamtentwurf sind in diesem Szenario

- die entsprechenden Vordokumente (Anforderungsliste, prinzipielle Lösung sowie ein Vorentwurf),
- die bei der Erstellung dieser Vordokumente eingesetzten Methoden und Werkzeuge und
- die für die Vordokumente und den Gesamtentwurf relevanten DfX-Kriterien

bekannt. Möchte der Entwickler wie in diesem Szenario aufgezeigt, zu Beginn der Arbeit am Gesamtentwurf nach Informationen zur Absicherung seiner Entscheidungen suchen, so stehen neben manuell eingegebenen Suchbegriffen auch die oben aufgezählten Informationen zur Anreicherung der Suche bereit. Man kann also z. B. im Unternehmensrepository nach in anderen Projekten erstellten Gesamtentwürfen suchen, die für „ähnliche“ Vordokumente mit „vergleichbaren“ DfX-Einschätzungen erstellt wurden. Die Hinzunahme der DfX-Kriterien kann dabei zweifellos eine Verbesserung der Ergebnisqualität bewirken, da bei einer hohen Bedeutung des geringen Materialeinsatzes im aktuellen Projekt sicher Gesamtentwürfe aus Projekten mit ähnlichen DfX-Einschätzungen relevanter sein werden als Gesamtentwürfe aus Projekten mit konkurrierenden DfX-Präferenzen.

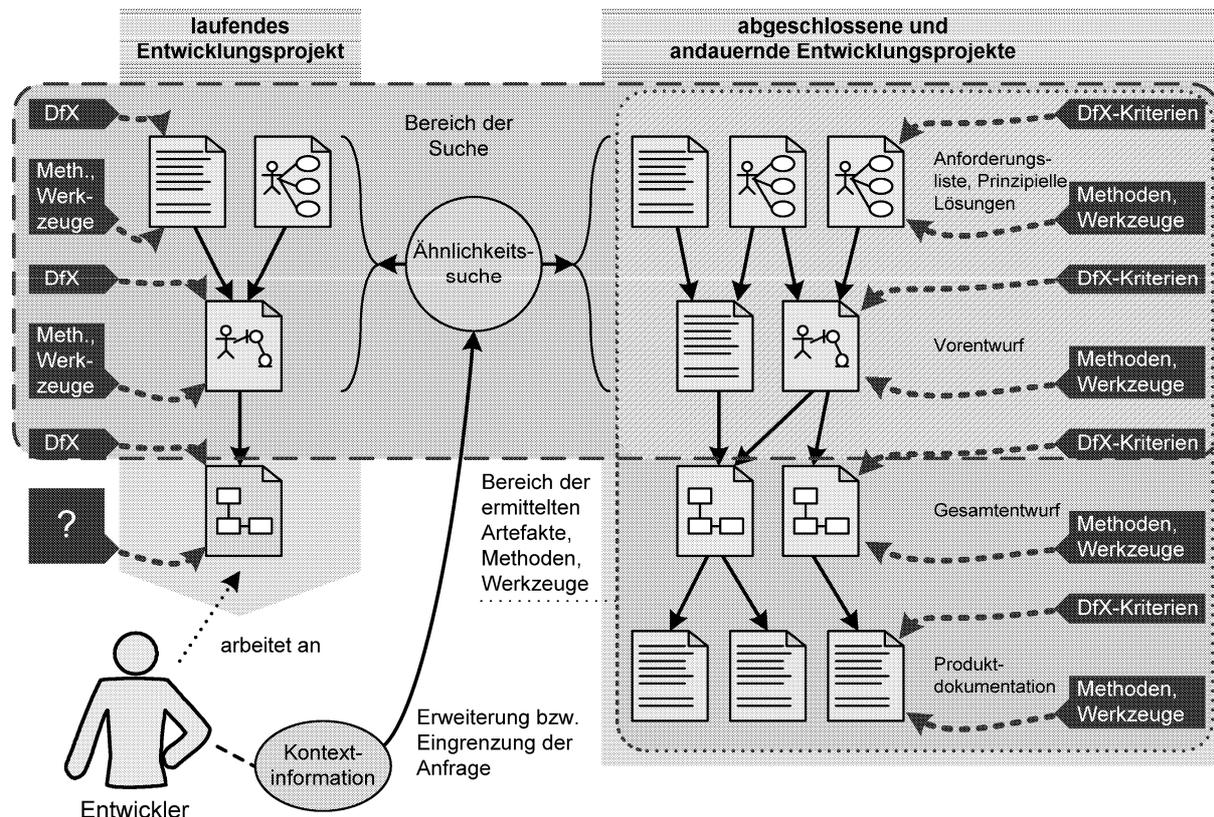


Bild 1: Szenario eines Entwicklungsingenieurs, der mit einem Gesamtwurf beginnt und dafür über den Einsatz geeigneter Methoden und Werkzeugen zu entscheiden hat.

Der Einsatz des in diesem Papier beschriebenen Ansatzes ist dabei nicht auf die Suche im Unternehmensrepository beschränkt. Vielmehr könnte man z. B. auch bei einer Suche, die der Entwicklungsingenieur aus Bild 1 im Internet absetzt, automatisch einige Schlüsselbegriffe ergänzen, die den relevanten DfX-Kriterien entsprechen.

Unser in diesem Papier präsentierter Ansatz sieht daher eine zusätzliche Berücksichtigung der DfX-Kriterien für die Suche im Rahmen eines Entwicklungsprozesses vor. Neben einer Übersicht zu den dafür in Frage kommenden Ansätze und Methoden aus dem Bereich des Information Retrieval wird in Abschnitt 2.6 eine exemplarische Grobarchitektur für die Berücksichtigung von DfX-Kriterien bei der Suche aufgezeigt.

2 Mögliche Ansätze für die Berücksichtigung von DfX-Kriterien bei der Suche

Eine Einbeziehung der DfX-Kriterien bei der Suche kann mit verschiedenen Ansätzen und Methoden aus dem Bereich des Information Retrieval realisiert werden. Welche dieser Ansätze in welcher Situation sinnvoll sein können, werden wir nach Vorstellung der verschiedenen Ansätze in Abschnitt 2.6 diskutieren.

2.1 Kombination mehrerer Kriterien für die Relevanz eines Dokumentes

Grundsätzlich sollen die DfX-Kriterien in der Regel zusätzlich zu weiteren Anfragekriterien bei der Beurteilung der Relevanz der Ergebnisdokumente berücksichtigt werden. So könnte z. B. im Rahmen eines Konstruktionsprozesses von einem Entwicklungsingenieur eine Anfrage „Getriebe, Drehmoment, Material“ gestellt werden, bei deren Bearbeitung zusätzlich

zum Bezug der gesuchten Dokumente zu den drei Anfragebegriffen auch die Passung der Dokumente zu den im aktuellen Anfragekontext relevanten DfX-Kriterien betrachtet werden soll. Für die potentiellen Ergebnisdokumente gibt es damit zwei Kriterien für Relevanz, die beide üblicherweise durch einen Wert (den so genannten *Retrieval Status Value*, RSV) zwischen 0 (= keine Relevanz) und 1 (= perfekte Relevanz) repräsentiert werden. Zur Kombination dieser Werte können nun verschiedene Verfahren eingesetzt werden:

- Man kann aus beiden Kriterien z. B. einen gewichteten Mittelwert bilden (d stehe für ein Dokument und q für eine Anfrage):

$$RSV_{\text{gesamt}}(q,d) = \alpha \cdot RSV_{\text{Anfrage}}(q,d) + (1 - \alpha) \cdot RSV_{\text{DfX}}(q,d)$$

Diese Formel unterstellt offensichtlich eine gewisse Substituierbarkeit des Erfüllungsgrades beider Kriterien. Wichtig ist dabei die geeignete Wahl des Gewichtungsparameters α , die ggf. in Testreihen ermittelt werden muss.

- Eine Alternative ist, eines der Kriterien als Filter zu nutzen und für die Rangordnung im Ergebnis nur das verbleibende Kriterium zu verwenden. So könnte man z. B. nur Dokumente d im Ergebnis präsentieren, für die gilt: $RSV_{\text{DfX}}(q,d) > 0,7$. Diese Dokumente würden dann im Ergebnis nach $RSV_{\text{Anfrage}}(q,d)$ sortiert ausgegeben. Natürlich könnte man auch umgekehrt verfahren und nur Dokumente mit $RSV_{\text{Anfrage}}(q,d) > 0,8$ sortiert nach ihren $RSV_{\text{DfX}}(q,d)$ Werten im Ergebnis präsentieren.

Man beachte, dass das letztgenannte Vorgehen im Prinzip von vielen Internet-Suchmaschinen eingesetzt wird. RSV_{DfX} wird dabei allerdings durch ein Kriterium ersetzt, das auf Basis der auf eine Webseite verweisenden Links deren Popularität abschätzt. So werden dann nur Dokumente im Ergebnis präsentiert, die von den Suchbegriffen her hinreichend gut zur Anfrage passen (in der Regel also alle Anfragebegriffe enthalten). Diese Dokumente werden nach der durch die eingehenden Links approximierten Popularität sortiert.

In der Literatur gibt es zu den oben skizzierten Verfahren noch zahlreiche Varianten und Alternativen (vgl. z. B. [29]), die allerdings keine grundsätzlichen Neuerungen mit sich bringen. Für uns steht daher im Folgenden die Frage im Vordergrund, wie $RSV_{\text{DfX}}(q,d)$ berechnet werden kann.

2.2 Suche nach Begriffen

Auch ohne die oben beschriebene Kombination mehrerer Kriterien könnte ein Entwicklungsingenieur natürlich versuchen, die relevanten DfX-Kriterien explizit in die Anfrage einzubeziehen, indem er seine Suchanfrage selbst mit entsprechenden Suchbegriffen erweitert. Um das obige Beispiel aufzugreifen, wäre dies eine Erweiterung der Suchanfrage „Getriebe, Drehmoment, Material“ bspw. um die Begriffe „kostengünstige Fertigung“. Wenn man diesen Ansatz auf zwei getrennte Kriterien RSV_{Anfrage} und RSV_{DfX} übertragen will, dann könnte man mit den Suchbegriffen „Getriebe, Drehmoment, Material“ den Wert RSV_{Anfrage} und mit den Begriffen „kostengünstige Fertigung“ den Wert RSV_{DfX} berechnen.

Im Bereich des Information Retrieval wurden in den vergangenen 40 Jahren verschiedenste Ansätze entwickelt, um aufgrund von Suchbegriffen RSV-Werte für Dokumente zu berechnen [7]. Im einfachsten Fall gilt für eine Anfrage q und ein Dokument d , dass $RSV(q,d) = 1$ ist, wenn alle Suchbegriffe im Dokument enthalten sind. Anderenfalls gilt $RSV(q,d) = 0$. Dies entspricht einem Booleschen Retrieval-Modell mit UND-Verknüpfung der Anfragebegriffe. Alternativ könnte man $RSV(q,d)$ gleich dem Anteil der im Dokument d enthaltenen Begriffe aus q wählen. Als weiteres Beispiel repräsentiert das Vektorraummodell [28] Dokumente und Anfragen als Vektoren in einem Raum, in dem die Begriffe/Wörter die einzelnen Dimensio-

nen aufspannen. Die Anzahl der Dimensionen entspricht damit der Anzahl der unterschiedlichen in der Dokumentensammlung vorkommenden Begriffe und liegt in der Regel weit über 10.000. Jede einzelne Vektorkomponente beschreibt, wie gut der Begriff bzw. das Wort die Anfrage bzw. das Dokument beschreibt. Die Formel zur Berechnung der Komponente $w_{d,k}$ im Beschreibungsvektor eines Dokumentes d für den Begriff k könnte z. B. $w_{d,k} = tf_{d,k} \cdot n_k/N$ lauten, wobei $tf_{d,k}$ angibt, wie oft der Begriff k im Dokument d vorkommt, und N der Anzahl der Dokumente insgesamt und n_k der Anzahl der Dokumente, die den Begriff k enthalten, entspricht. n_k/N steht damit für die „Trennschärfe“ des Begriffs k . Den Wert $RSV(q,d)$ kann man nun z. B. durch das so genannte Kosinusmaß berechnen. Es entspricht dem Kosinus des Winkels zwischen Anfragevektor und Dokumentvektor und ist umso größer, je ähnlicher die Richtung der Vektoren ist.

Auch wenn eine in der obigen Weise aus den beiden Anfragen „Getriebe, Drehmoment, Material“ sowie „kostengünstige Fertigung“ kombinierte Anfrage aus Sicht des Entwicklungsingenieurs sein Informationsbedürfnis wiedergibt, so werden die zu erwartenden Suchergebnisse kaum diesem Informationsbedürfnis gerecht werden, da Suchmaschinen im Wesentlichen über Wortvorkommen arbeiten und kein semantisches Verständnis der Anfrage haben. Zwar verwenden heute übliche Information Retrieval Systeme, die bspw. in modernen PDM, Dokumenten- oder Wissensmanagementsystemen integriert sind, fortgeschrittene Methoden für die Verarbeitung natürlicher Sprache, wie die Reduktion von Wörtern auf ihre Stammform oder Grundform sowie eine Behandlung von Mehrwortgruppen. Das bedeutet, dass z. B. Begriffe wie „kostengünstige“ auf ihre Stammform „kostengünstig“ und durch eine Behandlung von Mehrwortgruppen in die einzelnen Bestandteile „kosten“ und „günstig“ aufgetrennt werden. Gleichwohl können damit immer noch nur die Dokumente gefunden werden, die genau diese Begriffe enthalten. Werden in relevanten Dokumenten statt dessen Begriffe wie bspw. „effizient“ verwendet, so landen diese Dokumente kaum unter den ersten Treffern unserer Suchanfrage.

Eine Maßnahme zur Erweiterung von Suchanfragen in diesem Sinne kann der Einsatz eines Thesaurus sein, der sinn- und artverwandte Wörter (Synonyme) in Verbindung setzt. Dabei wird üblicherweise die Suche um Begriffe aus dem Thesaurus erweitert, die in Verbindung mit den angegebenen Suchbegriffen stehen. In einem Suchergebnis sind damit zumeist mehr Dokumente enthalten (die Vollständigkeit des Ergebnisses, der sog. Recall, steigt) wobei aber fast zwangsläufig die Genauigkeit absinkt (die sog. Precision), da auch nach Begriffen gesucht wurde, die nicht mehr mit der expliziten Anfrage angegeben wurden (vgl. z. B. [10]). Typische Thesauri umfassen allerdings nur einen grundlegenden, allgemeinen Wortschatz, so dass je nach Spezialgebiet besondere Fachthesauri teilweise erst erstellt und kontinuierlich gepflegt werden müssten. Sind in der vorzufindenden Fachsprache viele englischsprachliche Begriffe enthalten, so gilt es darüber hinaus Thesauri verschiedener Sprachen zu verknüpfen – ein komplexes Unterfangen, das Gegenstand des „Cross Language Information Retrieval“ ist [21].

Allerdings gilt es an dieser Stelle zu konstatieren, dass die in heute üblichen PDM, Dokumenten- oder Wissensmanagementsystemen integrierten Information Retrieval Systeme in der praktischen Nutzung häufig nur einen Teil ihrer eigentlichen Fähigkeiten ausschöpfen können. So werden bspw. die sehr häufig verwendeten Information Retrieval Komponenten von Verity¹ bzw. Autonomy² oft nur in einer sehr rudimentären Standardkonfiguration eingesetzt. So dass die oben aufgezeigten Methoden, wie eine Reduktion auf Stammform oder Grundform oder eine Behandlung von Mehrwortgruppen häufig nicht stattfindet bzw. nicht aktiviert ist.

¹ http://www.verity.com/de/products/k2_arch.html, Abruf: 25.8.2006

² http://www.autonomy.com/content/Products/IDOL/f/Enterprise_Search.en.html, Abruf: 25.8.2006

2.3 Suche über Ontologien

Über den Einsatz von Thesauri hinaus geht eine Integration von Ontologien in die Suche. Ontologien definieren nicht nur die Beziehungen zwischen verwandten Begriffen. Viel mehr werden einzelne Konzepte beschrieben und untereinander in semantische Beziehungen gesetzt (vgl. z. B. [9]). Für eine Berücksichtigung der DfX-Kriterien in einer Suche auf Basis von Ontologien bedeutet dies zunächst eine Definition einer entsprechenden Ontologie, die neben der Abbildung einer hierarchischen Strukturierung verschiedener DfX-Kriterien (vgl. z. B. [3]) auch deren wechselseitige Beziehungen sowie wichtige Schlüsselbegriffe berücksichtigt. Die Beziehungen können dabei auch mit Gewichten annotiert sein, die zur Berechnung einer „ontologischen Distanz“ zwischen Konzepten genutzt werden können. So könnte z. B. die Beziehung zwischen den DfX-Kriterien „Design for Manufacture“ und „Design for Assembly“ (vgl. [3]) die relativ hohe Ähnlichkeit der Kriterien ausdrücken.

Die Einbindung von Ontologien in den Suchvorgang sollte im Idealfall sowohl bei der Erfassung (Indexierung) von Dokumenten als auch bei der Durchführung des eigentlichen Suchvorgangs erfolgen. So gilt es bei der Erfassung von Dokumenten diesen verschiedene Konzepte der Ontologie zuzuordnen. Erfolgt eine Anfrage in Form von natürlichsprachlichen Begriffen, so muss auch hier eine möglichst automatisierte Zuordnung von Konzepten erfolgen. Eine Präsentation der gesamten Ontologie in der Suchmaske einer Suchmaschine und eine manuelle Auswahl von präferierten Konzepten bzw. in diesem Fall von DfX-Kriterien durch einen Fragesteller sind zwar denkbar, allerdings werden derartige Funktionalitäten nur sehr wenige, geübte Anwender nutzen. Hier zeigen gerade die sehr schlichten Benutzeroberflächen gebräuchlicher Internetsuchmaschinen den Weg auf, der zu einer hohen Akzeptanz beim Nutzer führt.

Für unser Szenario ist die Ermittlung der die relevanten DfX-Kriterien beschreibenden Konzepte für die *Anfrage* allerdings ggf. aus dem zur Projektverwaltung verwendeten Softwaresystem möglich. Im Idealfall sind in diesem System die für die aktuelle Aufgabe priorisierten DfX-Kriterien hinterlegt, so dass diese von einer integrierten Suchfunktion genutzt werden können. Ein wesentlich komplexeres Problem stellt aber die Zuordnung von Konzepten zur Beschreibung der relevanten DfX-Kriterien bei den *Dokumenten* dar. Hier kann nicht davon ausgegangen werden, dass die Begrifflichkeit eines Konzeptes genau so auch im Dokument auftritt. Typische Ansätze begegnen diesem Problem entweder mittels des Vektorraummodells (z. B. [4], [27]) oder mit probabilistischen Ansätzen. Sie versuchen, entweder auf Basis der in einer Ontologie bei Konzepten hinterlegten textuellen Beschreibungen oder einer manuell vorgenommenen Zuordnung einer Reihe von Beispieldokumenten, neue Dokumente einem oder mehreren Konzepten zuzuordnen. Dies führt dann in die Richtung lernender Verfahren, die wir in Abschnitt 2.5 ansprechen werden.

Zusammenfassend erscheint der Einsatz von Ontologien in unserem Kontext durchaus vielversprechend, weil in einer Ontologie die Beziehungen der einzelnen DfX-Kriterien zueinander sowie die Beziehungen zwischen DfX-Kriterien und wichtigen Schlüsselbegriffen modelliert werden können. Allerdings bleiben als Probleme der Erstellungs- und Pflegeaufwand für die Ontologie sowie die Zuordnung der in der Ontologie verwalteten Konzepte zu Dokumenten.

2.4 Einsatz von Metadaten

Natürlich kann man insbesondere für die unternehmensintern verwalteten Dokumente auch daran denken, dass diesen in einem manuellen Prozess Daten zu den relevanten DfX-Kriterien zugeordnet werden (so genannte Metadaten). Die Möglichkeiten reichen hier von der Auszeichnung von Dokumenten mit allgemeinen Metadaten und Stichwörtern, wie dies

bspw. Dublin Core³ vorsieht, bis zum Einsatz eines kontrollierten Vokabulars bzw. von Ontologien. Im Sinne der DfX-Kriterien sollte eine Auszeichnung der Dokumente dabei mit eben denjenigen Kriterien erfolgen, die in einem Dokument berücksichtigt werden. Für eine Suche stellen vorhandene Metadaten den Idealfall dar, wenn die Qualität der Metadaten entsprechend gewährleistet ist. Doch eben in einer qualitativ hochwertigen Auszeichnung von Metadaten liegt das Hauptproblem, weil in einer praktischen Umsetzung, die Metadaten häufig unvollständig und mit zweifelhafter Qualität gepflegt werden.

In Produktentwicklungsprozessen würde dies bei einfacher Umsetzung einen zusätzlichen Aufwand für die Mitarbeiter bedeuten, die ein Dokument erstellen oder weiter bearbeiten, da diese Dokumente mit entsprechenden Metadaten ausgezeichnet werden müssten. Im Sinne einer Qualitätssicherung müsste diese Auszeichnung ggf. auch einem Reviewprozess unterliegen. Darüber hinaus wäre ein direkter Nutzen aus dieser Auszeichnung erst in späteren Phasen oder gar anderen Projekten zu ziehen, wenn eben entsprechende Dokumente zur Absicherung von Entscheidungen gesucht würden. Insgesamt darf man davon ausgehen, dass dieses Unterfangen unter den typischerweise in Projekten anzutreffenden zeitlichen wie auch finanziellen Rahmenbedingungen ein kaum zu realisierendes Verfahren darstellt. Eine interessante Alternative kann in diesem Zusammenhang wieder der Einsatz von Software zum Projektmanagement bieten, die „DfX-aware“ ist. Wenn bei der Ablage von Dokumenten durch einen Entwicklungsingenieur immer automatisch die für die aktuelle Aufgabe relevanten DfX-Kriterien als Metadaten annotiert werden, dann entfällt der manuelle Zusatzaufwand und auch die Qualität scheint – zumindest auf einem grundlegenden Niveau – gewährleistet [13].

2.5 Einsatz lernender Verfahren

Während eine manuelle oder automatische Zuordnung von DfX-Kriterien zu einzelnen Dokumenten im Bereich der unternehmensinternen Dokumente vielleicht noch möglich sein mag, scheidet ein solcher Ansatz für Suchen im Internet offensichtlich aus. In diesem Umfeld bieten sich verschiedene lernende Verfahren aus dem Bereich des Information Retrieval an. Ziel ist es dabei, eine Verbindung von Dokumenten zu den entsprechenden DfX-Kriterien herzustellen, indem man entsprechende Zuordnungsvorschriften aus Beispieldokumenten „lernt“. Grundlage aller lernenden Verfahren ist damit ein Satz von Trainingsdaten. In unserem Fall also eine Reihe von Dokumenten, die manuell mit den jeweiligen DfX-Kriterien verknüpft werden, die durch sie repräsentiert bzw. in ihnen behandelt werden.

Für eine Zuordnung verschiedener Kriterien zu Texten haben sich in der Vergangenheit im Bereich der lernenden Verfahren z. B. der Naive Bayes'sche Klassifikator [33] oder Support Vector Machines [14] bewährt. Mittels des Naiven Bayes'schen Klassifikators lassen sich die Wahrscheinlichkeiten für ein Dokument berechnen, mit denen es bestimmte Kriterien (z. B. „niedrige Kosten in der Produktion“, „hohe Belastbarkeit“) erfüllt bzw. sich mit den jeweiligen Kriterien befasst. Im Gegensatz dazu kann beim Einsatz von Support Vector Machines für ein Dokument immer nur entschieden werden, ob dieses sich einem Kriterium zuordnen lässt oder eben nicht. Im praktischen Einsatz bedeutet dies bspw. für die DfX-Kriterien, dass für jedes DfX-Kriterium entschieden werden muss, ob sich ein Dokument diesem oder der Gesamtheit der restlichen Kriterien zuordnen lässt (siehe Bild 2).

³ <http://dublincore.org/>, Abruf: 25.8.2006

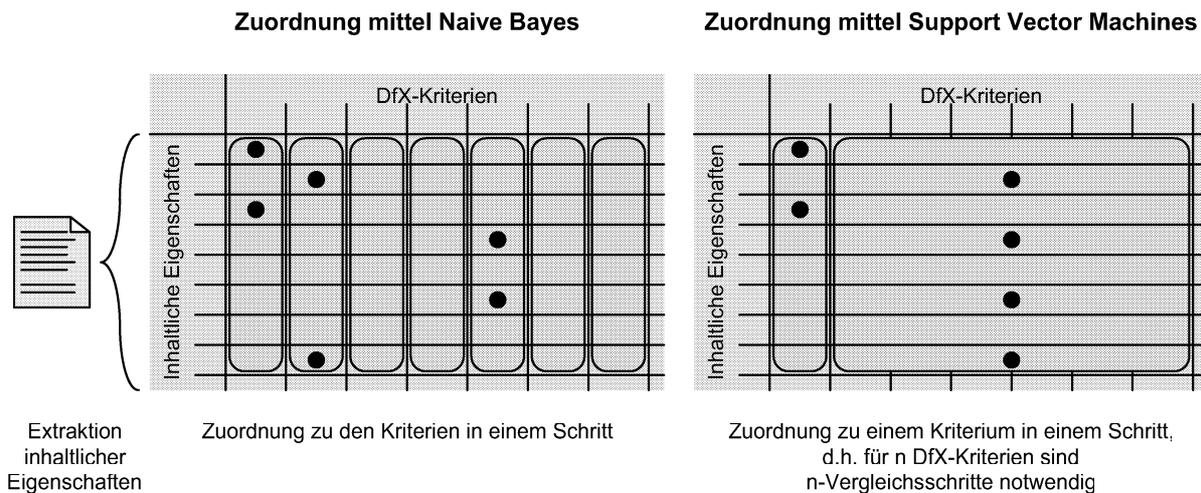


Bild 2: Zuordnung von DfX-Kriterien zu einem Dokument auf Basis inhaltlicher Eigenschaften

Für ein lernendes Verfahren ergibt sich dabei in jedem Fall das Problem, dass für die Lernphase repräsentative Beispieldokumente für die in Bild 2 durch abgerundete Rechtecke umrissenen Dokumentmengen gefunden werden müssen. Hätte man z. B. für die Situation auf der rechten Seite für das erste DfX-Kriterium zufällig nur Beispiele, die sich mit den oberen beiden inhaltlichen Eigenschaften beschäftigen, während sich die Beispieldokumente für die anderen DfX-Kriterien z. B. mit den unteren inhaltlichen Eigenschaften beschäftigen, so würde die Support Vector Machine nicht auf die Unterscheidung der DfX-Kriterien sondern zumindest auf einen unerwünschten Mix aus DfX-Kriterien und inhaltlichen Eigenschaften trainiert.

Trotz dieser Probleme bei der Bestimmung einer „sauberen“ Trainingsmenge könnte sich hier ein praktikabler Ansatz zur Zuordnung von DfX-Kriterien zu Dokumenten im Internet ergeben. Inhaltlich steht hinter diesem Ansatz vereinfacht formuliert, dass die Verfahren auf Basis der Trainingsmenge „lernen“, welche Begriffe/Konzepte Dokumente charakterisieren, die bei Priorisierung bestimmter DfX-Kriterien relevant sein können. Wenn dann ein neues Dokument ebenfalls diese charakteristischen Begriffe enthält, dann wird es auch mit diesen DfX-Kriterien annotiert.

2.6 Einsatzmöglichkeiten der Ansätze

Mit den oben skizzierten Ansätzen aus dem Information Retrieval lässt sich nun ein erstes Gesamtkonzept für die Berücksichtigung von DfX-Kriterien bei Suchanfragen konzipieren:

- Man verwendet eine geeignete Strategie zur Kombination von inhaltlichen Kriterien und von DfX-Kriterien für die Beurteilung der Relevanz von Dokumenten. Hierzu kann z. B. eine Formel wie $RSV_{\text{gesamt}} = \alpha \cdot RSV_{\text{Anfrage}} + (1 - \alpha) \cdot RSV_{\text{DfX}}$ genutzt werden.
- Die für eine Anfrage relevanten DfX-Kriterien werden nach Möglichkeit aus der zur Projektverwaltung eingesetzten Software gewonnen und automatisch bei Suchanfragen im Rahmen dieser Umgebung genutzt.
- Bei Dokumenten, die im Unternehmen selbst entstehen, wird nach Möglichkeit eine automatische Annotation mit den bei der Erstellung relevanten DfX-Kriterien durchgeführt, so dass hier kein manueller Zusatzaufwand entsteht. Die automatischen Zuord-

nungen sollten dabei ggf. im Rahmen einer Qualitätssicherung manuell überschrieben werden können.

- Für alle anderen Dokumente erfolgt eine Annotation mit DfX-Kriterien durch lernende Verfahren. Dabei könnte eine interessante Idee sein, als Trainingsmenge die automatisch annotierten im Unternehmen selbst entstandenen Dokumente zu nutzen.
- Um nicht von den konkreten Begrifflichkeiten abhängig zu sein, sollte eine Ontologie für die verwendeten DfX-Kriterien zum Einsatz kommen.

3 Verwandte Ansätze

Es existiert eine Reihe von Forschungsansätzen, die sich mit der Erfassung und Beschreibung weiterer Kriterien mit dem Ziel einer Verbesserung oder der automatischen Ermittlung von Suchergebnissen beschäftigen (siehe auch [20]). Dabei wird der aktuelle und historische Kontext von Anwendern erfasst und als Grundlage zur Verbesserung oder automatischen Ausführung von Suchanfragen verwendet.

Eine Gruppe von Ansätzen bilden Systeme, die den Text eines aktuell von einer Person bearbeiteten Dokumentes zur Formulierung von Anfragen verwenden (vgl. [25], [6], [8]). Gesucht werden dabei relevante Dokumente, wobei der betrachtete Kontext sich aber im Wesentlichen über den aktuell bearbeiteten Text definiert. Andere Ansätze beobachten einen Benutzer beim „Surfen“ im Web und geben Empfehlungen zu thematisch als ähnlich erachteten Webseiten (vgl. [16], [15], [22]). Darüber hinaus kommen Methoden des Collaborative Filtering sowohl in Empfehlungssystemen für Texte als auch für Produkte im E-Commerce zum Einsatz (vgl. [26], [30], [31]). Eine Integration entsprechender kontextbasierter Suchdienste zeigen bspw. [18] für Managementinformationssysteme und [23] für Portalsysteme. Im Gegensatz zu dem in diesem Papier verfolgten Ansatz beziehen sich diese Ansätze auf spezielle, wohl abgrenzte Einsatzszenarien, wie z. B. das Verfassen von Dokumenten mittels einer Textverarbeitung. Gleiches gilt für die berücksichtigten Quellen für die Erfassung von Kontextinformation.

Auf dem Gebiet des Workflow-Management setzen u. a. das Forschungsprojekt KnowMore, das sich daran anschließende DECOR-Projekt [1, 2] und das PreBIS-Projekt [5] auf existierende Workflow-Schemata, die um weitere Metadaten ergänzt wurden. In diesen sind explizit die Informationsbedarfe der entsprechenden Aufgaben in Form von Anfragen formuliert, die bei Erreichen einer Aufgabe gegenüber verschiedenen Suchdiensten zur Ausführung gebracht werden [17]. Allerdings setzen diese Ansätze auf eine starre Prozessmodellierung, die der dynamischen Entfaltung üblicher Produktentwicklungsprozesse entgegensteht.

Ein weiteres, hier relevantes Forschungsgebiet stellt die Kombination verschiedener Ähnlichkeitskriterien dar. Auf diesem Gebiet wurden in den vergangenen Jahren zahlreiche Ansätze und Algorithmen entwickelt (vgl. z. B. [24, 11]). Darüber hinaus existieren auch Ansätze zur Übertragung von Rangordnungen mit unterschiedlichen Semantiken [12]. Wie bereits die in Abschnitt 2 gezeigten Ansätze aus dem Bereich des Information Retrieval stellen auch diese Methoden zur Kombination und Übertragung von Ähnlichkeitskriterien und Rangordnungen Basiswerkzeuge dar, die in dem hier skizzierten Anwendungsfeld zum Einsatz kommen können.

4 Zusammenfassung und Ausblick

In diesem Papier wurden Ansätze und Methoden präsentiert, die eine Integration der DfX-Kriterien in die Informationsversorgung von Ingenieuren innerhalb eines Entwicklungsprozesses ermöglichen. Der in [19] aufgezeigte und im Rahmen des FORFLOW-Projektes ver-

folgte Prozessnavigator stellt gleichsam die Rahmenarchitektur bereit, die sowohl die Entscheidungsdeterminanten der Entwicklungsingenieure als auch die verschiedenen Quellen für eine derartige Informationsversorgung umfassen. Ein Projekt im FORFLOW-Forschungsverbund widmet sich einer kontextsensitiven Suche, die Mitarbeiter mit zu ihrer aktuellen Tätigkeit relevanter Information besser versorgen soll, als dies heute vorherrschende Suchdienste erlauben. In diesem Forschungsprojekt werden die hier aufgezeigten Methoden und Ansätze aus dem Bereich des Information Retrieval dazu eingesetzt werden, die Informationsversorgung kontextsensitiv u. a. auch auf Basis der für eine Projektphase vorherrschenden DfX-Kriterien zu verbessern.

5 Literatur

- [1] Abecker, A.; Bernardi, A.; Hinkelmann, K.; Kühn, O.; Sintek, M.: Context-Aware, Pro-active Delivery of Task-Specific Information: The KnowMore Project. *Information Systems Frontiers*, 2(3/4):253–276, 2000.
- [2] Abecker, A.; Bernardi, A.; Ntioudis, S.; Herterich, R.; Houy, C.; Legal, M.; Mentzas, G., Müller, S.: The DECOR Toolbox for Workflow-Embedded Organizational Memory Access. In: ICEIS 2001, 3rd International Conference on Enterprise Information Systems, Setúbal, Portugal, 7.–10. Juli 2001.
- [3] Bauer, S.: Design For X – Ansätze zur Definition und Strukturierung. In: Meerkamm, H. (Hrsg): Beiträge zum 15 Symposium „Design for X“, Neukirchen, 2003.
- [4] Bernstein, A.; Kaufmann, E.; Buerki, C.; Klein, M.: Object Similarity in Ontologies: A Foundation for Business Intelligence Systems and High-Performance Retrieval. In *Proceedings of the International Conference on Information Systems (ICIS 2004)*, S. 741–756, Washington DC, USA, 12.–15. Dezember 2004.
- [5] Böhm, K.; Härtwig, J.: Prozessorientiertes Wissensmanagement durch kontextualisierte Informationsversorgung aus Geschäftsprozessen. In: Ferstl, O. K.; Sinz, E. J.; Eckert, S.; Isselhorst, T. (Hrsg.): *Wirtschaftsinformatik 2005*. Heidelberg, 2005.
- [6] Budzik, J.; Hammond, K.: Watson: Anticipating and Contextualizing Information Needs. In: 62nd Annual Meeting of the American Society for Information Science, Medford, New Jersey, USA, 1999.
- [7] Baeza-Yates, R.; Ribeiro-Neto, B.: *Modern Information Retrieval*. Addison Wesley Longman, 1999.
- [8] Dumais, S.; Cutrell, E.; Sarin, R.; Horvitz, E.: Implicit Queries (IQ) for Contextualized Search. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2004)*, Sheffield, UK, 25.–29. Juli 2004.
- [9] Fensel, D.: *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*. Springer, Berlin Heidelberg, 2. Auflage, 2004.
- [10] Ferber, R.: *Information Retrieval: Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web*. dpunkt Verlag, Heidelberg, 2003.
- [11] Fagin, R.; Lotem, A.; Naor, M.: Optimal aggregation algorithms for middleware. *Journal of Computer and System Sciences*, 66(4):614–656, 2003.
- [12] Henrich, A.; Robbert, G.: Ein Ansatz zur Übertragung von Rangordnungen bei der Suche auf strukturierten Daten. In *Tagungsband der 10. GI-Fachtagung Datenbanksysteme für Business, Technologie und Web (BTW'03)*, Leipzig, 2003.
- [13] Jablonski, S.; Mayer, U.; Müller, S.: Integration of Different Information Sources using a Multidimensional Process based Knowledge Portal. 7. Konferenz zum Einsatz von

- Wissensmanagement in Wirtschaft und Verwaltung (KnowTech 2005), München, 24.–25. Oktober 2005.
- [14] Joachims, T.: Text categorization with support vector machines: learning with many relevant features. Proceedings of ECML-98, 10th European Conference on Machine Learning, Chemnitz, 1998.
- [15] Joachims, T.; Freitag, D.; Mitchell, T.M.: Web Watcher: A Tour Guide for the World Wide Web. In Proceedings of the 15th International Joint Conference on Artificial Intelligence (IJCAI), S. 770–777, 1997.
- [16] Lieberman, H.: Letizia: An Agent That Assists Web Browsing. In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95), S. 924–929, Montreal, Quebec, Kanada, 1995. Morgan Kaufmann.
- [17] Maus, H.: Workflow Context as a Means for Intelligent Information Support. In Proceedings of 3rd International Conference on Modeling and Using Context (CONTEXT'01), S. 261–274, 2001.
- [18] Meier, M.: Integration externer Daten in Planungs- und Kontrollsysteme – Ein Redaktions-Leitstand für Informationen aus dem Internet. Gabler, Wiesbaden, 2000.
- [19] Meerkamm, H.; Paetzold, K. (Hrsg.): FORFLOW. Bayerischer Forschungsverbund für Prozess- und Workflowunterstützung zur Planung und Steuerung der Abläufe in der Produktentwicklung. Erlangen, 2006.
- [20] Morgenroth, K.: Kontextbasiertes Information Retrieval – Modell, Konzeption und Realisierung kontextbasierter Information Retrieval Systeme. Dissertation, Otto-Friedrich-Universität Bamberg, 2006.
- [21] Oard, D.W.; Diekema, A.: Cross-Language Information Retrieval, in Annual Review of Information Science and Technology, Vol. 33, S. 223–256, 1998.
- [22] Pazzani, M.J.; Muramatsu, J.; Billsus, D.: Syskill & Webert: Identifying Interesting Web Sites. In Proceedings of the Thirteenth National Conference on Artificial Intelligence, S. 54–61, Portland, Oregon, USA, 1996.
- [23] Priebe, T.; Pernul, G.; Krause, P.: Ein integrativer Ansatz für unternehmensweite Wissensportale. In Tagungsband 6. Internationale Tagung Wirtschaftsinformatik (WI 2003), Dresden, September 2003.
- [24] Pfeifer, U.; Pennekamp, S.; Fuhr, N.: Incremental Processing of Vague Queries in Interactive Retrieval Systems. In Norbert Fuhr, Gisbert Dittrich, und Klaus Tochtermann, Hrsg., Tagungsband Hypertext - Information Retrieval - Multimedia '97: Theorien, Modelle und Implementierungen integrierter elektronischer Informationssysteme (HIM'97), Dortmund, 1997. Universitätsverlag Konstanz.
- [25] Rhodes, B.J.; Starner, T.: Remembrance Agent: A continuously running automated information retrieval system. In The Proceedings of The First International Conference on The Practical Application Of Intelligent Agents and Multi Agent Technology (PAAM '96), S. 487–495, London, UK, April 1996.
- [26] Resnick, P.; Iacovou, N.; Suchak, M., Bergstrom, P.; Riedl, J.: GroupLens: an open architecture for collaborative filtering of netnews. In Proceedings of the 1994 ACM conference on Computer supported cooperative work, S. 175–186, Chapel Hill, North Carolina, USA, 1994. ACM Press.
- [27] Rodríguez, M.A.; Egenhofer, M.J.: Determining semantic similarity among entity classes from different ontologies. IEEE Transactions on Knowledge and Data Engineering, 15(2):442–456, März/April 2003.
- [28] Salton, G.: The SMART Retrieval System – Experiments in Automatic Document Processing. Prentice-Hall, Englewood Cliffs, New Jersey, USA, 1971.

- [29] Schmitt, I.: Ähnlichkeitssuche in Multimedia-Datenbanken – Retrieval, Suchalgorithmen und Anfragebehandlung, Oldenbourg, München, 2005.
- [30] Schafer, J.B.; Konstan, J.A.; Riedl, J.: E-Commerce Recommendation Applications. *Data Min. Knowl. Discov.*, 5(1-2):115–153, 2001.
- [31] Shardanand, U.; Maes, P.: Social information filtering: algorithms for automating word of mouth. In *Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems*, S. 210–217, Denver, Colorado, USA, 1995.
- [32] Wilson, T.D.: Models in information behaviour research. *Journal of Documentation*, 55(3):249–270, 1999.
- [33] Wittig, F.: Maschinelles Lernen Bayes'scher Netze für benutzeradaptive Systeme. Dissertation, Universität des Saarlandes, 2002.

Prof. Dr. Andreas Henrich
Lehrstuhl für Medieninformatik
Fakultät Wirtschaftsinformatik und Angewandte Informatik
Otto-Friedrich-Universität Bamberg
Feldkirchenstraße 21, D-96045 Bamberg
Tel.: +49-951-863-2850
Fax: +49-951-863-2852
E-Mail: andreas.henrich@wiai.uni-bamberg.de
URL: <http://www.uni-bamberg.de/wiai/minf/>

Dr. Karlheinz Morgenroth
Lehrstuhl für Medieninformatik
Fakultät Wirtschaftsinformatik und Angewandte Informatik
Otto-Friedrich-Universität Bamberg
Feldkirchenstraße 21, D-96045 Bamberg
Tel.: +49-951-863-2853
Fax: +49-951-863-2852
E-Mail: karlheinz.morgenroth@wiai.uni-bamberg.de
URL: <http://www.uni-bamberg.de/wiai/minf/>