

September 2003

Such- und Klassifizierungsstrategien in elektronischen Produktkatalogen

Stefan Naumann

Fachhochschule Trier, naumann@umwelt-campus.de

Rolf Krieger

Fachhochschule Trier

Norbert Kuhn

Fachhochschule Trier

Cordula Schürmann

Fachhochschule Trier

Christian Sommer

Fachhochschule Trier

Follow this and additional works at: <http://aisel.aisnet.org/wi2003>

Recommended Citation

Naumann, Stefan; Krieger, Rolf; Kuhn, Norbert; Schürmann, Cordula; and Sommer, Christian, "Such- und Klassifizierungsstrategien in elektronischen Produktkatalogen" (2003). *Wirtschaftsinformatik Proceedings 2003*. 22.
<http://aisel.aisnet.org/wi2003/22>

This material is brought to you by the Wirtschaftsinformatik at AIS Electronic Library (AISeL). It has been accepted for inclusion in Wirtschaftsinformatik Proceedings 2003 by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

In: Uhr, Wolfgang, Esswein, Werner & Schoop, Eric (Hg.) 2003. *Wirtschaftsinformatik 2003: Medien - Märkte - Mobilität*, 2 Bde. Heidelberg: Physica-Verlag

ISBN: 3-7908-0111-9 (Band 1)

ISBN: 3-7908-0116-X (Band 2)

© Physica-Verlag Heidelberg 2003

Such- und Klassifizierungsstrategien in elektronischen Produktkatalogen

Stefan Naumann, Rolf Krieger, Norbert Kuhn, Cordula Schürmann, Christian Sommer

Fachhochschule Trier

Zusammenfassung: Die Suche nach geeigneten Produkten in Datenbanken und Internet stellt eine wesentliche Aufgabe in elektronischen Geschäftsprozessen dar. Dabei hängt die Qualität der Suchergebnisse im Wesentlichen von drei Faktoren ab: Genauigkeit der Anfrage, Struktur und Qualität des Produktdatenbestandes und Art des Suchverfahrens. In dieser Arbeit werden Methoden vorgestellt, mit denen merkmalsbasierte Produktkataloge durch die Kombination von Methoden des Information Retrieval mit fuzzyfizierten Ansätzen durchmustert werden. In einem mehrstufigen Verfahren wird zunächst die Anzahl der zu bewertenden Produkte eingeschränkt. Anschließend wird pro Merkmal jedes Ergebniskandidaten ein Gewicht anhand von Zugehörigkeitsfunktionen der Fuzzy-Technologie bestimmt, das schließlich zu einer Produktrelevanz zusammengefasst wird. Das Verfahren lässt sich auch zur automatischen Klassifikation vorhandener Produkte einsetzen.

Schlüsselworte: Produktsuche, Produkt-Retrieval, elektronischer Produktkatalog, Klassifikationssystem, automatische Klassifizierung, e-commerce

1 Einführung

Im Internet existiert mittlerweile eine Vielzahl von elektronischen Markt- und Verkaufsplätzen. Ein maßgeblicher Erfolgsfaktor dieser - im Regelfall datenbankgestützten - Marktplätze ist, wie leicht der Marktplatz für Kunden und Anbieter nutzbar ist. Aus Kundensicht ist vor allem wichtig, wie gut sich geeignete Produkte bewertet finden lassen. Ein Suchvorgang findet dabei üblicherweise merkmalsbasiert statt: Der Kunde legt die Eigenschaften eines Produkts fest, die durch Merkmale und ihre Werte vorgegeben werden. Aufgabe des Produkt-Retrieval ist es, anhand geeigneter statistischer und semantischer Methoden die Produkte in der Reihenfolge ihrer Relevanz aufzulisten. Die Qualität der Ergebnisse hängt dabei von der Genauigkeit der Anfrage, der Struktur des Produktkataloges und dem gewählten Suchverfahren ab.

Als Anwendungsbereich betrachten wir in dieser Arbeit Produktkataloge gebrauchter Industriemaschinen, da sich diese durch eine Vielzahl von Merkmalen und Merkmalsarten auszeichnen. Unsere Ergebnisse lassen sich aber auch auf andere merkmalsbasierte Produkt- und Dienstleistungskataloge übertragen.

1.1 Aufbau des Beitrages

Nachdem wir auf die Motivation unserer Untersuchungen und das wissenschaftliche Umfeld eingegangen sind, erläutern wir in Abschnitt 2 zunächst grundlegende Begriffe zur Klassifikation von Produkten. Abschnitt 3 stellt verschiedene Suchstrategien vor. Wir betrachten dabei insbesondere die Retrieval-Verfahren vektorielle Suche, erweiterte boolesche Suche und Fuzzy-Retrieval. Für die letzte Kategorie stellen wir einen Ansatz zur Bestimmung von Ähnlichkeiten zwischen Anfragen und Merkmalswerten vor. Wir schließen unseren Beitrag mit einer Zusammenfassung und einem Ausblick.

1.2 Motivation

Ein typischer Marktplatz für gebrauchte Industriemaschinen (z. B. [Surp03; FDM03]) enthält bei einem Datenbestand von etwa 20.000 Produkten rund 500 hierarchisch strukturierte Produktgruppen und über 500 verschiedene Merkmale. Allgemeinere elektronische Klassifikationssysteme wie eClass [IDW03] kommen auf über 5.000 Merkmale in 15.000 Produktgruppen.

Hinsichtlich der angebotenen Suchverfahren in elektronischen Handelsbörsen bieten die meisten von uns evaluierten Marktplätze und Anbieter allerdings lediglich eine vereinfachte Suche entlang der Produktgruppenhierarchie an oder beschränken sich auf wenige, allgemeingültige Merkmale wie Hersteller, EAN-Code oder Typ. Oft steht dem Anfragenden aber die genaue Kennzeichnung eines Produktes (wie beispielsweise der EAN-Code) nicht zur Verfügung. Er muss daher das Produkt über seine Merkmale beschreiben.

Um eine unscharfe Suchanfrage automatisiert beantworten zu können, bedarf es Techniken, die in hohem Maße die Struktur des Klassifikationssystems und das in den Produktkatalogen durch die Produktmerkmale und -werte gegebene Anwendungswissen berücksichtigen. Dazu müssen semantische Ansätze mit Algorithmen und Konzepten des auf unstrukturierten Daten basierenden Information Retrieval kombiniert werden.

In vorausgegangenen Untersuchungen [Kri⁺02] hat sich herauskristallisiert, dass es keine einheitlich standardisierten Formate und Beschreibungen für Klassifikationssysteme und elektronische Produktkataloge gibt. Suchstrategien, die weite Verbreitung finden sollen, können sich daher nicht auf einen Standard wie beispielsweise BMEcat [BME02] oder eClass beschränken.

Infolgedessen setzen wir unsere Such- und Klassifizierungsstrategien auf eine allgemeine, merkmalsbasierte Struktur auf, die Produkte durch eine Referenzhierarchie klassifiziert.

1.3 Verfahren des Information Retrieval

Verfahren des Information Retrieval haben mit der zunehmenden Nutzung des Internets stark an Bedeutung gewonnen. Ohne Suchmaschinen wie google, altavista, yahoo oder fireball ließe sich häufig die im Internet verstreute Information kaum finden. Bei Eingabe einer Liste von Suchbegriffen liefern diese Systeme eine Menge von URLs als Verweis auf Dokumente, die für den Suchenden relevant sein könnten. Gemäß einer Klassifikation von van Rijsbergen [Rijs79] handelt es sich damit um Verfahren zum Text-Retrieval. Kennzeichen dieser Suchverfahren ist - neben dem gelieferten Ergebnis in Form von kompletten Dokumenten oder Textteilen - die weitgehend unstrukturierte Repräsentation der Dokumentbasis. Aus den Dokumenten wird daher im Regelfall ein Index generiert, d.h. ein Stichwortverzeichnis bestehend aus Indextermen, die wesentliche Inhalte der Dokumente in komprimierter Form darstellen sollen. Dokumente werden als Liste von Indextermen beschrieben, wobei die Relation zwischen einem Indexterm und einem Dokument häufig zusätzlich gewichtet wird, um ein Maß der Wichtigkeit auszudrücken [Salt71; SaMc83].

Ein Problem der Verfahren des Text-Retrieval ist die oft unbekannt, weil beliebige Struktur der zu suchenden Dokumente. Dies ändert sich, wenn der Suchraum nicht mehr das Internet, sondern eine (oder auch mehrere) bekannte Datenbanken umfasst. Die Suche hierin bezeichnet van Rijsbergen als Data Retrieval, manchmal auch Faktenretrieval genannt [Fuhr00]. Die Anfrage ist gegeben als eine Menge von Attribut-/Wertpaaren, welche die Objekte in der angefragten Datenbank möglichst gut erfüllen sollen.

Für unsere vorliegende Arbeit gehen wir davon aus, dass Benutzer Produkte suchen, die in der Datenbank einer Handelsbörse oder eines Marktplatzes abgelegt sind. Insofern handelt es sich bei unserem Ansatz zum Produkt-Retrieval um eine Realisierung eines Data Retrieval Verfahrens, wobei wir jedoch auch Komponenten des klassischen Information Retrievals verwenden.

Das von uns vorgestellte Retrievalmodell setzt sich mit den drei Teilaufgaben auseinander, die generell in jedem Modell behandelt werden müssen. Dies sind die Konstruktion einer Repräsentationsform der Anfrage, die Konstruktion einer Repräsentation der Produktmenge sowie die Bereitstellung einer Funktion zur Berechnung der Übereinstimmung (Ähnlichkeitsmaß) zwischen einer Anfrage und einzelnen Produkten. Die Ähnlichkeitsmaße, die wir vorstellen, verwenden Zugehörigkeitsfunktionen (wie sie aus der Theorie der Fuzzymengen bekannt sind, siehe z. B. [Zade65; Grau95]) und Rankingfunktionen (wie sie im Information Retrieval Verwendung finden, z. B. in [RiBa99]).

2 Merkmalsbasierte Produktkataloge

Grundlage für die Darstellung der Produkte in einem Produktkatalog ist eine geeignete Klassifikation anhand von Merkmalen, mittels der die Produkte eindeutig den verschiedenen Produktgruppen zugeordnet werden können. Je genauer und detaillierter die Merkmalsdefinitionen und die Zuordnung der Merkmale zu den einzelnen Produktgruppen und damit zu den Produkten erfolgt, desto gezielter können für den Suchvorgang in Produktkatalogen geeignete Strategien und Algorithmen ausgewählt werden. Die für diese Suchstrategien relevanten Informationen werden in den Merkmalsdefinitionen und deren Ausprägungen hinterlegt. Detailliertere Ausführungen zu Klassifikationssystemen und Produktkatalogen bieten [Leu⁺02; Kri⁺02; OtBe01].

2.1 Notationen

Definition 1: Produkt-Klassifikationssystem

Ein *Klassifikationssystem* \mathbf{K} ist ein Tupel $(\mathbf{G}, \mathbf{M}, \mathbf{W})$ mit

1. \mathbf{G} ist die Menge aller im Klassifikationssystem vorhandenen *Produktgruppen*. Die einzelnen Produktgruppen sind als Baum hierarchisch aufgebaut. Den Produktgruppenbaum bezeichnet man als *Referenzhierarchie*.
2. \mathbf{M} ist die Menge der im System vorhandenen *Merkmale*. Die Merkmale werden den Produktgruppen zugeordnet und bilden gemeinsam eine *Sachmerkmaleiste* der Produktgruppe. Eine Produktgruppe erbt alle Merkmale von Vorgänger-Produktgruppen innerhalb der Referenzhierarchie.
3. Die Menge \mathbf{W} enthält für jedes Merkmal \mathbf{m} eine Menge \mathbf{W}_m der möglichen *Merkmalswerte*. Ein Merkmalswert kann semantisch einem *Einzelwert*, einem (ggf. mehrdimensionalen) *Bereich* oder einer *Aufzählung von Einzelwerten* entsprechen.

Definition 2: Produktkatalog

Ein *Produktkatalog* besteht aus einem Klassifikationssystem \mathbf{K} und einer Menge \mathbf{P} von Produkten. Die Produkte sind den Produktgruppen des Klassifikationssystems zugeordnet. Sie werden über die Merkmale beschrieben, die dieser Produktgruppe zugeordnet sind. Für ein Produkt \mathbf{p} bezeichnen wir den Merkmalswert für ein Merkmal \mathbf{m} mit $\mathbf{w}_m(\mathbf{p})$. Wird ein Produkt angefragt, bezeichnen wir den Merkmalswert in der Anfrage \mathbf{q} mit $\mathbf{w}_m(\mathbf{q})$.

2.2 Merkmale, Skalierungen und Wertebereiche

Unsere Such- und Klassifizierungsstrategien nutzen das Klassifikationssystem und insbesondere die Produktmerkmale mit ihren Werten als indizierende Kennzeichen eines Produktes. Hierfür ist es erforderlich, dass die Merkmale selbst eindeutig definiert werden. Die Basis bietet hierfür eine Beschreibung anhand von Attributen, wie sie beispielsweise [DIN98] vorsieht.

In statistischen Modellen wird einem Merkmal zusätzlich eine Skalierung (Nominalskala, Ordinalskala, Intervallskala und Verhältnisskala [Ble⁺96]) zugeordnet. Als Sonderfall der Skalierung betrachten wir die Merkmale, die Auflistungen von Einzelwerten oder Bereiche als Werte annehmen können, wobei die Einzelwerte bzw. die Bereichsgrenzen ordinal-, intervall- oder verhältnisskaliert sind. Das Merkmal "Drehzahlbereich" beispielsweise hat als Ausprägungen Bereiche, deren Grenzen Ausprägungen des Merkmales "Drehzahl" sind.

Attribut		Ausprägung		
Kennung		XXX003003		
Name		Zustand		
Synonym		Beschaffenheit		
Definition		Genauere Beurteilung des Zustandes der Maschine		
Wertkodierung		Gut	Mittel	Schlecht
Wertbedeutung		Regelmäßige Wartung, sofort einsatzfähig	Unvollständiger Wartungszyklus, grundsätzlich einsatzfähig	Offensichtliche Mängel, reparaturbedürftig
Skalierung		ordinalskaliert		
Zugehörig- keitsfunktion	Typ	Trapez		
	Glättungs- bereich	Absolut		

Tabelle 1: Merkmal „Zustand“ mit ausgewählten Attributen und deren Ausprägungen

Wie wir in Abschnitt 3 sehen werden, spielt die Skalierung bei der Realisierung von Suchstrategien eine wesentliche Rolle und muss als zusätzliches Attribut bei der Definition eines Merkmals berücksichtigt werden. Des Weiteren wird die Attributmenge eines Merkmals um das Attribut „Zugehörigkeitsfunktion“, das für das fuzzybasierte Produkt-Retrieval benötigt wird, ergänzt. Das Attribut legt den Typ der Funktion und ihren Glättungsbereich fest. Tabelle 1 zeigt auszugsweise die Definition eines Merkmals, das den Zustand einer Maschine beschreibt.

3 Suchstrategien in Produktkatalogen

Ausgehend von einer einheitlichen Darstellung der Anfrage - zur allgemeineren Nutzung kodiert in ein XML-Schema - stellen wir in diesem Abschnitt unsere Suchstrategien vor. Die Strategien können auch verwendet werden, um vorhandene Produkte zu klassifizieren.

Wir stellen zwei Ansätze vor: Der erste Ansatz basiert auf dem Vektorraum-Modell, wobei Merkmals-/Wertpaare als Indexterme aufgefasst werden. Der zweite Ansatz nutzt das erweiterte boolesche Modell als Rankingverfahren, wobei die Ähnlichkeit für einzelne Merkmale über Fuzzy-Zugehörigkeitsfunktionen bestimmt wird. Abschließend werden die beiden Ansätze vergleichend betrachtet.

3.1 Gestaltung der Anfrage

Die Beschreibung einer Produkthanfrage stützt sich auf eine Liste von Merkmalen mit entsprechenden Wertausprägungen. Für eine genaue Beschreibung komplexer Produkte und eine Einschränkung des Suchraumes muss es möglich sein, die Merkmale mit ihren Werten über logische Operatoren zu verknüpfen. Eine Gewichtung einzelner angefragter Merkmale muss ebenfalls möglich sein. Als Beispiel kann die folgende Suchanfrage dienen.

```
"Drehmaschine"
  and {"Drehzahlbereich (U/min)" = "[1000 .. 5000]"}erfüllt;1.0
  and {"Drehlänge(mm)" = "800"}1.0
  and {"Gesamtantriebsleistung (kW)" = "[20 .. 25]"}1.0
  and {"Hersteller" = "Okuma"}1.0
  and {"Herstellungsjahr" = "1985"}0.7
  and {"Preis (EUR)" = "40000"}1.0
```

Abbildung 1: Beispielanfrage einer gebrauchten Maschine

Diese Suchanfrage beschreibt eine Drehmaschine mit folgenden Eigenschaften:

- Der Drehzahlbereich soll 1000 bis 5000 U/min betragen. "Erfüllt" ("fulfilled") bedeutet, dass mindestens dieser Bereich abgedeckt werden muss.
- Die Drehlänge soll 800 mm betragen.
- Die Gesamtantriebsleistung soll zwischen 20 kW und 25 kW liegen.
- Der Hersteller der Maschine soll die Fa. Okuma sein.
- Das Herstellungsjahr der Maschine soll 1985 sein.

- Der Preis soll bei 40.000 EUR liegen.

Der logische Operator `and` drückt aus, dass alle Merkmals-/Wertpaare erfüllt sein sollen. Bis auf das Herstellungsjahr, das mit 0.7 gewichtet ist, sollen sie bei der Suche mit dem Gewicht 1.0 als gleichgewichtet betrachtet werden.

Für die maschinelle Verarbeitung in E-Business-Szenarien ist es vorteilhaft, eine Suchanfrage als XML-Dokument zu kodieren. Hierzu wurde ein XML-Schema erstellt, das Abbildung 2 beschreibt. Das XML-Dokument für eine Suchanfrage beinhaltet folgende Informationen:

- Auflistung der Bezeichnungen der Produktgruppen, denen das gesuchte Produkt angehören könnte (<GROUPS>-Element).
- Angabe von Merkmalen mit Einzelwerten oder Wertebereichen. Ein Merkmal mit dem entsprechenden Wert ist dem <F_LITERAL>-Element zugeordnet und wird im Folgenden als Literal bezeichnet. Ein Merkmal muss eindeutig identifizierbar sein.
- Verknüpfung von Literalen zu Ausdrücken über die logischen Operatoren `and`, `or` und `not`.
- Gewichtung von Literalen und Ausdrücken. Über die Gewichtung kann die Bedeutung einzelner Literale oder Ausdrücke bei der Suche beeinflusst werden.
- Angabe, wie Bereichsanfragen semantisch zu handhaben sind: `FULFILLED` (der Anfragebereich muss mindestens ausgefüllt sein), `COMPRISED` (der Produktbereich soll im Anfragebereich liegen) oder `EXACT` (der Anfragebereich soll weder unter- noch überschritten, sondern genau getroffen werden).

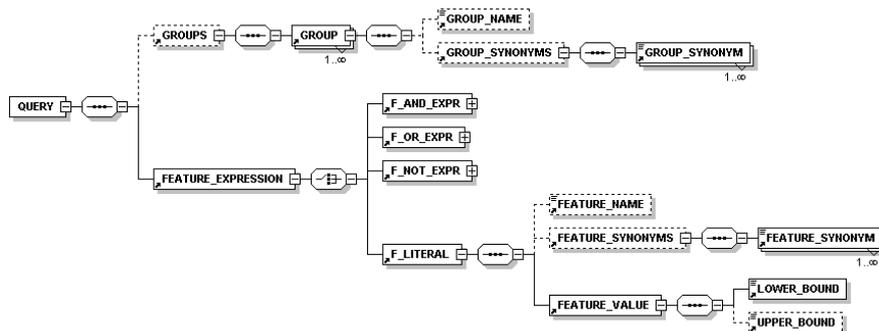


Abbildung 2: XML-Schema einer Anfrage

Zusätzlich können neben den Merkmals- und Gruppennamen auch Synonyme angegeben werden, da es keine einheitliche Normung für Produktmerkmalsnamen und ihre Eigenschaften gibt. Hier sind neben bestehenden Klassifizierungssysteme-

men wie eClass übergreifende Normierungsbemühungen zu erkennen [DIN02]. Bei bekannten Klassifikationssystemen können auch eindeutige Bezeichner für die Produktgruppen und Merkmale angegeben werden.

Abbildung 3 zeigt auszugsweise die Suchanfrage aus Abbildung 1 als XML-Dokument gemäß dem beschriebenen Schema.

```
<QUERY>
  <FEATURE_EXPRESSION>
    <F_AND_EXPR weight="1.0">
      <F_LITERAL feature_id="124" weight="1.0">
        <FEATURE_NAME>
          Drehzahlbereich(U/min)
        </FEATURE_NAME>
        <FEATURE_VALUE>
          <LOWER_BOUND>1000</LOWER_BOUND>
          <UPPER_BOUND>5000</UPPER_BOUND>
        </FEATURE_VALUE>
      </F_LITERAL>
      ...
      <F_LITERAL feature_id="290" weight="1.0">
        <FEATURE_NAME>
          Preis
        </FEATURE_NAME>
        <FEATURE_VALUE>
          <LOWER_BOUND>40000</LOWER_BOUND>
        </FEATURE_VALUE>
      </F_LITERAL>
    </F_AND_EXPR>
  </FEATURE_EXPRESSION>
</QUERY>
```

Abbildung 3: Auszüge einer XML-kodierten Maschinenanfrage entsprechend Abb. 1

3.2 Suchstrategie

Unser Ansatz zur Optimierung der Produktsuche berücksichtigt Klassifikationssysteme und Wissen über die Anwendungsdomäne und kann im Ablauf wie folgt beschrieben werden:

1. Erstellung einer Suchanfrage auf Basis des Klassifikationssystems, beispielsweise über ein Web-Formular oder einen Anfragegenerator.
2. Überführung der Suchanfrage in ein XML-Anfragedokument.
3. Vorauswahl der Produktgruppen und Produkte, für welche ein numerisches Ranking erstellt werden soll.
4. Ermittlung der Ähnlichkeiten zwischen Anfragewert und Produktwert für jedes Merkmal eines Produktes anhand hinterlegter Zugehörigkeitsfunktionen oder statistischer Werte. Dabei werden auch Informationen wie Skalierung oder Glättungsbereich berücksichtigt.
5. Zusammenfassung der Ähnlichkeiten auf Merkmalsebene zu einer Gesamtrelevanz für jedes Produkt. Dabei müssen anfragespezifische Gewichte berücksichtigt werden.
6. Sortierung der Produkte entsprechend ihrer Relevanz.

7. Optionale Bewertung des Ergebnisses durch den Anfragenden ("Relevance Feedback").

3.3 Notationen

Definition 3: Produkt-Retrieval-Modell

Ein *Produkt-Retrieval-Modell* besteht aus einem Tupel $(\mathbf{K}, \mathbf{P}, \mathbf{Q}, \mathbf{R}, \mathbf{I})$ mit

1. \mathbf{K} und \mathbf{P} repräsentieren den aus dem Klassifikationssystem und der Produktmenge bestehenden *Produktkatalog*.
2. \mathbf{Q} ist eine Menge von *Anfragen* nach Produkten. Eine Anfrage besteht aus gesuchten Produktgruppen, Merkmalen und Merkmalswerten, die über logische Operatoren verknüpft sind
3. \mathbf{R} repräsentiert eine Menge von *Rankingfunktionen* $\mathbf{r}: \mathbf{Q} \times \mathbf{P} \rightarrow [0,1]$, die jedem Produkt \mathbf{p} in Abhängigkeit von einer Anfrage \mathbf{q} eine Relevanz zwischen 0 und 1 zuweisen.
4. \mathbf{I} repräsentiert die Menge der *Indexterme* eines Retrieval-Systems. Jedes Produkt wird durch einen oder mehrere Indexterme charakterisiert.

3.4 Vektorraum-Modell

Eines der bekanntesten und anschaulichsten Verfahren im Information Retrieval ist das Vektorraum-Modell. Hierbei wird versucht, eine Ähnlichkeit zu bestimmen, indem Anfrage und Dokument als Vektoren aufgefasst werden. Als Ähnlichkeitsmaß wird der Winkel zwischen Dokument- und Anfragevektor verwendet, der über die Cosinusfunktion auf einen Wert im Intervall $[0,1]$ abgebildet wird [RiBa99]. In den folgenden Unterabschnitten übertragen wir dieses Modell auf die Produktsuche.

3.4.1 Indexterme

Im Information Retrieval werden Dokumente durch die in ihnen enthaltenen Worte indiziert. Jedes Dokument wird durch eine Liste von solchen Wort-Indextermen repräsentiert. Die Relevanz bezüglich eines Terms hängt von der Gesamthäufigkeit des Indexterms bezogen auf alle Dokumente ab und wird als inverse document frequency (idf) bezeichnet. Ein Dokument hat bezüglich eines Anfrageterms (Wortes) dann eine hohe Relevanz, wenn der Anfrageterm über alle Dokumente selten, aber in dem Dokument selbst besonders häufig im Verhältnis zu anderen Indextermen vorkommt.

Die Analyse von Produktkatalogen hat ergeben, dass das obige Konzept der Indizierung auf die Kataloge übertragbar ist. Dazu werden alle Produkte als Doku-

mente interpretiert. Wie in Abschnitt 2 erläutert, werden diese Produkte insbesondere durch ihre Zuordnung zu einer Produktgruppe, ihre Merkmale und den Ausprägungen dieser Merkmale charakterisiert. Es können daher sowohl die Merkmale als auch Merkmals-/Wertpaare der Produkte als Indexterme gewählt werden. In unserem Ansatz verwenden wir als Indexterme Merkmals-/Wertpaare.

Analog zur inverse document frequency (idf) beim Text-Retrieval definieren wir die „inverse product frequency“:

Definition 4: Inverse Product Frequency (ipf)

Unter der *ipf* verstehen wir die *inverse product frequency* und setzen $ipf_i = \log(|P|/f_i)$, wobei f_i die Häufigkeit des Indexterms i über alle Produkte und $|P|$ die Gesamtanzahl der Produkte im Produktkatalog angibt.

Tabelle 2 zeigt einen Auszug der Merkmalswerte aus einer Produktdatenbank mit etwa 8000 Maschinen. Hier ist zu erkennen, dass verschiedene Merkmals-/Wertpaare deutlich häufiger in Produkten vorkommen als andere und entsprechend die **ipf** beeinflussen.

Häufigkeit	Merkmalsname	Merkmalswert	ipf
5	Drehzahlbereich(U/min)	[3..2500]	3,204
7	Drehzahlbereich(U/min)	[25..5000]	3,058
7	Drehlänge(mm)	190	3,058
21	Drehlänge(mm)	800	2,581
38	Gesamtantriebsleistung(kW)	22	2,323
77	Gesamtantriebsleistung(kW)	30	2,017
15	Hersteller	Okuma	2,727
134	Herstellungsjahr	1983	1,776
162	Herstellungsjahr	1985	1,694
2	Preis	40.000	3,602

Tabelle 2: Auszug aus einem Produktindex mit Merkmals-/Wertpaaren und deren ipf

3.4.2 Rankingfunktion

Unter Verwendung der **ipf** wird an Anlehnung an [RiBa99] mit folgender Funktion die Gesamtrelevanz für ein Produkt bestimmt.

Definition 5: Rankingfunktion

1. Das *Gewicht* des k -ten Indexterms für ein Produkt p wird berechnet durch

$$weight_k(p) = \begin{cases} ipf_k & \text{falls Indexterm im Produkt } p \text{ enthalten} \\ 0 & \text{sonst} \end{cases} \quad (1)$$

und bestimmt den *Gewichtsvektor* des Produktes

$$\vec{p} = (weight_1(p), \dots, weight_{|I|}(p)) \quad (2)$$

2. Eine *Anfrage* (Query) \mathbf{q} besteht aus einem *Gewichtsvektor* $\vec{q} = (\text{weight}_1(q), \dots, \text{weight}_n(q))$, der die Gewichte der angefragten Merkmals-/Wertpaare repräsentiert. Der Anfragende kann einzelnen Merkmals-/Wertpaaren Gewichte zuordnen, wobei das Defaultgewicht 1 ist. Boolesche Operatoren finden keine Verwendung.

3. Die *Rankingfunktion* \mathbf{r} bestimmt den Cosinus zwischen Anfragevektor und

$$\text{Produktmerkmalsvektor und ist definiert durch } r(\vec{q}, \vec{p}) = \frac{\vec{q} \cdot \vec{p}}{|\vec{q}| \cdot |\vec{p}|}$$

Ausgehend von diesen Berechnungsgrundlagen zeigt Tabelle 3 das Ergebnis der vektoriellen Suche für die Beispielanfrage aus Abschnitt 3.1. Die Tabelle zeigt für jedes Produkt den Gewichtsvektor, der sich über die in Definition 5 angegebene Formel bestimmt. Die von 0 verschiedenen Gewichte entsprechen der **ipf** (siehe Tabelle 2).

	Drehzahlbereich (U/min)	Drehlänge (mm)	Gesamtantriebsleistung(kW)	Hersteller	Herstellungsjahr	Preis (€)	Ranking
Anfrage	[1000..5000]	800	[20..25]	Okuma	1985	40000	
<i>Gewichtsvektor</i>	1	1	1	1	0,7	1	
Maschine 7	[15..3000]	800	30	Okuma	1985	32000	
<i>Gewichtsvektor</i>	0	2,581	0	2,727	1,694	0	0,6728231
Maschine 5	[15..3000]	800	30	Okuma	1985	66539	
<i>Gewichtsvektor</i>	0	2,581	0	2,727	1,694	0	0,6728231
Maschine 4	[35..7500]	200	22	Index	1985	40000	
<i>Gewichtsvektor</i>	0	0	2,323	0	1,694	3,602	0,6584932
Maschine 3	[9..1800]	800	22	Gildemeister	1983		
<i>Gewichtsvektor</i>	0	2,581	2,323	0	0	0	0,6027381
Maschine 2	[20..5000]	190	11	EMCO	1992	26000	
<i>Gewichtsvektor</i>	0	0	0	0	0	0	0
Maschine 6	[20..4000]	490	30	Traub	1991	47500	
<i>Gewichtsvektor</i>	0	0	0	0	0	0	0
Maschine 1	[3..2500]	1000	15	Boehringer	1992	40900	
<i>Gewichtsvektor</i>	0	0	0	0	0	0	0

Tabelle 3: Suchergebnisse mit vektoriellem Ranking

Die Maschinen 7, 5 und 4 haben die meisten angefragten Merkmals-/Wertpaare und sind daher am höchsten bewertet. In den Suchergebnissen ist auch deutlich zu sehen, dass viele Indexterme mit 0 gewichtet werden, da der angefragte Merkmalswert nicht mit dem Produkt-Merkmalwert übereinstimmt.

3.5 Erweitertes Fuzzy-Retrieval

In der Fuzzy-Logik wird die Ermittlung einer Mengenzugehörigkeit über eine Zugehörigkeitsfunktion bestimmt. Im Gegensatz zur Mengenzugehörigkeit im rein booleschen Sinne (mit den Werten "gehört zur Menge" oder "gehört nicht zur Menge") kann eine Fuzzymenge auch eine Zugehörigkeit im Intervall [0..1] an-

nehmen. Fuzzyfizierung bietet sich daher an, um die Ähnlichkeit eines Produktes zu einer Anfrage bezüglich eines Merkmales zu bestimmen.

Diese Ähnlichkeiten können über eine Rankingfunktion zusammengefasst werden, die in unserem Fall auf dem erweiterten booleschen Modell [Sal⁺83] beruht.

3.5.1 Ähnlichkeitsbestimmung durch Zugehörigkeitsfunktionen

Definition 6: Zugehörigkeitsfunktionen

Z ist die Menge der *Zugehörigkeitsfunktionen* im Produkt-Retrieval-System. $z: W_m \times W_m \rightarrow [0,1]$ steht dabei für die Zugehörigkeitsfunktion, welche die Ähnlichkeit zwischen einem Anfragewert und einem Produktwert bezüglich eines Merkmals m in Abhängigkeit von seiner Produktgruppenzugehörigkeit berechnet. Der konkrete Funktionsverlauf der Basisfunktion wird bestimmt durch die Werte der Anfrage. *Glättungsfaktoren* geben zusätzlich an, wie stark die Funktion abfällt: Die Zugehörigkeitsfunktion wird in diesen ähnlichkeitsrelevanten Bereich gestreckt/gestaucht. Der Bereich ergibt sich prozentual oder absolut aus dem Anfragewert und wird auch als *Glättungsbereich* bezeichnet. Ein Merkmal Baujahr beispielsweise erfordert einen Glättungsbereich, der absolut angegeben wird (z. B. +/- 5 Jahre), eine Drehzahl kann durchaus prozentual abgewertet werden (z. B. +/- 30%). Auch der Betrag einer Bereichsanfrage kann den Glättungsbereich bestimmen.

Für einen Anfragewert $w_m(q)$ schreiben wir vereinfacht für die angepasste Zugehörigkeitsfunktion $z_q: W_m \rightarrow [0,1]$. Diese Funktion errechnet direkt für einen Produktwert die Ähnlichkeit zum Anfragewert.

Der Typ der Zugehörigkeitsfunktion wird neben der domänenabhängigen Semantik beeinflusst durch die Skalierung des Merkmals.

Nominalskalierte Merkmale

Nominalskalierte Merkmalswerte stehen nicht in einer Reihenfolge zueinander. Die Zugehörigkeitsfunktion wird durch eine Korrelationsmatrix [Oga⁺91, RiBa99] definiert.

Merkmale mit linear geordneten Werten

Bei ordinal-, intervall- und verhältnisskalierten Merkmalswerten kann als Zugehörigkeitsfunktion beispielsweise die Rechteck- und Trapezfunktion, die Hyperbelfunktion oder die Gauss'sche Glocke [Grau95] bzw. Kombinationen davon verwendet werden.

Generell unterscheiden wir folgende Fälle:

1. Die Anfrage liegt als Einzelwert vor. Es soll also eine Ähnlichkeit zwischen zwei Einzelwerten bestimmt werden.

2. Die Anfrage liegt bereits benutzerseitig als Bereich vor (ist daher vereinfacht fuzzyfiziert). Ein maximale Ähnlichkeit liegt vor, wenn der gesuchte Produktwert innerhalb des gesuchten Bereiches liegt.
3. Der Produktwert soll oberhalb oder unterhalb des Anfragewertes liegen.

In allen Fällen kann die Zugehörigkeit unmittelbar durch die Zugehörigkeitsfunktion z bestimmt werden, wobei für den zweiten Fall innerhalb des Bereichs der Anfrage eine Ähnlichkeit von 1 angenommen wird. Im dritten Fall ist die Ähnlichkeit oberhalb (unterhalb) des Anfragewertes mit 1 zu bewerten.

Beispielsweise lässt sich die Zugehörigkeitsfunktion zum Merkmal Preis, die basierend auf einer geeigneten Funktion z die Zugehörigkeit eines Produktes p zu einer Anfrage q berechnet, wie folgt definieren:

$$z_{preis}(w_{preis}(p)) = \begin{cases} 1 & \text{falls } w_{preis}(p) \leq w_{preis}(q) \\ z(w_{preis}(p)) & \text{sonst} \end{cases} \quad (3)$$

Unterhalb des Anfragewertes ist ein Produktpreis immer ein „Volltreffer“, oberhalb wird er gemäß der Zugehörigkeitsfunktion abgewertet.

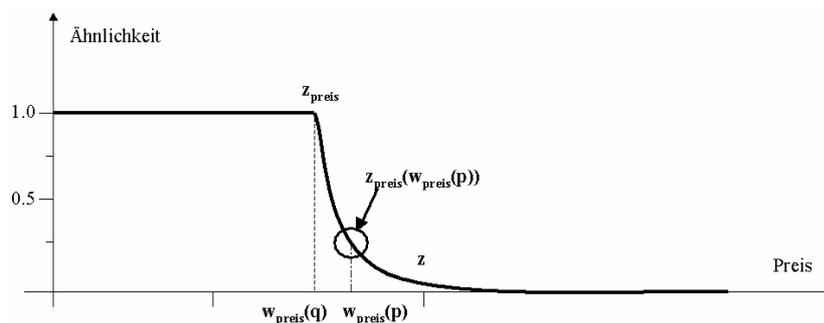


Abbildung 4: Zugehörigkeitsfunktion des Merkmals "Preis"

Bereichsfähige Merkmale

Als ein Sonderfall der nominalskalierten Merkmale können Merkmale aufgefasst werden, deren Werte Bereiche bilden können. Das ist grundsätzlich möglich, wenn die Merkmalswerte linear geordnet werden können. Beispiel hierfür ist das Merkmal Drehzahlbereich. Bereiche an sich sind nominalskaliert, da nicht gesagt werden kann, ob beispielsweise der Bereich 3.000-5.000 Umdrehungen pro Minute größer ist als der Bereich 2.000-4.000 U/min.

Da jedoch die Bereichsgrenzen vergleichbar sind (siehe 2.2), lassen sich die im vorigen Abschnitt beschriebenen Zugehörigkeitsfunktionen auch zur Bestimmung einer Ähnlichkeit zwischen Bereichen verwenden. Um dies zu verdeutlichen, be-

trachten wir den Bereich $w_m(q)=[u_q, o_q]$, der dem Merkmal m in der Anfrage zugeordnet ist, und den Bereich $w_m(p)=[u_p, o_p]$, der den Merkmalswert des Produktes repräsentiert. Den Bereichsgrenzen werden die Zugehörigkeitsfunktionen z_u und z_o zugeordnet. Folgende Fälle können unterschieden werden (angegeben ist jeweils, welche Bedingung die höchste Ähnlichkeit hat, in Klammern der in 3.1 eingeführte Begriff):

1. $w_m(q) \subseteq w_m(p)$ (fulfilled)

Der Merkmalsbereich soll die Anfrage komplett einschließen, die Anfrage soll also mindestens erfüllt werden. Wie stark die Übererfüllung ist, ist irrelevant.

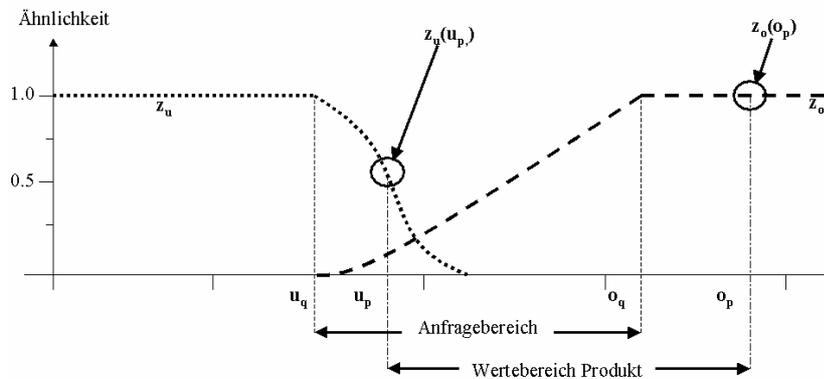


Abbildung 5: Ähnlichkeit bei Erfüllung des Anfragebereiches

2. $w_m(q) \supseteq w_m(p)$ (comprised)

Der Merkmalswert soll innerhalb der Anfrage liegen, die Anfrage ist also bereits vereinfacht fuzzyfiziert.

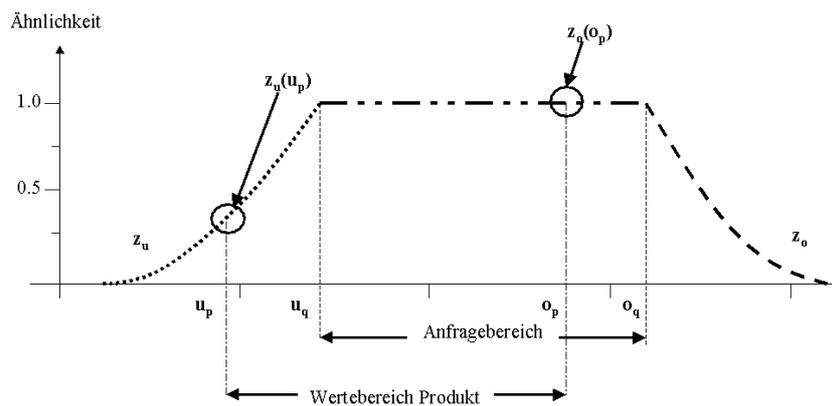


Abbildung 6: Ähnlichkeit bei Enthaltung des Produktes im Anfragebereich

3. $w_m(q) = w_m(p)$ (exact)

Der Produktbereich soll den Anfragebereich möglichst genau treffen, Über- und Untererfüllung führt zu Abwertung.

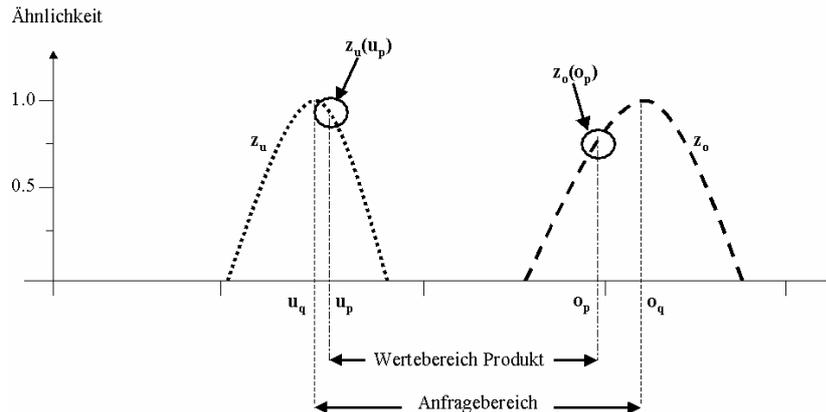


Abbildung 7: Ähnlichkeit bei exakter Übereinstimmung

Die Ähnlichkeit zwischen Anfragebereich und Produktbereich ergibt sich in allen 3 Fällen mit

$$z(u_p, o_p) = z_u(u_p) \cdot z_o(o_p) \quad (4)$$

Neben der Produktbildung sind auch andere Verfahren wie Flächenverhältnisse oder Mittelwertbildung zur Berechnung der Ähnlichkeit zwischen Anfrage- und Merkmalsbereichen einsetzbar. Welches Verfahren verwendet wird, hängt auch von der Anwendungsdomäne ab.

Entsprechend lassen sich auch Ähnlichkeiten zwischen komplexen Merkmalswerten (wie beispielsweise einem Raummaß) bestimmen.

Mehrfachwerte

Einen weiteren Sonderfall der Nominalskalierung stellen Mehrfachwerte eines Produktes dar. Auch hier können Ähnlichkeiten zwischen Anfrage und Merkmalswerten bestimmt werden, indem die einzelnen Produktwerte mit der Anfrage verglichen werden. Als Ähnlichkeit wird in diesem Fall das Maximum gesetzt (ausgehend von n Einzelwerten im Produktmerkmal):

$$z(w_m(p)) = \max(w_m(p)_1, \dots, w_m(p)_n) \quad (5)$$

Mit den vorgestellten Verfahren lassen sich Zugehörigkeitsfunktionen für alle Skalierungs- und Anfragearten der im Klassifikationssystem vorhandenen Merk-

male definieren. Im nächsten Schritt müssen diese errechneten Einzelgewichte mittels einer Rankingfunktion zu einer gesamten Produktrelevanz zusammengefasst werden.

3.5.2 Rankingfunktion

Um die merkmalsbezogenen Ähnlichkeiten zu einem Gesamtranking zusammenzufassen, verwenden wir das erweiterte boolesche Modell [Sal⁺83].

Definition 7: Erweitertes boolesches Modell

Das *erweiterte boolesche Modell* unterscheidet die Rankingfunktion $\mathbf{r}(\mathbf{q}, \mathbf{p})$ in Abhängigkeit von den innerhalb der Anfrage verwendeten booleschen Operatoren and, or und not. Die Formeln lassen sich rekursiv anwenden. Wir setzen (6)

$$r(q, p) = 1 - \left(\frac{\text{weight}_1(q)^k (1 - z_1(w_1(p)))^k + \dots + \text{weight}_n(q)^k (1 - z_n(w_n(p)))^k}{\text{weight}_1(q)^k + \dots + \text{weight}_n(q)^k} \right)^{1/k}$$

im Fall einer and-Verknüpfung der Merkmals-/Wertepaare in der Anfrage \mathbf{q} . Dabei repräsentiert $\text{weight}_i(\mathbf{q})$ das (nutzerdefinierte) Gewicht eines Merkmals-/Wertepaars der Anfrage und $z_i(w_i(\mathbf{p}))$ das Ergebnis der Zugehörigkeitsfunktion von Merkmalswert der Anfrage und Merkmalswert des Produktes für das Merkmal i . n bezeichnet die Anzahl der Merkmale in der Anfrage \mathbf{q} . (Die Formeln für or und not können der Literatur [Sal⁺83] entnommen werden.) Im erweiterten booleschen Modell können durch Variation des Parameters \mathbf{k} mehrere Rankingverfahren abgebildet werden.

Tabelle 4 zeigt beispielhaft das bewertete Ergebnis für die Anfrage aus Abschnitt 3.1. Dabei wurde generell die Trapezfunktion als Zugehörigkeitsfunktion gewählt, der Glättungsbereich wurde außer bei dem Merkmal Herstellungsjahr, wo mit +/- 5 Jahren gerechnet wurde, mit +/- 50% bestimmt. Alle Produktpreise unterhalb des Anfragewertes von 40.000 EUR wurden mit 1 bewertet. Der Parameter \mathbf{k} in der Rankingfunktion wurde auf 1 gesetzt.

	Drehzahlbereich (U/min)	Drehlänge (mm)	Gesamtantriebsleistung (kW)	Hersteller	Herstellungsjahr	Preis (€)	Ranking
Anfrage	[1000..5000]	800	[20..25]	Okuma	1985	40000	
Gewichtsvektor	1	1	1	1	0,7	1	
Maschine 7	[15..3000]	800	30	Okuma	1985	32000	
Gewichtsvektor	0,2	1	0,6	1	1	1	0,7894737
Maschine 4	[35..7500]	200	22	Index	1985	40000	
Gewichtsvektor	1	0	1	0	1	1	0,6491228
Maschine 5	[15..3000]	800	30	Okuma	1985	66539	
Gewichtsvektor	0,2	1	0,6	1	1	0	0,6140351
Maschine 3	[9..1800]	800	22	Gildemeister	1983		
Gewichtsvektor	0	1	1	0	0,6	0	0,4245614
Maschine 2	[20..5000]	190	11	EMCO	1992	26000	
Gewichtsvektor	1	0	0,1	0	0	1	0,3684211
Maschine 6	[20..4000]	490	30	Traub	1991	47500	
Gewichtsvektor	0,6	0,23	0,6	0	0	0,63	0,3614035
Maschine 1	[3..2500]	1000	15	Boehringer	1992	40900	
Gewichtsvektor	0	0,5	0,5	0	0	0,96	0,3438596

Tabelle 4: Ranking-Ergebnisse mit Fuzzy-Retrieval

In der Tabelle ist zu sehen, dass sich im Vergleich zur rein statistischen Ähnlichkeitsermittlung im Vektorraum-Modell ein differenzierteres Ranking ergibt. Jedes Merkmals-/Wertpaar wird nur dann mit 0 gewichtet, wenn es nach der Zugehörigkeitsfunktion keine Ähnlichkeit gibt, der Produktwert also außerhalb des Glättungsbereiches liegt. Die Bewertung der aus einer Produktgruppe stammenden Maschinen – vom angefragten Typ „Drehmaschine“ her sind alle potentiell geeignet – variiert zwischen 79% und 34%.

3.6 Interpretation der Ergebnisse

Betrachtet man die beiden Suchstrategien, so lässt sich feststellen, dass genauere Suchergebnisse als bei einer reinen booleschen Anfrage oder der Suche entlang der Referenzhierarchie möglich sind. Für die betrachtete Anfrage würde eine rein boolesche Suche für die in Tabelle 3 und 4 gefundenen Produkte beispielsweise keinen Treffer liefern. Ein Vergleich der Tabellen zeigt, dass die Suchergebnisse des Fuzzy-Verfahrens der Gewichts-bildung im vektorialen Verfahren qualitativ überlegen sind. Zu erklären ist dies durch die stärkere Einbeziehung der Anwendungsdomäne bei dem fuzzyfizierenden Suchverfahren. Die besten Treffer allerdings ermitteln beide Ansätze zumindest qualitativ gleich.

Beide Verfahren lassen gewichtete Anfragen zu und ermitteln eine Gesamtbewertung aufgrund von Einzelgewichten der Merkmals-/Wertpaare.

Vorteile des erweiterten Fuzzy-Retrievals sind:

- Anfragen können unscharf verarbeitet werden, Produktwerte müssen nicht genau getroffen werden

- Einzelgewichte der Merkmals-/Wertpaare werden anwendungsspezifisch bestimmt, eine genauere Anpassung der Zugehörigkeitsfunktionen an die Anwendungsdomäne ist möglich
- Beliebige logische Verknüpfungen können in der Anfrage verwendet werden

Die Vorteile des Vektorraum-Modells sind:

- Der Wartungsaufwand für den Marktplatzbetreiber ist geringer, da keine Hinterlegung von Skalen und Zugehörigkeitsfunktionen notwendig ist
- Die Auswertung ist schneller, da Einzelgewichte unabhängig von der Suchanfrage bereits über die ipf -Berechnung vorliegen.

Das Vektorraum-Modell schneidet vor allem bei kleinen Datenbeständen schlecht ab, da in diesem Fall für viele potentielle Anfragewerte keine passenden Produktwerte vorliegen. Eine Verbesserung kann hier durch Relevance Feedback erreicht werden, in dem für nicht als Indexterm vorhandene Anfragewerte Vorschläge aus dem Index gemacht werden. Allerdings stellt dies faktisch eine Fuzzyifizierung dar.

Generell wird das erweiterte Fuzzy-Retrieval die bessere Wahl sein. Weitergehende Untersuchungen zur Verbesserung dieses Ansatzes sollten zum einen Variationen des k -Parameters umfassen. Erste Tests zeigen hier höhere Genauigkeiten bei kleineren Werten von k . Zum anderen sollten die Variationsmöglichkeiten bei der Bildung der Zugehörigkeitsfunktion für (mehrdimensional) bereichsfähige Merkmale untersucht werden, die derzeit durch Multiplikation erfolgt. Tests haben gezeigt, dass auch Flächenvergleiche durch Integralbildung zu relevanten Ergebnissen führen. Auch Variationen von Bestimmung und Größe der Glättungsbereiche ergeben unterschiedliche Ähnlichkeiten und müssen anwendungsspezifisch evaluiert werden.

3.7 Automatische Klassifizierung

Zur automatischen Klassifizierung von Produkten in ein gegebenes Klassifikationssystem können die vorgestellten Verfahren herangezogen werden, indem das zu klassifizierende Produkt als Suchanfrage aufgefasst wird:

1. Die Merkmale des zu klassifizierenden Produkts werden über das vektorielle Retrieval (wobei nur Merkmale als Indexterme genutzt werden) mit den bestehenden Sachmerkmalsleisten der Produktgruppen verglichen und ergeben Ähnlichkeiten zwischen Produkt und Produktgruppen.
2. Die Produktgruppen, die zu den durch erweitertes Fuzzy-Retrieval ermittelten besten Suchergebnissen gehören, werden als Vorschläge zur Klassifizierung angegeben. Hinzu kommen die gemäß Schritt 1 gefundenen Produktgruppen,

denn ein Produkt kann auch in eine Produktgruppe klassifiziert werden, wenn sie keine oder keine ähnlichen Produkte enthält.

4 Zusammenfassung und Ausblick

In dieser Arbeit haben wir untersucht, wie die Suche in Produktkatalogen als auch die automatisierte Klassifizierung von Produkten verbessert werden können. Dazu wurden Verfahren aus dem Information Retrieval für das Produkt-Retrieval erweitert, angepasst und prototypisch implementiert. Aus einer exakten Anfrage wurde mit Hilfe von Fuzzy-Verfahren eine unscharfe Produktbeschreibung für die Suche generiert. Hierzu haben wir die verschiedenen Skalierungen der Merkmale zur Beschreibung von Produkten klassifiziert, und für die einzelnen Klassen wurden geeignete Zugehörigkeitsfunktionen beschrieben. Die Zugehörigkeitsfunktionen legen die Relevanz einzelner Produktmerkmale bezüglich der Anfrage fest. Um zu einer Gesamtbewertung zu kommen, wurden das erweiterte boolesche Modell und das Vektorraum-Modell adaptiert.

Empirische Auswertungen haben gezeigt, dass die vorgestellten Ansätze erhebliche Vorteile bringen. Diese Ergebnisse sollen nun in weiteren Experimenten vertieft analysiert werden. Ein weiteres Thema wird die Zuordnung und Anpassung der Zugehörigkeitsfunktionen sein, die derzeit noch manuell erfolgen muss, was ein tiefgehendes Wissen in der Anwendungsdomäne erfordert. Insbesondere soll untersucht werden, wie eine Bestimmung und Anpassung von Zugehörigkeitsfunktionen für einzelne Merkmale über Lernmethoden in Verbindung mit Relevance Feedback automatisiert erfolgen kann.

Literatur

- [Ble⁺96] Bley Müller, Josef; Gehlert, Günther; Gülicher, Herbert (1996): Statistik für Wirtschaftswissenschaftler. Verlag Franz Vahlen, München, 10. Auflage.
- [BME02] Bundesverband Materialwirtschaft, Einkauf und Logistik (2002): BMEcat 1.2, <http://www.bmecat.org/>, Abruf am 12. Februar 2003
- [DIN98] Deutsches Institut für Normung e.V. (DIN) (1998): Grundlagen für den Aufbau eines Merkmal-Lexikons. Beuth Verlag, Berlin
- [DIN02] Deutsches Institut für Normung e.V. (DIN) (2002): Optimierung elektronischer Geschäftsprozesse - Neuer Arbeitskreis 'Produktklassifikation und -merkmale' entwickelt umfassendes Konzept, Mitteilung vom 13. Dezember 2002 auf <http://www.din.de>, Abruf am 23. Januar 2003

- [FDM03] Fachverband des Deutschen Maschinen- und Werkzeug-Großhandels e.V. (FDM) (2003): <http://www.machinestock.com/>, Abruf am 10. Februar 2003
- [Fuhr00] Fuhr, Norbert (2000): *Models in Information Retrieval*, Veröffentlichung der Universität Dortmund
- [Grau95] Grauel, Adolf (1995): *Fuzzy-Logic. Einführung in die Grundlagen mit Anwendungen*. BI Wissenschaftsverlag, Mannheim
- [IDW03] Institut der deutschen Wirtschaft Köln (2002): eCI@ss 4.0, <http://www.eclass.de/>, Abruf am 12. Februar 2003
- [Kri+02] Krieger, Rolf; Kuhn, Norbert; Mees, Michael; Naumann, Stefan; Schürmann, Cordula; Sommer, Christian (2002): *Verwendbarkeit von Klassifikationssystemen und Katalogstandards zum Aufbau von elektronischen Handelsbörsen für gebrauchte Maschinen*. In: Weinhardt, Christof; Holtmann, Carsten (Hrsg.): *E-Commerce: Netze, Märkte, Technologien*. Physica-Verlag, Heidelberg
- [Leu+02] Leukel, Jörg; Schmitz, Volker; Dorloff, Frank-Dieter (2002): *Referenzmodell für den Austausch von Produktklassifikationssystemen im E-Business*. In: Loos, Peter; Gronau, Norbert (Hrsg.): *E-Business - Integration industrieller ERP-Architekturen*. Cuvillier Verlag, Göttingen
- [Oga+91] Ogawa, Yasushi; Morita, Tetsuya; Kobayashi, Kiyohiko (1991): *A fuzzy document retrieval system using the keyword connection matrix and a learning method*. In: *Fuzzy Sets and Systems* 39:163-179.
- [OTBe01] Otto, Boris; Beckmann, Helmut (2001): *Klassifizierung und Austausch von Produktdaten auf elektronischen Marktplätzen*. In: *Wirtschaftsinformatik* 43 Nr. 4, S.351 – 361
- [RiBa99] Ribeiro-Neto, Berthier; Baeza-Yates, Ricardo (1999): *Modern Information Retrieval*. Addison-Wesley, Harlow
- [Rijs79] van Rijsbergen, Cornelis J. (1979): *Information Retrieval*, Butterworths
- [Salt71] Salton, Gerald (ed.) (1971): *The SMART Retrieval System - Experiments in Automatic Document Processing*. Prentice Hall, Englewood, New Jersey
- [Sal+83] Salton, Gerard; Fox, Edward A.; Wu, Harry (1983): *Extended Boolean Information Retrieval*. In: *Communications of the ACM* 26(11), 1022-1036
- [SaMc83] Salton, Gerald; McGill, Michael J. (1983): *Introduction to Modern Information Retrieval*. McGraw Hill Publications
- [Surp03] Surplex AG (2003): <http://www.surplex.com/>, Abruf am 9. Februar 2003
- [Zade65] Zadeh, Lotfi (1965): *Fuzzy sets*. *Information Control* 8, 338-353