E-Procurement und Catalog Content Management

Seminar "CRM und SRM: Customer und Supplier Relationship Management"

Lehrstuhl für Informationssysteme in Zusammenarbeit mit SAP Retail Solutions

Michael Biwer
mbiwer@studcs.uni-sb.de

Jörg Diesinger
j.diesinger@vizzavi.de

Patrick Zimmer
pzimmer@studcs.uni-sb.de

Universität des Saarlandes

WS 2002/2003, 11. Februar 2003

Inhaltsverzeichnis

1 Einleitung **3**

2 E-Procurement im Allgemeinen **4**

2.1 Was versteht man unter E-Procurement? 4
2.2 Direkte und indirekte Materialien 4
2.3 Die Beschaffungsprozesskosten 5
2.4 Der Beschaffungsprozess 6
2.4.1 Der traditionelle Einkauf 6
2.4.2 Der E-Procurement-Zyklus 6
2.5 Direkte und indirekte Materialien – Auswirkungen auf den Beschaffungsprozess 7
2.6 Einkaufsverfahren – Ebenen der elektronischen Beschaffung 9
2.7 Erfolgsfaktoren von E-Procurement-Projekten 10
2.8 Komponenten eines E-Procurement-Systems 11

3 Catalog Content Management **12**

3.1 Allgemeines zu Katalogen 12
3.2 Katalogarten 13
3.2.1 Single-Supplier Katalog 13
3.2.2 Multi-Supplier Katalog 14
3.2.3 Externer Katalog 14
3.2.4 Interner Katalog 15
3.2.5 Dienstleister-Prinzip 15
3.3 Katalog-Integration 16
3.3.1 BMEcat 17
3.3.2 eCl@ss 21
3.3.3 Automatisierte Kategorieintegration 22
3.4 e-Catalogs – Kriterien und Anforderungen 24
3.4.1 Allgemeine Kriterien für die Auswahl eines Kataloges 24
3.4.2 Anforderungen an die Katalogsoftware 25
3.4.3 Anforderungen an die Content-Management Software 25
3.4.4 Anforderungen an ein „gutes“ Katalog-Suchsystem 25

4 Preference SQL **25**

4.1 Einführung 25
4.2 Preference SQL Sprach-Design 26
4.2.1 Integrierte Präferenz Basis-Typen 27
4.2.2 Zusammenbau von komplexen Präferenzen 28
Gleiche Gewichtung: „Pareto Akkumulation“ 28
Geordnete Gewichtung: Kaskadierung von Präferenzen 28

Kombination von Pareto Akkumulation und Kaskadierung 28
4.2.3 Qualitätsfunktionen 29
4.2.4 Qualitätskontrolle 30
4.3 Implementation von PSQL 30
4.3.1 Plug-and-Go Application Integration 30
4.3.2 Preference SQL Optimizer 31
4.4 Anwendungsbeispiel 31

5 Fazit **34**

Abbildungsverzeichnis **35**

Literatur **36**

1 Einleitung

Die zunehmende Globalisierung, die wachsende Dynamik der Beschaffungsmärkte sowie ein weiterhin steigender Kostendruck stellen erhöhte Anforderungen an Einkauf und Beschaffung in Unternehmen. Das Internet bietet dazu die elektronische Grundlage, auf der konventionelle Geschäftsmöglichkeiten und -prozesse jedoch völlig neu überdacht werden müssen. In immer stärkerem Maße erkennen Unternehmen, daß der Einsatz des Internets in der Beschaffung enorme Verbesserungspotentiale freisetzt – sei es beim Einkauf von Büroklammern, dem Auftrag an die Werbeagentur oder die nächste Rohöllieferung. Je effizienter Waren und Leistungen beschafft werden, desto größer sind die Einsparungen bei den Materialien und Dienstleistungen selbst und im Beschaffungsprozess.

Die vorliegende Arbeit, die in drei Kapitel gegliedert ist, geht nach einigen Begriffserläuterungen im ersten Teil auf die unterschiedlichen Ausprägungen der elektronischen Beschaffung ein und versucht weiterhin, die möglichen Einsparpotentiale zu erkennen und offenzulegen.

Das zweite Kapitel beschäftigt sich mit einer zentralen Komponente von E-Procurement-Systemen, nämlich dem elektronischen Katalog. Dabei wird insbesondere auf die verschiedenen Arten solcher Kataloge sowie auf das Problem der Integration fremder Produktdaten ins eigene System eingegangen.

Im dritten Kapitel wird schließlich anhand von PSQL eine interessante Erweiterung der integrierten Suchfunktionalität elektronischer Kataloge – einer der wichtigsten Qualitätsfaktoren dieser Systeme – vorgestellt.

2 E-Procurement im Allgemeinen

2.1 Was versteht man unter E-Procurement?

Der Begriff des E-Procurement bezeichnet "die Nutzung von Informations- und Kommunikationstechnologien zur elektronischen Unterstützung von Beschaffungsprozessen und deren Integration in den Arbeitsablauf eines Unternehmens.
Beschaffungsprozesse dienen dazu, einem Unternehmen alle diejenigen Produktionsfaktoren zur Verfügung zu stellen, die zur Erfüllung eines bestimmten Sachzieles benötigt werden, aber nicht selbst produziert werden können."[1]
Dabei kann es sich sowohl um Produkte aus verschiedenen Warengruppen handeln als auch um Dienstleistungen.

Es wird dabei von möglichen Kosteneinsparungen für das Unternehmen von insgesamt etwa 15% ausgegangen, die sich wie folgt zusammensetzen: die elektronische Beschaffung senkt die Transaktionskosten für den Bestellvorgang um bis zu 70%. Die Einkaufspreise für die Produkte können durch eine Nachfragebündelung um 5 bis 15% gesenkt werden. Zahlungsvorgänge reduzieren sich um bis zu 25%, die Beschaffungsgemeinkosten um bis zu 20% und die Lagerhaltungskosten lassen sich um bis zu 10% verringern. [PWC02]

Diesen Schätzungen für einen optimalen elektronischen Einkauf liegen allerdings die größtmöglichen Aufwendungen beim traditionellen Beschaffungsprozess zugrunde, die beispielsweise die vorherige Angebotsanfrage und Informationsbeschaffung bei verschiedenen Anbietern sowie die sorgfältige Auswahl des optimalen Lieferanten beinhalten.

Welches enorme Einsparpotential der konventionelle Einkaufsprozess birgt, verdeutlicht auch die Aussage der US National Association of Purchasing Professionals:

> *"Eine Reduzierung der Beschaffungskosten um 5% hat den gleichen Effekt auf die Gewinn- und Verlustrechnung eines Unternehmens wie ein 30%iger Umsatzzuwachs."*

Da eine Umsatzsteigerung von 30% nicht nur in Zeiten schwacher Konjunktur für ein Unternehmen schwer zu realisieren ist, erscheint die Reduzierung der Beschaffungskosten um 5% als die einfachere Möglichkeit zum Erreichen desselben Zieles.

2.2 Direkte und indirekte Materialien

Im Zusammenhang mit der elektronischen Beschaffung ist es wichtig, zwiscchen direkten und indirekten Materialien zu differenzieren, da sich hier große Unterschiede im Beschaffungsprozess ergeben, auf die in den folgenden Abschnitten noch weiter eingegangen wird.

Direkte Materialien definiert man als diejenigen Materialien, die direkt in die Produktion einer Ware eingehen, d.h. Bestandteil des Endproduktes eines Fertigungsprozesses sind. Dazu zählt man Rohstoffe wie beispielsweise Holz oder Metall, Hilfsstoffe wie Leim oder Schrauben und Bauteile wie zum Beispiel Metallfassungen.

Indirekte Materialien werden auch als MRO-Güter (Maintenance, Repair and Operations) bezeichnet und definieren die Waren, die zur Wartung und Reparatur benötigt werden sowie zum laufenden Betrieb eines Unternehmen, zum Beispiel Büromaterial, Maschinenteile oder Dienstleistungen. Indirekte Materialien sind kein Bestandteil des Endproduktes.

[1] Quelle: KPMG Consulting, 2003

2.3 Die Beschaffungsprozesskosten

Die entscheidenden Stellgrößen für die Gesamtkosten in der Beschaffung sind einerseits der Einkaufspreis der benötigten Ware und andererseits die Kosten, die aus dem Beschaffungsvorgang heraus entstehen.

Die Einkaufspreise werden dabei im Wesentlichen bestimmt durch die Marktposition, die der jeweilige Anbieter bzw. Nachfrager vertritt, die Eigenschaften bzw. die Substituierbarkeit des Artikels sowie das strategische Verhalten des Einkäufers.

Die Prozesskosten in der Beschaffung ergeben sich größtenteils aus administrativen Vorgängen der Angebotseinholung und Bestellabwicklung, der Terminverfolgung, aber auch der Kontrolle am Wareneingang (Qualitätskontrolle) sowie schließlich der Rechnungsprüfung und Zahlung der Ware.

Die folgende Abbildung aus [ACT01] illustriert die Diskrepanz, die sich diesbezüglich bei einem Vergleich von direkten und indirekten Gütern ergibt. Ein Balken stellt hier die jeweiligen Beschaffungsgesamtkosten dar.

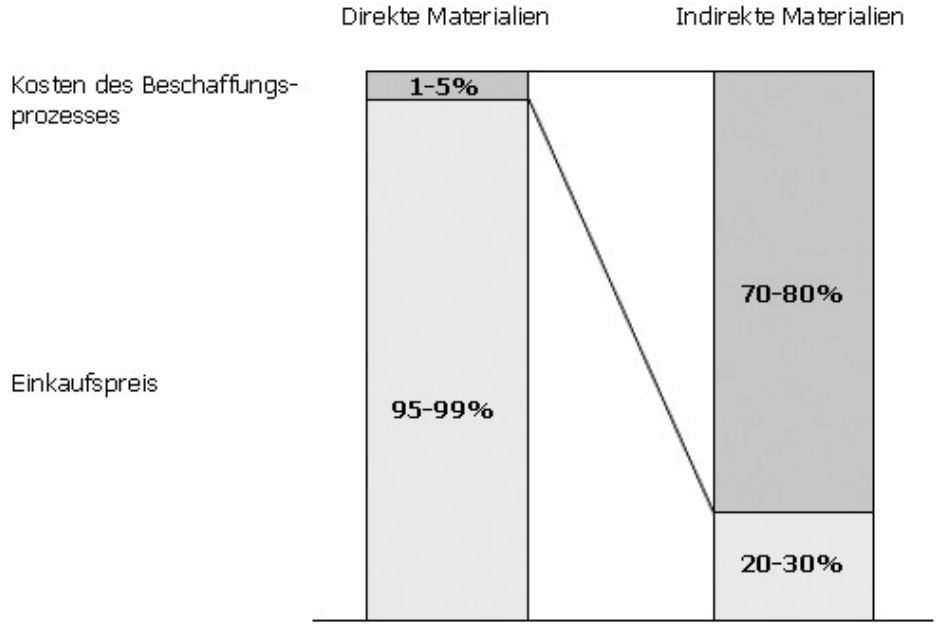

Abbildung 1: Kostenstrukturen für direkte und indirekte Güter

Es ist eine Tatsache, daß trotz der sehr viel höheren Prozesskosten bei der Beschaffung von indirekten Gütern in Relation zu deren Einkaufspreisen Beschaffungsentscheidungen nach wie vor fast ausschließlich durch den Einkaufspreis bestimmt werden. Dieses enorme Einsparpotential auf Einkaufsebene, sowie die Unterschiede von indirekten zu direkten Materialien, die sich aufgrund ihrer Definition ergeben, sind die maßgeblichen Faktoren, die die besondere Eignung von indirekten Gütern für die elektronische Beschaffung ausmachen. In Abschnitt 2.5 werden diese explizit einander gegenübergestellt.

Im Folgenden werden der traditionelle Beschaffungsprozess und der Beschaffungszyklus beim Einkauf unter Verwendung eines E-Procurement-Systems dargestellt.

2.4 Der Beschaffungsprozess

2.4.1 Der traditionelle Einkauf

Beim traditionellen und nicht oder lediglich teilweise automatisierten Beschaffungsprozess werden Einzelprozesse in Papierform abgewickelt. Informationen müssen zwischen vielen am Prozess beteiligten Personen per Telefon, Fax, E-Mail oder persönlich übermittelt werden. Zwangsläufig entstehen langwierige Geschäftsprozesse, die eine redundanz- und fehlerfreie Datenhaltung nahezu unmöglich machen.

Die Folge ist ein Geschäftsprozess, der durch lange Zykluszeiten und hohe Bearbeitungs- und Transaktionskosten geprägt ist. Eine Kontrolle der Lieferanten und Logistikdienstleister ist nicht oder nur teilweise möglich (vgl. Abbildung 2).
Die manuelle Abwicklung einer Bestellung nimmt durchschnittlich 7,3 Tage in Anspruch.[2]

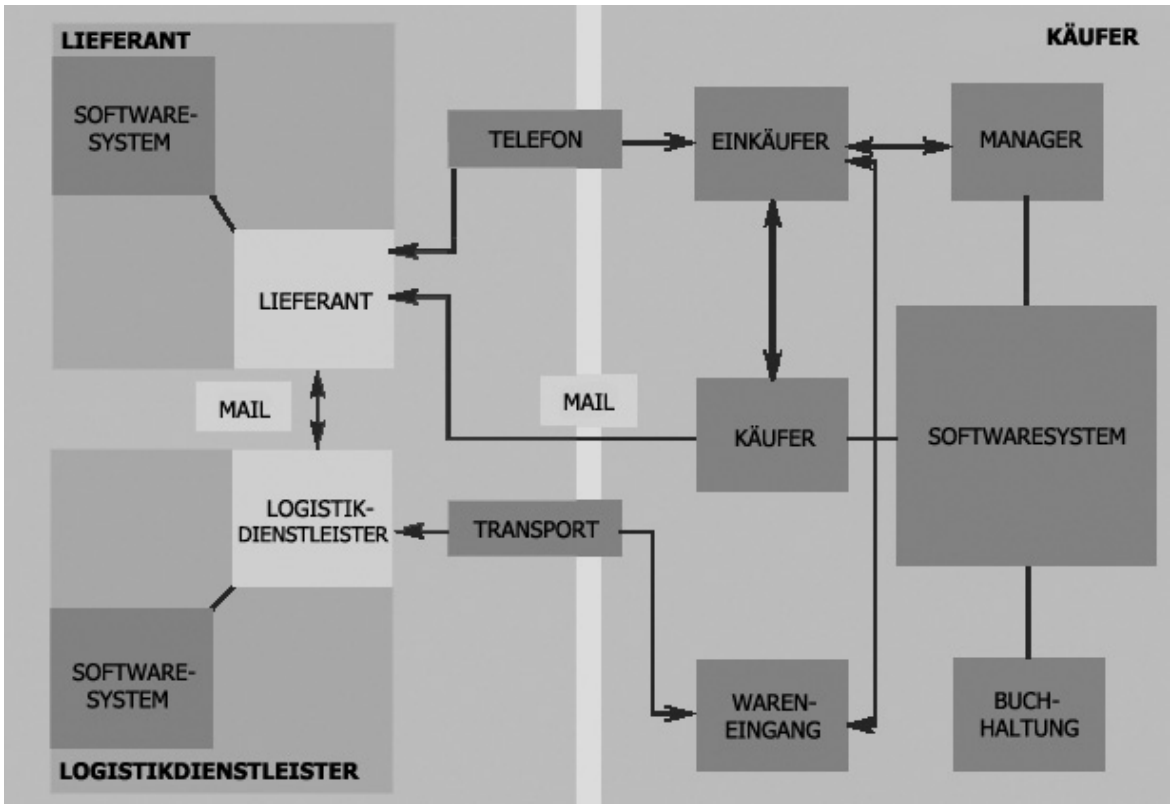

Abbildung 2: Der traditionelle Beschaffungsprozess

2.4.2 Der E-Procurement-Zyklus

Der Einkauf erfolgt im E-Procurement-System nach dem sogenannten “Self-Service”-Verfahren, d.h. die Warenbestellung geht direkt von dem Mitarbeiter aus, der den Bedarf feststellt. Die Einkaufsabteilung ist an diesem Prozess weitestgehend unbeteiligt.

Der Vorgang durchläuft im Wesentlichen drei Stationen:

- **Bezugsquellenfindung**: Der Mitarbeiter wählt in einem elektronischen Katalog die gewünsch-

[2] Quelle: Aberdeen Group

ten Artikel aus. Über ein Warenkorb-System wird seine Anforderung verwaltet und gespeichert. Er hat außerdem die Möglichkeit, eine aktuelle Statusabfrage (Verfügbarkeitsabfrage) für seine Artikel beim Lieferanten durchzuführen.

- **Einkauf**: Nachdem die Bedarfsanforderung optional einen Genehmigungsworkflow durchlaufen hat (was beispielsweise bei hochpreisigen Produkten unumgänglich sein kann), wird sie automatisch in eine Bestellung umgewandelt und an den Lieferanten übertragen. Hier erfolgt die eigentliche Abwicklung der Bestellung.
- **Abwicklung**: Die Auslieferung der Ware wird dem Mitarbeiter als Lieferbestätigung mitgeteilt (Kontrolle der Logistikkette). Entsprechend erfolgt die Rechnungsstellung. Die Qualitätskontrolle und Rechnungsprüfung wird unter Umständen ebenfalls vom Käufer selbst durchgeführt. Der Beschaffungszyklus ist beendet (vgl. Abbildung 3).

Der Käufer hat durch das System jederzeit Kontrollmöglichkeiten und vollen Zugriff auf den Prozess. Die Reduzierung der Beschaffungszeit und Transaktionskosten ist offensichtlich.
Die Abwicklung einer Bestellung, gestützt durch ein E-Procurement-System, benötigt nur noch zwei Tage.[3]

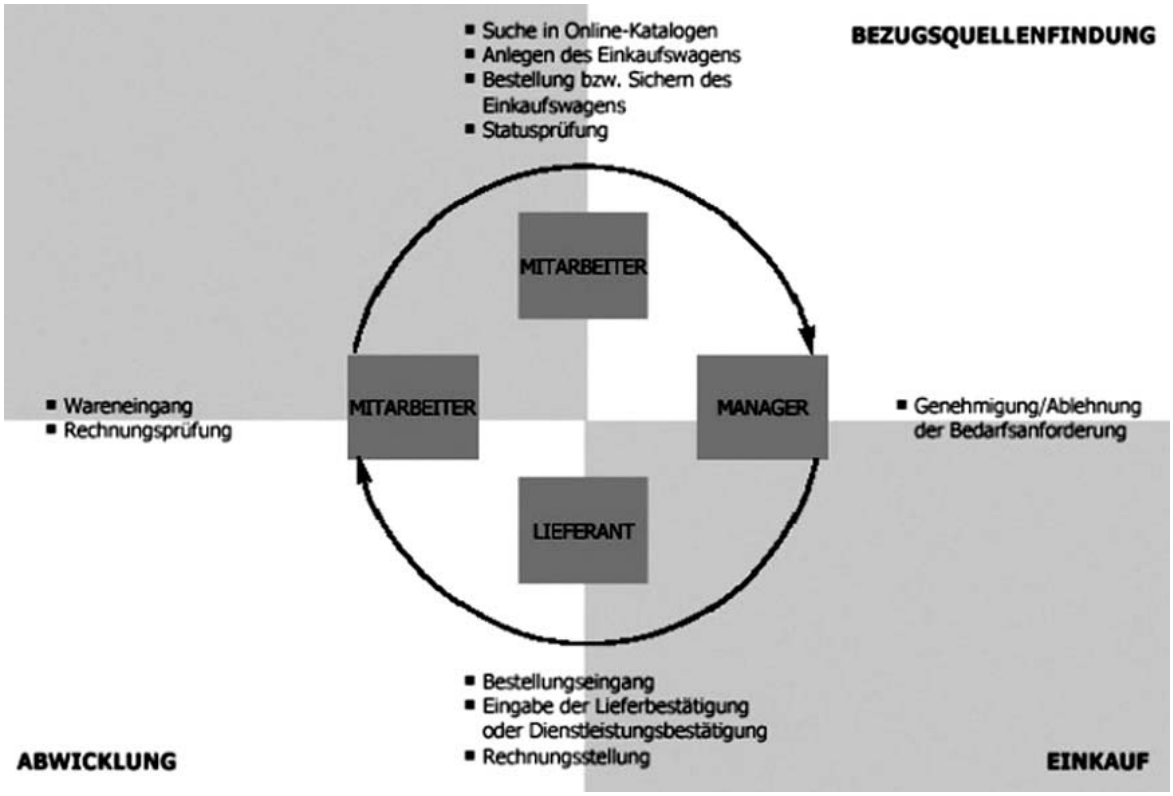

Abbildung 3: Der Beschaffungszyklus durch E-Procurement

2.5 Direkte und indirekte Materialien – Auswirkungen auf den Beschaffungsprozess

Bedarfsanforderungen für indirekte Materialien können – wie der letzte Abschnitt dargestellt hat – sehr einfach und effektiv über eine Beschaffungssoftware und interne Produktkataloge in einem

[3] Aberdeen Group

Unternehmen automatisiert werden. Verschiedene Faktoren sind hier zu nennen, die den Einsatz des "Self-Service"-Verfahrens für indirekte Güter begründen:

- Der Wert von indirekten Materialien macht nur etwa 20% des Gesamtwertes der Einkäufe eines Unternehmens aus, aber etwa 80% des Volumens ("High volume, low value").
- Indirekte Materialien fließen nicht in die Produktions- bzw. Kerngeschäftsabläufe eines Unternehmens mit ein ("Low risk").
- Es besteht kein Beratungsbedarf.

Forrester Research führt in diesem Zusammenhang weiterhin an:

- Die Artikel sind für administrative Zwecke bestimmt und der Bedarf ist vorhersehbar (betrachtet man beispielsweise Verbrauchsmaterialien im Bürobereich).
- Indirekte Waren sind Standardartikel (Produkte "von der Stange"), deren Preis (relativ) niedrig ist und kaum schwankt.
- Aufgrund der geringen Komplexität der indirekten Materialien lassen sich Angebote sehr einfach untereinander vergleichen.
- Der Endverbraucher ist der interne Mitarbeiter.

Zusammenfassend zeichnet sich laut Morgan Stanley, Dean Witter Internet Research eine E-Procurement-Lösung zur Beschaffung indirekter Materialien durch folgende Merkmale aus:

- Alle Mitarbeiter eines Unternehmens sind berechtigt, das System zu nutzen.
- Auslöser für eine Bestellung ist der Bedarf eines Einzelnen (Operativer Einkauf, vgl. Abschnitt 2.6).
- Die Integration in die IT-Landschaft des Unternehmens ist hier nur begrenzt erforderlich. Es werden lediglich Workflow-Funktionen (zum Beispiel für den Genehmigungsprozess) und Anbindungen an Finanz- und Controllingsysteme sowie die Katalogkomponente benötigt.
- Der Einkauf erfolgt über angeschlossene Online-Kataloge.
- Sonderkonditionen oder Rabatte müssen mit den Lieferanten im Vorfeld ausgehandelt werden.

Im Gegensatz dazu stellt sich die Beschaffung der Direktmaterialien sehr viel komplexer dar.
Sie machen den größten Anteil an den Gesamtausgaben eines Unternehmens aus (Forrester Research spricht von 80%) und können dem Endprodukt direkt zugerechnet werden. Direktmaterialien wirken sich somit unmittelbar auf den Umsatz, den Marktanteil und die Produktqualität aus.

- Ihr Bedarf ist nicht vorhersehbar. Die Nachfrage nach dem Endprodukt am Markt steuert letztlich die Beschaffungsentscheidung für die Direktmaterialien.
- Ebenso wird der Preis für Direktmaterialien durch Angebot und Nachfrage bzw. die Verfügbarkeit am Markt beeinflußt.
- Der Endverbraucher definiert sich als externer Kunde (Käufer des Endproduktes).

Beim Kauf von Direktmaterialien kommt es somit vielmehr auf die Produktqualität, die Zuverlässigkeit der Lieferanten und die Lieferkonditionen an. Ein entsprechendes E-Procurement-System muß also über den bloßen Bestellprozess hinaus Funktionalitäten zur Kontrolle und Steuerung von Lieferantenanfragen, Angeboten, Verträgen und Nachbestellungen bieten.[4]
Morgan Stanley, Dean Witter Internet Research stellen der Beschaffung indirekter Güter folgende Auswirkungen der Direktbeschaffung auf die E-Procurement-Lösung gegenüber:

- Die Benutzer des Systems sind ausschließlich Facheinkäufer, die Lieferantenbeziehungen pflegen, Verträge und Konditionen aushandeln und auf detailliertes Wissen über die Produktionsvorgänge und (Roh-) Materialien zurückgreifen können.
- Bestellungen werden sorgfältig geplant und über feste Lieferpläne terminiert (Strategischer Einkauf, vgl. Abschnitt 2.6).
- Es muß eine vollständige Integration des E-Procurement-Systems in die bestehende IT-Landschaft des Unternehmens erfolgen. So muß beispielsweise ein Zugriff auf die Produktions- und Logistikplanung ebenso möglich sein wie eine Erstellung detaillierter Auswertungen über die Analyse-Werkzeuge eines Business-Information-Warehouse.
  Eine direkte Anbindung der Lieferanten an das Backend-System erlaubt zudem die volle Bestandskontrolle seitens des Lieferanten, wodurch Vereinbarungen über Nachbestellungen beispielsweise ohne zusätzlichen Bestellaufwand eingehalten werden können (Supply Chain Management).

2.6 Einkaufsverfahren – Ebenen der elektronischen Beschaffung

Die verschiedenen Anforderungen an direkte und indirekte Materialien und die damit verbundenen spezifischen Zielsetzungen für den Einkauf eines Unternehmens führen zu unterschiedlichen Ausprägungen des jeweiligen Beschaffungsprozesses. Man unterscheidet drei Level:

- **Level 1 - Der operative Einkauf**

  Dieser folgt im Wesentlichen der Prozessreihenfolge

  - Auswahl des Produktes/der Dienstleistung,
  - Auftragsabwicklung,
  - Rechnungsbearbeitung

  und beschreibt somit die katalogbasierte Beschaffung von standardisierten Waren oder Dienstleistungen, also den Einkauf indirekter Güter. Der Bestellvorgang wird sofort zu dem Zeitpunkt abgesetzt, wenn der Bedarf auftritt und erfordert keine längerfristige Planung. Die elektronische Unterstützung dieser gleichartigen und wiederkehrenden Beschaffungsvorgänge führt zwangsläufig zu einer Erhöhung der Rahmenvertragsquote und Verbesserung der Beschaffungskonditionen beim Lieferanten und verringert so das sogenannte Maverick-Buying (bezeichnet die Beschaffung von Waren und Dienstleistungen ohne Nutzung von Rahmenverträgen, obwohl diese mit Lieferanten vereinbart wurden).

  Die Einsparpotentiale dieses Verfahrens fassen sich zusammen in

  - (Prozess-) Kostenreduzierung insgesamt durch
    * Verringerung der mauellen administrativen Tätigkeiten,
    * Beschleunigung (Verkürzung) des Beschaffungsprozesses

    → Einsparungen von bis zu 20% der jährlichen Gesamtkosten eines Unternehmenes sind denkbar.[5]

[4] Forrester Research
[5] Quelle: PricewaterhouseCoopers Unternehmensberatung GmbH

  - Reduktion der Übertragungsfehler,
  - Dezentrale Bestellung am Arbeitsplatz und/oder bei Geschäftspartnern,
  - Lagerbestandsreduktion,
  - Senkung der Einkaufspreise durch Markttransparenz,
  - Hohe Aktualität und Transparenz durch Anbindung an Online-Kataloge

- **Level 2 - Der strategische Einkauf**

  Diese Ebene des Einkaufs hat letztlich die Bestimmung des besten Preis-Leistungs-Verhältnisses am Markt zum Ziel. Durch

  - Planung der Beschaffung und Prognosenerstellung,
  - Marktanalyse,
  - Lieferantenauswahl und
  - Vertragsabschlüsse

  möchte der Einkäufer die bestmöglichen Konditionen für eine Einzeltransaktion (beispielsweise in Form von Spot-Käufen) für nicht-standardisierte Waren, also Direktmaterialien erreichen. Hier liegen die Einsparpotentiale begründet in

  - der Verbesserung der Kontroll- und Auswertungsmöglichkeiten,
  - den gegebenen erweiterten Beschaffungsmöglichkeiten wie elektronische Marktplätze, Auktionen, Generierung öffentlicher und nicht-öffentlicher Ausschreibungen,
  - den besseren Konditionen durch (Single-) Sourcing bzw. Konsolidierung von Lieferanten und Vereinbarungen (Defragmentierung von Lieferanten)
  - der Verlagerung von operativen Tätigkeiten auf den Bedarfsträger ("Self-Service", siehe Level 1), wodurch eine Entlastung der Einkaufsabteilung und eine Fokusierung auf die strategischen Ziele (Lieferantenauswahl, Rahmenverträge) erfolgt.

- **Level 3 - Die langfristige Entwicklung**

  Diese dritte Ebene der Einkaufsverfahren berücksichtigt die langfristige Entwicklung sowie den künftigen Umfang der Produktentwicklung eines Unternehmens. Partner und Lieferanten sollen im optimalen Fall bei der Entwicklung neuer Produkte mit dem Unternehmen kollaborieren, d.h. in die Forschung, Planung und das Produktdesign direkt miteinbezogen werden.

2.7 Erfolgsfaktoren von E-Procurement-Projekten

Zusammenfassend lassen sich die wichtigsten Faktoren beschreiben, deren Beachtung für einen erfolgreichen Einsatz von E-Procurement-Systemen unerlässlich ist.

- **Lieferant**: Es muß sichergestellt sein, daß der Lieferant e-fähig ist, d.h. daß er Online-Kataloge zur Verfügung stellt und diese auch in das System integrierbar sind bzw. eventuell eine Anbindung seines Backend-Systems an das unternehmensspezifische E-Procurement-System erfolgen kann.

- **Prozess**: Sämtliche Geschäfts- und Beschaffungsprozesse müssen über alle Bereiche und Abteilungen des Unternehmens hinweg einheitlich und durchgängig elektronisch sein.

- **Software**: Das Zusammenspiel der einzelnen E-Procurement-Komponenten Backend-System, E-Procurement-System und Katalogkomponente (vgl. Abschnitt 2.8) muß gegeben sein. Datenredundanz an Front- und Backend-System ist zu vermeiden.

- **Change Management**: Unter Change Management versteht man die gleichzeitig mit der Einführung der E-Procurement-Lösung stattfindende Anpassung der internen Organisation des Unternehmens. Das **interne Change Management** umfasst hierbei die konsequente Durchsetzung der Prozessanpassungen, sowohl innerhalb der Einkaufsabteilung als auch beim Anwender. Eventuell müssen unternehmensweite Schulungen organisiert werden. Dagegen spricht das **externe Change Management** den Lieferanten an, der möglichst schon zu Beginn in den Prozess eingebunden wird, damit Verträge und Konditionen angepasst, die logistischen Abläufe vereinbart und nicht zuletzt auch seine E-Tauglichkeit geprüft werden kann.
- **Organisationseinheiten**: Die Wirtschaftlichkeit der Einführung einer E-Procurement-Lösung im Unternehmen muß gegeben sein, d.h. es ist eine kritische Masse (eine Mindestanzahl) an Mitarbeitern erforderlich, die das System nutzt. Nur so kann die kostenintensive E-Procurement-Umsetzung getragen werden.
- **Einkauf**: Im Zuge des Change Managements entstehen für den Einkauf neue Aufgabenbereiche: Die Katalogstrukturen im elektronischen Katalogsystem müssen definiert und angelegt, Lieferantendaten müssen integriert und Kataloginhalte angepasst und gepflegt werden (Content Management). Darüber hinaus sind die üblichen strategischen Aufgaben wie Preis- und Konditionsverhandlungen notwendig (Rahmenvertragsmanagement).

2.8 Komponenten eines E-Procurement-Systems

Abbildung 4 zeigt die technische Realisierung einer E-Procurement-Lösung am Beispiel des SAP Enterprise Buyer Professional (EBP). Hervorzuheben sind dabei:

- Das E-Procurement-System (EPS), zum Beispiel SAP EBP.
- Die Web-Komponenten, wie Internet Transaction Server (ITS) und Web-Server.
- Das Enterprise Resource Planning (ERP) Backend-System, zum Beispiel SAP R/3.
- Die Kataloganbindung.
- Optional die direkte Anbindung des Lieferanten-Backend-Systems an das eigene EPS, z.B. über den SAP Business Connector (dies macht vor allem bei Supply Chain Management Sinn)

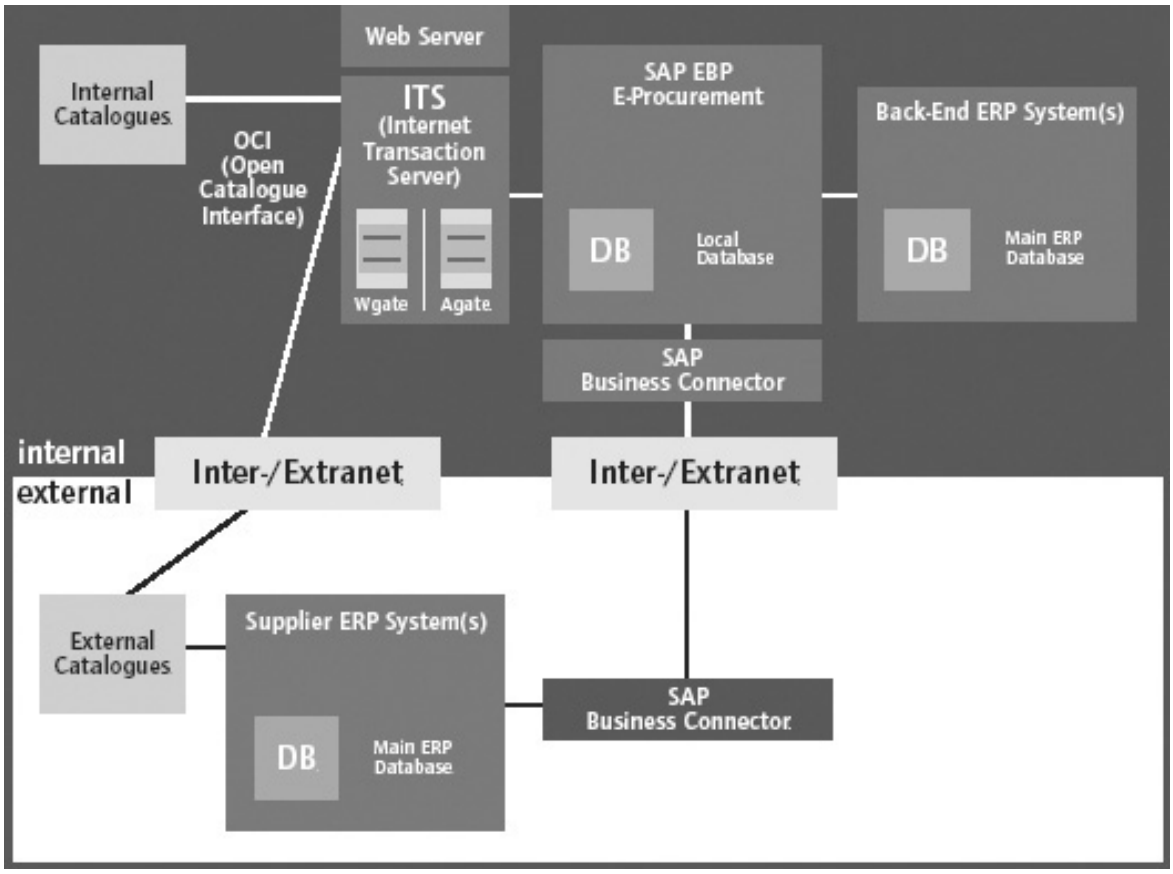

Abbildung 4: Die E-Procurement-System Architektur am Beispiel von SAP EBP

Das folgende Kapitel behandelt die Katalog-Komponente als wichtigen Bestandteil und Erfolgsfaktor eines E-Procurement-Systems. Dabei wird auf die Frage nach Format-Standards und Klassifizierungsmethoden bei der Fremdkatalogintegration ebenso eingegangen wie auf allgemeine Anforderungen an Katalogsysteme.

3 Catalog Content Management

3.1 Allgemeines zu Katalogen

Der Katalog ist der zentrale Ort, an dem kundenrelevante Informationen zu den verfügbaren Produkten und Dienstleistungen eines Unternehmens vorgehalten werden. Dazu gehören z.B.

- eine strukturierte Beschreibung (Kurz- und Langtext), soweit möglich inkl. Bebilderung
- die Einordnung in eine Produktkatgorie zur Strukturierung des Angebots
- Preisinformationen wie Einzelpreis und ggf. Staffelpreise
- erweiterte Referenzen auf ein Produktvideo, ein PDF-Datenblatt oder einen Link auf die Webseite des Herstellers
- Bestellintervalle, z.B. im Rahmen von Supply Chain Management (SCM); die Idee hierbei ist, daß ein externer Lieferant über eine geeignete Schnittstelle (i.a. ein Webinterface) auf den

Katalog zugreift und anhand der hinterlegten Bestellintervalle seine Lieferungen besser planen kann (Stichwort “Regelmäßigkeit der Nachfrage”).

Insbesondere bei kleineren Einkaufsabteilungen kann zumeist nur ein geringer Teil der Lieferanten durch den strategischen Einkauf intensiv betreut und einem laufenden Controlling unterzogen werden. Wie bereits im Kapitel “E-Procurement” erwähnt, wird der Fokus dabei eher auf Lieferantenbeziehungen für direkte Güter liegen, da diese wertmäßig und für den Unternehmenserfolg von größerer Bedeutung sind. Um auch für die Beschaffung indirekter Güter eine Prozeßbeschleunigung zu erreichen, sowie darüberhinaus optional eine Einkaufssteuerung und Lieferantenoptimierung durchführen zu können, werden elektronische Kataloge benötigt, die einen effizienten Zugang zu den o.a. Produktinformationen ermöglichen.

Die Überlegenheit elektronischer Kataloge im Vergleich zu traditionellen Formen wurde unter 2.4.2 bereits angedeutet; zur Verdeutlichung hier noch eine Gegenüberstellung nach einer Studie von Forrester Research [For98]:

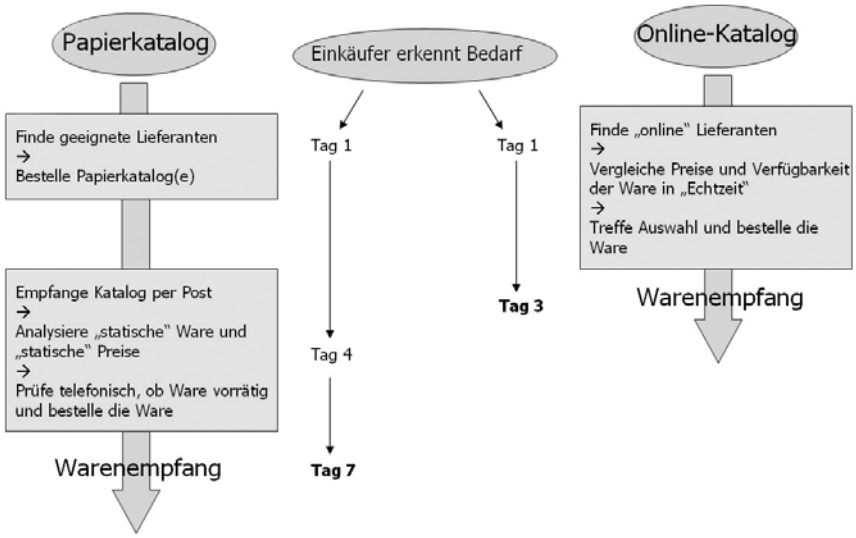

Abbildung 5: Papier -vs. Online-Katalog

Innerhalb eines E-Procurement-Systems stellt die Katalogkomponente ein relativ eigenständiges Modul dar. Dennoch ergibt sich eine Rückkopplung zu einigen übergordneten Systemen, insbesondere zum Beschaffungssytem, in das die über virtuelle Warenkörbe vom Katalogbenutzer erzeugten Bedarfsanforderungen übertragen werden. Hier werden die Bedarfsanforderungen optional durch einen Genehmigungsworkflow geschleust und anschließend(bzw. nach erteilter Genehmigung) in entsprechende Bestellungen umgewandelt.

3.2 Katalogarten

Die elektronischen Kataloge lassen sich grob in zwei Arten einteilen, die sich anhand der Verwaltung des Kataloginhalts voneinander unterscheiden.

3.2.1 Single-Supplier Katalog

Bei sog. Single-Supplier Katalogen wird je Lieferant ein eigener Katalog eingerichtet; dieser kann “inhouse” gepflegt oder -was eher der Regel entspricht- vom Lieferanten selbst oder einem Out-Sourcing

Provider vorgehalten werden; in letzterem Fall entstehen innerhalb der beschaffenden Organisation keine bzw. geringere Kosten für das Management der Kataloginhalte.

3.2.2 Multi-Supplier Katalog

Ein Multi-Supplier Katalog integriert die Produktdaten mehrerer Lieferanten. Dies eröffnet u.a. die Möglichkeit, eine Konkurrenzsituation zwischen den eingebundenen Lieferanten zu schaffen, da ein direkter Vergleich der "normalisierten" (d.h. einheitlich präsentierten) Produkte anhand von Ausstattungsmerkmalen oder Preisen vereinfacht wird. Man beachte dabei aber, daß aufgrund bestehender Verträge u.U. bestimmte Lieferanten explizit aus dieser Konkurrenzsituation ausgeklammert werden müssen; zu diesem Zweck kann man beispielsweise festlegen, daß bestimmte Produktgruppen nur von einem Lieferanten angeboten werden.

Generell gilt bei diesem Katalogtyp, daß alle Daten unter einer einheitlichen Oberfläche zur Verfügung stehen, was zu höherer Ergonomie und in der Folge zu besserer Endbenutzerakzeptanz führt. Auch die Möglichkeit, transparent über die Produkte verschiedener Lieferanten suchen zu können, trägt entscheidend zur Benutzerfreundlichkeit und Effizienz des Gesamtsystems bei.

Die genannten Vorteile werden allerdings durch deutlich höhere Kosten erkauft, da die Integration der verschiedenen Kataloge -abhängig von der Datenqualität- eine mehr oder weniger aufwendige Content Management (CM) Software erforderlich macht; dabei ist es auch zunächst einmal unerheblich, ob der Katalog vom beschaffenden Unternehmen selbst oder einem Dienstleister verwaltet wird.

Hier noch einmal die Vorteile der beiden Katalogtypen in einer Gegenüberstellung

- Single-Supplier Katalog
  - keine bzw. geringere Kosten für das CM (Wegfall des Integrationsaufwands)
  - relativ einfache Anbindung neuer Lieferanten (grundsätzlich Unabhängigkeit von deren Katalogformaten bzw. -strukturen)
- Multi-Supplier Katalog
  - Strategische Gründe (wie das Schaffen einer Konkurrenzsituation)
  - einfacher Vergleich normalisierter Produkte
  - einheitliche Struktur und Oberfläche, damit hohe Enbenutzerakzeptanz
  - Transparente Suche

Die Grobeinteilung in Single-Supplier- und Multi-Supplier Katalog kann nun nach verschiedenen Kriterien verfeinert werden; im folgenden wird nach der Organisation bzw. dem Ort der Datenpflege differenziert.

3.2.3 Externer Katalog

Diese Organisationsform basiert auf dem Single-Supplier Modell. Die Lieferanten bauen hierbei jeweils eigene Kataloge auf, die sie ihren Kunden zur Verfügung stellen. Damit ergeben sich aus Kundensicht folgende Vor- und Nachteile:

- **Vorteile**
  - die Kosten für das Katalogmanagement trägt der Lieferant
  - ständiger Zugriff auf die aktuellsten Artikel- bzw. Preisinformationen
  - Online-Verfügbarkeitsprüfungen und -Statusabfragen sind möglich

- **Nachteile**
  - die Datenhoheit liegt beim Lieferant, d.h. es besteht eine Abhängigkeit von der vom Lieferanten gewählten Struktur und Funktionalität (man denke z.B. an Suchfunktionen).
  - aus dem letzten Punkt ergibt sich auch, daß der Katalog nicht bzw. nur sehr eingeschränkt um kundenspezifische Informationen erweitert werden kann.
  - der Katalogzugriff erfolgt über das Internet, woraus sich die bekannten Probleme, z.B. was Sicherheit und Zuverlässigkeit angeht, ergeben.

3.2.4 Interner Katalog

Hierbei handelt es sich um das Gegenstück zum externen Katalog. Interne Kataloge basieren auf dem Multi-Supplier-Modell, wobei die Einkaufsorganisation die von ihren Lieferanten zur Verfügung gestellten Daten in einen eigenen Katalog integriert. Einige Vor- und Nachteile dieser Vorgehensweise sind im folgenden aufgeführt:

- **Vorteile**
  - Diese Organisationsform bietet die größtmögliche Kontrolle über die angzeigten Produktdaten, da die Datenhoheit bei der Einkaufsorganisation liegt.
  - da der Katalog lokal gehostet wird, kann den Sicherheits- und Zuverlässigkeitsaspekten des Katalogzugriffs eher Rechnung getragen werden.
  - Die Daten kleinerer Lieferante ohne eigene Kataloge sind ebenfalls integrierbar.
  - Zusätzliche Informationen (wie eigene Materialnummern) können leicht im internen Katalog hinterlegt werden.
- **Nachteile**
  - Kostenintensivste und zeitaufwendigste Lösung, da der Aufbau und die Pflege eines eigenen CM-Systems erfoderlich sind.
  - Aufgrund dieses hohen Aufwands sind interne Kataloge i.a. nur für eine relativ geringe Anzahl von Produkten handhabbar.
  - es ist kein unmittelbarere Zugriff auf die Daten der Lieferanten möglich, d.h. es stehen nicht ständig die aktuellsten Informationen zur Verfügung.

3.2.5 Dienstleister-Prinzip

Das Dienstleister-Prinzip stellt in diesem Zusammenhang eine komplexere Ausprägung des Multi-Supplier-Modells dar. Beim Dienstleister handelt es sich dabei i.a. um den Betreiber eines elektronischen Marktplatzes, der die Produkt-Informationen der angeschlossenen Lieferanten sammelt und in einen Multi-Supplier-Katalog integriert, auf den die teilnehmenden Einkaufsorganisationen zugreifen können; optional werden dabei zusätzliche Konvertermodule zwischengeschaltet, so daß die Daten transparent im jeweils benötigten Format zur Verfügung stehen. Das Dienstleister-Prinzip zeichnet sich durch folgende Vor- und Nachteile aus:

- **Vorteile**
  - es stehen relativ kostengünstig (im Vergleich zu internen Katalogen) standardisierte und damit vergleichbare Produkt-Informationen zur Verfügung.
  - das vorhandene Know-How des Dienstleisters kann genutzt werden
  - wie beim externen Katalog ist kein CM-System auf Seiten der Einkaufsorganisation erforderlich.

- **Nachteile**
  - es besteht eine direkte Abhängigkeit vom Marktplatzbetreiber und seiner Lieferantenauswahl.
  - kundenspezifische Informationen und die o.g. Konvertermodule sind i.a. mit zusätzlichen Kosten verbunden.
  - analog zum externen Katalog wird man mit den dort bereits genannten Problemen des Zugriffs via Internet konfrontiert.

3.3 Katalog-Integration

Wie schon mehrfach erwähnt, beruhen alle Varianten des Multi-Supplier-Modells auf derselben Kernidee, nämlich der Integration von Produktdaten verschiedener Lieferanten in einen Katalog. Zu deren Realisierung setzt man häufig Import-Tools ein, die in einem sog. Staging-Prozeß arbeiten.

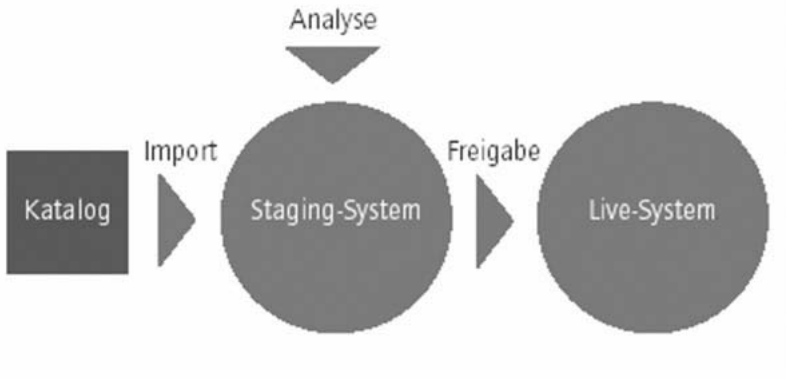

Abbildung 6: Import-Tool mit Staging-Komponente

Dabei werden die externen Daten zunächst in das Staging-System geladen, wo sie mithilfe der CM-Software technisch aufbereitet werden. Daran schließt sich optional ein Genehmigungsworkflow an, an dessen Ende –bei positivem Resultat– die Freigabe der Daten zur Aufnahme ins Live-System bzw. den Produktivkatalog steht.

Bei der technischen Datenaufbereitung sind verschiedene Aufgaben zu erfüllen, wie

- eine **Normalisierung** der Daten, d.h. Aufbereitung lieferantenspezifischer Abkürzungen und Begriffe entsprechend einer standardisierten Terminologie; ein einfaches Beispiel wäre die Ersetzung von “schw.” durch “schwarz”.
- **Lokalisierungsmaßnahmen**, d.h. Anpassung an lokale bzw. nationale Gegebenheiten, z.B. bei der Darstellung von Fließkommazahlen (→ “12.5” wird zu “12,5”).
- eine **Rationalisierung** der Daten; in diesem Zusammenhang versteht man darunter eine Anordnung der Artikelbezeichnungen und -merkmale in einer definierten Reihenfolge, z.B. die Umwandlung von “schw. Tacker” in “Klammerhefter, schwarz” (wobei hier zuätzlich eine Normalisierung der Begriffe durchgeführt wurde).
- die **Integration der Daten in das eigene Klassifikationsschema**. Bei der Klassifizierung werden Produkte mit ähnlichen Eigenschaften zu Klassen bzw. Kategorien zusammengefaßt, die in einem hierarchischen Schema (Taxonomie) angeordnet werden; durch die Anwendung

dieses Verfahrens werden die Kataloge strukturierter und damit auch für den Anwender leichter zugänglich.

Die Katalog-Integration kann durch die Definition bzw. Verwendung von Standards erheblich erleichtert werden. Dies gilt insbesondere für die o.g. Klassifizierung sowie die Übertragung der Katalogdaten. Das folgende Beispiel soll die Wichtigkeit solcher Standards verdeutlichen:

> Ein Unternehmen hat 100 Lieferanten, die ihre Kataloge allesamt in individuellem Format verbreiten. Für jeden dieser Kataloge ist nun ein eigenes Konvertermodul erforderlich, das die Daten extrahiert und den eigenen Anforderungen entsprechend aufbereitet[6]. Dabei muß im ungünstigsten Fall für jedes einzelne Produkt das Mapping in die passende Kategorie des eigenen Katalogs definiert werden, was bei einem Update des Lieferantenkatalogs evtl. wieder angepaßt werden muß.

Ein Lösungsansatz für das Klassifizierungsproblem wird in einem späteren Teil dieses Papiers dargestellt, jedoch lassen sich alle geschilderten Schwierigkeiten durch Verwendung von Standards auch einfach umgehen. Zwei solche Standards, die jeweils von diversen, teils führenden Unternehmen getragen bzw. unterstützt werden, sind nun Gegenstand der folgenden beiden Abschnitte.

3.3.1 BMEcat

BMEcat[7] ist ein Standard für den Austausch von Katalogen zwischen Lieferanten und beschaffenden Organisationen. Der Lieferant stellt dazu die Katalogdaten in einem Dokument zusammen, das dem BMEcat-Format entpricht. Dieses basiert vollständig auf XML und definiert über DTDs [BMEDTD] neben der generellen Dokumentstruktur u.a.

- Muß- und Kannfelder,
- die zugehörigen Datentypen und Feldlängen,
- 3 Katalogtransaktionen, nämlich T_NEW_CATALOG, T_UPDATE_PRODUCTS und T_UPDATE_PRICES

Die Katalogtransaktionen bestimmen das Aussehen des sog. Transaktionsteils des BMEcat-Dokuments; die folgende Abbildung veranschaulicht dies am Beispiel von T_NEW_CATALOG:

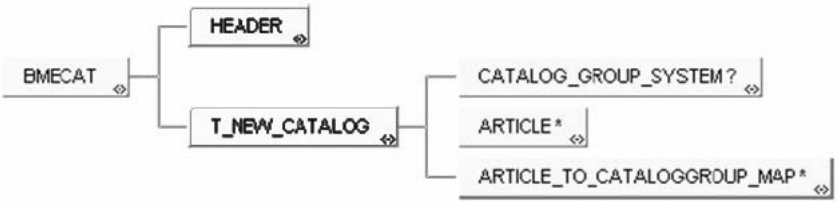

Abbildung 7: XML-Struktur von T_NEW_CATALOG

Jedes Kästchen repräsentiert ein XML-Element, wobei optionale Elemente mit " ? " gekennzeicnet sind. Man erkennt, daß der Transaktionsteil über die Kindknoten CATALOG_GROUP_SYSTEM, ARTICLE und ARTICLE_TO_CATALOGGROUPMAP verfügt[8], wobei durch " * " spezifiziert wird,

[6] bei jeder Erweiterung der Lieferantenmenge kommen neue Konverter hinzu.

[7] BME ist eine Abkürzung für "Bundesverband Materialwirtschaft, Einkauf und Logistik e.V.". Prominente Unterstützer und Mitentwickler von BMEcat sind z.B. Siemens, BMW, DaimlerChrysler und Philips.

[8] daneben existieren noch weitere mögliche Kindknoten, deren vollständige Erläuterung aber den Rahmen dieses Papiers sprengen würde.

daß ARTICLE und ARTICLE_TO_CATALOGGROUPMAP mehrfach auftreten dürfen. Den einzelnen Elementen kommen dabei folgende Aufgaben zu:

- CATALOG_GROUP_SYSTEM:

  Hierüber kann eine Taxonomie von Produktklassen für das Katalogdokument definiert werden. Dazu wird jede einzelne Klasse mithilfe des Elements CATALOG_STRUCTURE angelegt, wobei PARENT_ID die GROUP_ID des Vorgängers in der (Baum-)Hierarchie angibt.

  Über GROUP_SYSTEM_ID und GROUP_SYSTEM_NAME kann die definierte Taxonomie eindeutig gekennzeichnet bzw. beschrieben werden, wobei der Lieferant eine für sich selbst eindeutige Kennung verwenden sollte.

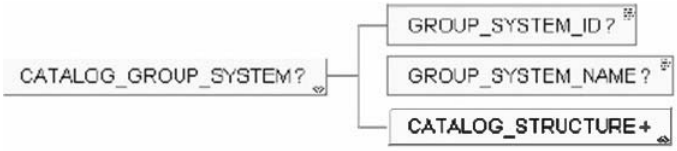

Abbildung 8: Element CATALOG_GROUP_SYSTEM

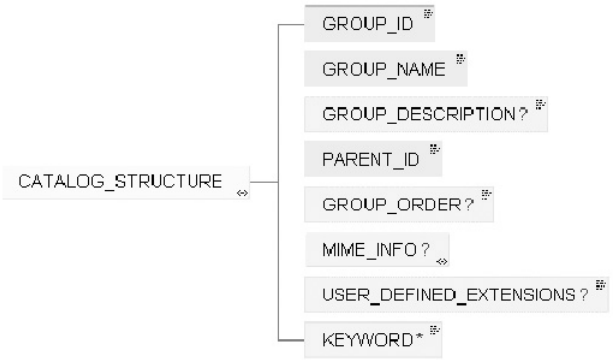

Abbildung 9: Element CATALOG_STRUCTURE

- ARTICLE:

  Mithilfe dieses Elements können die einzelne Produkte des Kataloges beschrieben werden; dazu gehören die Nummer des jeweiligen Artikels (d.h. eine eindeutige Kennung aus Sicht des Lieferanten), Details zum Artikel selbst und zu Bestell- und Lieferkonditionen, sowie diverse Preisinformationen (Einzel- und Staffelpreise, Besteuerung, u.ä). Darüberhinaus können über das Element MIME_INFO optional erweiterte Artikeldaten referenziert werden, wie z.B. ein PDF-Datenblatt oder ein Produktvideo.

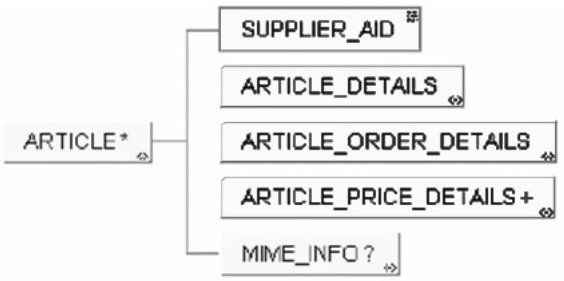

Abbildung 10: Element ARTICLE

- ARTICLE_TO_CATALOGGROUP_MAP:

  Hiermit werden die einzelnen (durch ARTICLE repräsentierten) Produkte den (via CATALOG_STRUCTURE) definierten Kataloggruppen zugeordnet. Das optionale Element ARTICLE_TO_CATALOGGROUP_MAP_ORDER legt dabei über einen Integer-Wert die Reihenfolge fest, in der die Artikel einer Kataloggruppe im Zielsystem dargestellt werden sollen.

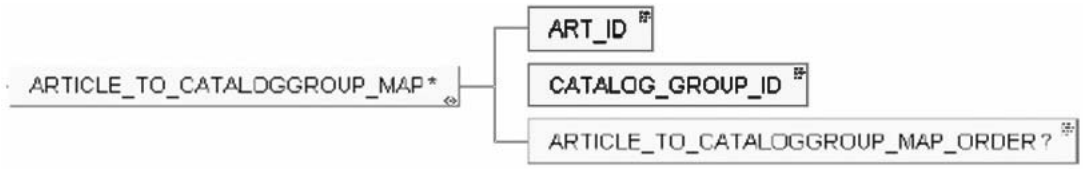

Abbildung 11: Element ARTICLE_TO_CATALOGGROUP_MAP

Die beiden anderen Transaktionen sind analog aufgebaut:

Abbildung 12: XML-Struktur von T_UPDATE_PRODUCTS und T_UPDATE_PRICES

Die einzelnen Elemente haben dieselbe Struktur und Semantik wie bei T_NEW_CATALOG, enthalten jetzt aber nur noch die jeweils zu aktualisierenden Daten.

Neben dem Transaktionsteil verfügt jedes BMEcat-Dokument noch über einen HEADER, der als Container für “globale” Daten dient (vgl. Abbildung 13).

Dabei ist insbesondere das CATALOG-Element von Bedeutung, das über CATALOG_ID auf einen bereits beim Empfänger bestehenden Katalog referenzieren kann; so kann mittels T_NEW_CATALOG beispielsweise eine neue Version angelegt (spezifiziert durch CATALOG_VERSION) oder das Ziel für eine Update-Transaktion festgelegt werden.

Die beiden anderen Elemente BUYER und SUPPLIER können in ihrer Funktion mit der Anschrift bzw. dem Absender des Katalogdokuments verglichen werden: BUYER enthält Daten zum Empfänger, SUPPLIER zum erstellenden Lieferanten.

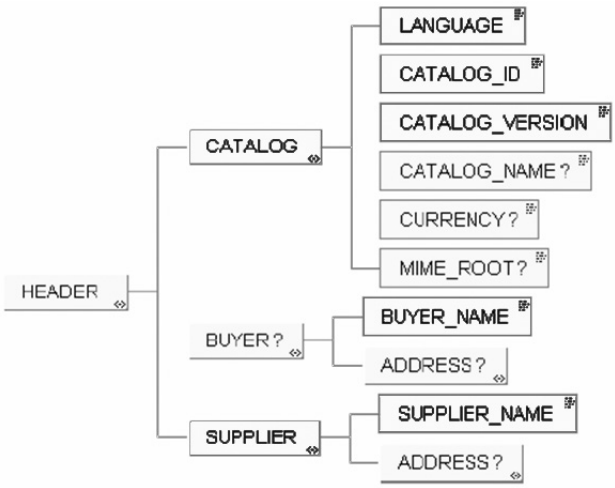

Abbildung 13: XML-Struktur des BMEcat-Headers

Zum Abschluß noch zwei Anmerkungen zu BMEcat:

- Die Codierung von Produktdaten in BMEcat-konformem XML stellt für viele Lieferanten (v.a. kleinere Unternehmen) einen hohen Aufwand dar; daher ist das ausschlaggebende Kriterium für die Verwendung dieses Standards i.a. ein hohes Kundeninteresse, d.h. eine große Zahl von Kunden muß die Übertragung von Katalogdokumenten in entsprechendem Format fordern. Daraus ergibt sich, daß nur eine schnelle Verbreitung die Überlebensfähigkeit von BMEcat auf Dauer sicherstellen wird.
- BMEcat bietet die Möglichkeit, jeden beliebigen Standard zur Produktklassifikation zur verwenden. Dazu kann im oben beschriebenen Transaktionsteil-Element ARTICLE ein zusätzlicher Kindknoten ARTICLE_FEATURES angegeben werden, wo über REFERRENCE_FEATURE_-SYSTEM_NAME das gewünschte Klassifikationssystem spezifizierbar ist; Abbildung 14 macht dies am Beispiel von eCl@ss deutlich, das auch Gegenstand des nächsten Abschnitts sein wird[9].

```
<ARTICLE>
  <SUPPLIER_AID>Bohrer 124242</SUPPLIER_AID>
      ....
  <ARTICLE_FEATURES>
    <REFERENCE_FEATURE_SYSTEM_NAME>eclass
        </REFERENCE_FEATURE_SYSTEM_NAME>
    <REFERENCE_FEATURE_GROUP_ID>21-01-01-01
        </REFERENCE_FEATURE_GROUP_ID>
    <FEATURE>
      <FNAME>Werkstoff, Kurzname</FNAME>
      <FVALUE>X5CrNi18-10</FVALUE>
    </FEATURE>
    <FEATURE>
      <FNAME>Gewicht</FNAME>
      <FVALUE>0,13</FVALUE>
      <FUNIT>kg</FUNIT>
    </FEATURE>
      ....
  </ARTICLE_FEATURES>
      ....
</ARTICLE>
```

Abbildung 14: BMEcat in Verbindung mit eCl@ss

[9]zur Verwendung von eCl@ass in BMEcat exisitiert auch ein Kataloggenerator-Tool, verfügbar unter *http://ocf.cacontent.com/ocf/ocflogin.asp*

3.3.2 eCl@ss

eCl@ass ist nach eigener Definition ein “Hierarchisches System zur Gruppierung von Materialien [...] entsprechend der Produktcharakteristik oder -merkmale” [ECLPRES]. Die dabei verwendete Hierarchie gliedert sich in “Sachgebiete”, “Hauptgruppen”, “Gruppen” sowie auf unterster Ebene die “Untergruppen”, denen die eigentlichen Produkte zugeordnet werden; Abbildung 15 skizziert die so entstehende Struktur:

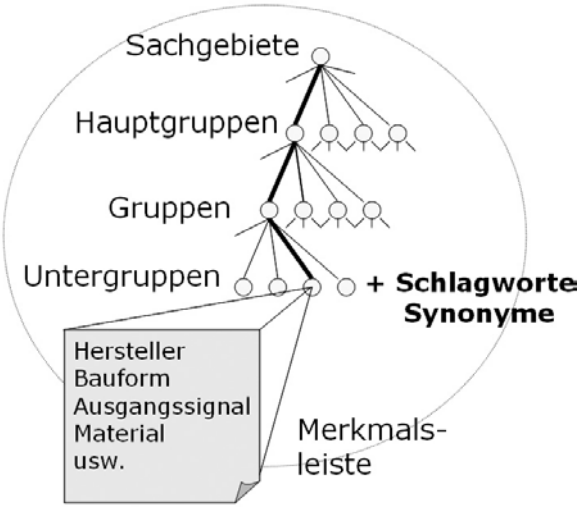

Abbildung 15: Schema der eCl@ss-Hierarchie

Jede Untergruppe verfügt neben einer Sammlung von Schlagworten (i.a. Synonyme der jeweiligen Untergruppenbezeichnung) über eine sog. Merkmalsleiste, die als Obermenge der Attribute aller zuordbaren Produkte fungiert. Die Zuordnung selbst erfolgt über einen vierteiligen Klassifikationsschlüssel, wobei die einzelnen Bestandteile den Ebenen des Hierarchiebaums entsprechen; Abbildung 16 zeigt ein Beispiel dieses Konzepts:

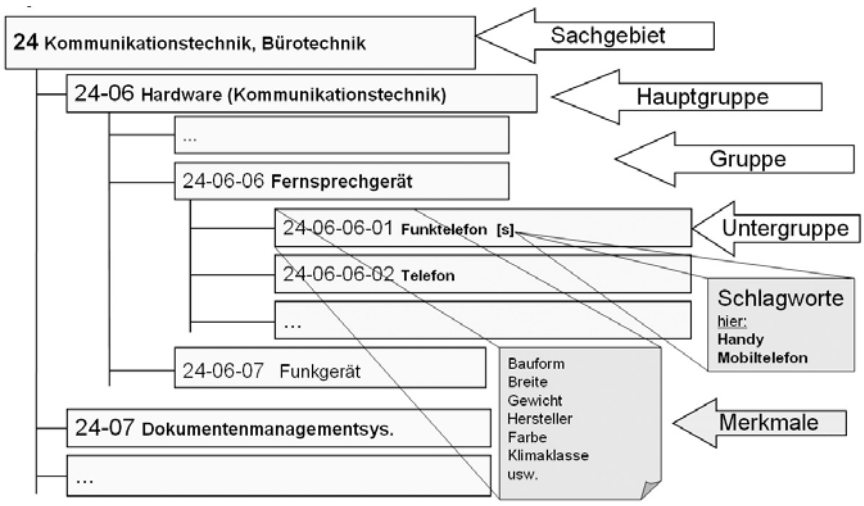

Abbildung 16: Skizze eines eCl@ss-Sachgebiets

Die Verwendung von eCl@ss folgt nun grob folgendem Schema (nach [ECLAPP]):

- Öffnen der Internet-Homepage von eCl@ss: *www.eclass.de*
- Eingabe eines Schlagwortes in die entsprechende Zeile und Auslösen der Suche, z.B. “Schraube”. Das Ergebnis beim Beispiel Schraube zeigt mehrere Sachgebiete (21; 23; 36)
- Auswahl des Sachgebietes, das das eigene Angebot am besten trifft (das Sachgebiet 21 wäre z.B. zutreffend, wenn man Schraubenschlüssel, Schraubendreher, u.ä. anbieten will)
- Überprüfung des gefundenen Sachgebietes anhand der Hierarchie (damit stellt man sicher, dass die Begrifflichkeiten mit den eigenen Vorstellungen übereinstimmen).
- Öffnen der entsprechenden Merkmalleisten zur Überprüfung der für die eigenen Produkte zutreffenden Attribute (man findet hier neben einer Definition auch weitere wichtige Informationen, wie z.B. die zu verwendende Einheit des jeweiligen Merkmals).
- Durchführung der eigentlichen Produkt-Untergruppen-Zuordung und Belegung der entprechenden Attribute.

Abschließend soll noch einmal der Vorteil der Verwendung von Standards für die Katalogintegration am Beispiel von eCl@ss verdeutlicht werden; dazu betrachten wir zwei Szenarien:

Kunde und Lieferant vewenden beide eCl@ss
In diesem Fall sind alle Produktdaten des übertragenen Katalogs mit dem jeweiligen vierteiligen Klassifikationsschlüssel versehen und über eine Teilmenge der standardisierten Merkmalsleiste der entsprechenden Untergruppe beschrieben; der Kunde kann diese Daten nun ohne Änderung in seinen Produktivkatalog übernehmen.

Der Lieferant verwendet keinen Klassifikationsstandard[10]
Dieses Szenario wurde unter 3.3 bereits erläutert; wichtig ist dabei die Tatsache, daß die Kategorieintegration größenordnungsmäßig manuell kaum handhabbar ist (bei größeren Katalogen finden sich durchaus mehrere hundert Klassen mit hoher Produktanzahl), aber auch keine einfache Datenübernahme wie bei Verwendung von eCl@ss möglich ist. Dieses Problem ist nun Gegenstand des folgenden Abschnitts, welcher auf [Ag01] basiert.

3.3.3 Automatisierte Kategorieintegration

An dieser Stelle soll zunächst einmal das schon mehrfach erwähnte Problem der Katgorieintegration genauer formuliert werden:

- gegeben sind ein Ziel-Katalog Z und ein Quell-Katalog Q mit jeweiligen Produktklassen Z_1, ..., Z_m bzw. Q_1, ..., Q_n.
- finde für jedes Produkt *Prod* aus Q eine Klasse in Z. Dabei soll die implizite Information der ursprünglichen Klassenzuordung von *Prod* in Q genutzt werden, d.h. man argumentiert, daß Produkte, die in Q in eine gemeinsame Klasse eingordnet wurden, sich in gewisser Hinsicht ähnlich sein müssen und somit auch in Z eher in dieselbe Klasse fallen sollten als Produkte aus unterschiedlichen Kategorien.
- Annahmen und Einschränkungen:
  - man geht von semantisch verwandten Klassifikationsmustern in Z und Q aus, da sonst die o.g. implizite Information nicht nutzbar bzw. vorhanden wäre; so würde eine Klassifikation nach Herstellern in Q keine oder nur geringe Aussagekraft für einen nach Produktart gegliederten Ziel-Katalog haben.

[10] würde er einen anderen Klassifikationsstandard als der Kunde verwenden, so wäre die Integration ebenfalls problematischer als im ersten Szenario; man beachte aber, daß man in diesem Fall i.a. auf entsprechende Mapping-Tools zwischen den entsprechenden Standards zurückgreifen kann.

- man betrachtet die Taxonomien als "flach", d.h. die Hierarchiestruktur wird in diesem Zusammenhang ignoriert.
- unterschiedliche Detaillierungsgrade der Taxonomien von Z und Q wirken sich wie folgt aus:
  - ist Z detaillierter als Q, so nützt die implizite Information von Q nur noch bedingt für die Einordung in Z: angenommen, Z hat getrennte Klassen für "Sportwagen" und "Limousinen", während Q nur eine Klasse "Autos" verwaltet, so ist keine implizite Information im obigen Sinne zur Unterscheidung zwischen "Sportwagen" und "Limousinen" gegeben; dagegen könnte eine Differenzierung dieser beiden Kategorien und "LKWs" durch Q implizit unterstützt werden.
  - bei höherem Detaillierungsgrad von Q ist es denkbar, zunächst eine Verschmelzung bestimmter Quell-Klassen durchzuführen, um so eine vergleichbare Granularität der beiden Kataloge zu erreichen.

Die Idee ist nun, einen Classifier zu konstruieren, der die Produktdaten von Z und das Klassifikationsschema *beider* Kataloge als Grundlage für die Klassenzuordnung von Produkten aus Q in Z nutzt.

Als Ausgangspunkt dazu dient hier die **naive Bayes-Klassifikation**: die Wahrscheinlichkeit, daß ein Produkt $Prod$ zur Klasse Z_i gehört, wird danach abgeschätzt durch

$$P(Z_i|Prod) = \frac{P(Z_i)P(Prod|Z_i)}{P(Prod)}$$

$P(Prod)$ kann dabei auch ignoriert werden, da sie von Z_i unabhängig ist, d.h. die relativen Wahrscheinlichkeiten für die Klassenzugehörigkeit von $Prod$ dadurch nicht beeinflußt werden.

Als Abschätzung für $P(Z_i)$ wird

$$P(Z_i) = \frac{\#\ Produkte\ in\ Z_i}{\#\ Produkte\ in\ Z\ insgesamt}$$

verwendet, was sich leicht berechnen läßt.

Zur Ermittlung von $P(Prod|Z_i)$ betrachtet man die Termzerlegung der Produktbeschreibung von $Prod$ (=: $T(Prod)$) und nimmt zur Vereinfachung an, daß die einzelnen Terme unabhängig voneinander auftreten[11]; damit ergibt sich

$$P(Prod|Z_i) = \prod_{Term\ \in\ T(Prod)} P(Term|Z_i)$$

Sei nun $n(Z_i, Term)$ die absolute Häufigkeit des Terms $Term$ in Z_i und $n(Z_i) = \sum_{Term} n(Z_i, Term)$ die Gesamtzahl von Termen in Z_i. Dann erhält man als mögliche Abschätzung[12] für $P(Term|Z_i)$:

$$P(Term|Z_i) \ = \ \frac{n(Z_i, Term)}{n(Z_i)}$$

Auf dem naiven Ansatz aufbauend wird nun eine "**erweiterte Bayes-Klassifikation**" konstruiert. Dabei soll die Wahrscheinlichkeit, daß ein Produkt $Prod$ aus der Quell-Klasse Q_j zur Zielklasse Z_i gehört, abgeschätzt werden, also

$$P(Z_i|Prod,\ Q_j) \quad = \quad \frac{P(Z_i,\ Prod,\ Q_j)}{P(Prod,\ Q_j)}$$

[11] genauer müßte man eigentlich noch zwischen den Termen der Produktbeschreibung und den –meist getrennt aufgeführten– Attributen des Produktes unterscheiden.

[12] darüberhinaus müßte man eigentlich noch den Fall betrachten, daß ein Term t gar nicht in Z_i auftritt, und somit nach obiger Schätzung $P(Prod|Z_i) = 0$ gelten, falls $t \notin T(Prod)$. Zur Vereinfachung werden wir dieses Problem hier aber nicht weiter verfolgen.

$$
\begin{aligned}
&= \frac{P(Z_i)P(Q_j,\ Prod|Z_i)}{P(Prod,\ Q_j)} \\
&= \frac{P(Q_j)P(Z_i|Q_j)P(Prod|Z_i)}{P(Prod,\ Q_j)},\ da\ P(Z_i|Q_j)P(Q_j)\ =\ P(Q_j|Z_i)P(Z_i) \\
&= \frac{P(Z_i|Q_j)P(Prod|Z_i)}{P(Prod|Q_j)}
\end{aligned}
$$

Hier kann $P(Prod|Q_j)$ ignoriert werden, da sich die relativen Wahrscheinlichkeiten –anlog zu $P(Prod)$ bei der naiven Klassifikation– dadurch nicht verändern.

Eine einfache Abschätzung für $P(Z_i|Q_j)$ wäre

$$
P(Z_i|Q_j)\ =\ \frac{\#\ Dokumente\ aus\ Q_j,\ die\ nach\ der\ naiven\ Klassifikation\ in\ Z_i\ fallen}{\#\ Dokumente\ in\ Q_j\ insgesamt}
$$

Diese kann nun noch um eine Gewichtung erweitert werden, die die semantische Änlichkeit der Klassifikationsmuster von Z und Q widerspiegelt und damit entscheidet, wie stark die implizite Information von Q berücksichtigt wird; eine ausführliche Betrachtung dieses Konzepts würde allerdings den Rahmen dieses Papiers sprengen und wird daher hier übergangen.

Basierend auf den obigen Überlegungen kann man nun folgenden Classifier konstruieren:

foreach($q\ \in \{Q_1, ..., Q_n\}$)
{
  foreach($Prod\ \in\ q$)
  {
    klassifiziere $Prod$ temporär mithilfe der naiven Bayes-Klassifikation
    ermittle damit $P(Z_i|Q_j)$ wie oben beschrieben
  }
  reklassifiziere jedes Produkt aus q mithilfe der erweiterten Bayes-Klassifikation
}

Die Fähigkeit einer e-Catalog-Lösung, standardisierte und evtl. auch nicht-standardisierte Fremdaten –z.B. mithilfe des oben beschriebenen Classifiers– zu integrieren, stellt allerdings nur einen Gesichtspunkt bei der Entscheidung über ihre Verwendung in einem Unternehmen dar. Im folgenden sollen daher weitere wichtige Kriterien und Anforderungen für elektronische Kataloge dargelegt werden.

3.4 e-Catalogs – Kriterien und Anforderungen

3.4.1 Allgemeine Kriterien für die Auswahl eines Kataloges

Das wichtigste allgemeine Kriterium ist wohl die Anwenderfreundlichkeit, da die beste Software einem Unternehmen nichts nützt, wenn sie von den Mitarbeitern nicht akzeptiert bzw. nicht benutzt wird. Die Kosten spielen natürlich auch eine sehr große Rolle, wobei darunter v.a. Aufwendungen für Lizenzgebühren, sowie Katalogaufbau und -pflege fallen. Die Ausfallsicherheit sollte möglichst hoch sein, damit dem Unternehmen keine unnötigen Kosten entstehen. Das System sollte skalierbar, d.h. nicht schon bei einer gleichzeitigen Nutzung von 10 Mitarbeitern zusammenbrechen; darüberhinaus sollte es vor Fremdeingriffen von außen geschützt sein. Zu guter letzt sollte Investitionssicherheit für das Unternehmen bestehen, damit beispielsweise kein Geld für die Lizensierung eines Systems bezahlt wird, das nach kurzer Zeit nicht mehr gewartet werden kann.

3.4.2 Anforderungen an die Katalogsoftware

Eine „gute“ Katalogsoftware sollte über umfangreiche Suchfunktionen verfügen, auf die wir aber später noch etwas genauer zu sprechen kommen. Artikellisten gleichartiger Produkte sind sinnvoll, damit diese leicht verglichen werden können. Es sollten gängige Standards unterstützt werden, um für die Zukunft flexibel zu sein. Eine Verfügbarkeitsprüfung ist sinnvoll, damit der Benutzer prüfen kann, ob das gewünschte Produkt vorhanden ist und wann es voraussichtlich geliefert werden kann. Bei Unterstützung von Stücklistenfunktionalität werden z.B. bei der Bestellung eines PCs automatisch dessen Einzelkomponenten vom System bestellt. Schliesslich sollte die Katalogsoftware über ein Rollenkonzept verfügen, damit schon von vornherein bestimmt werden kann, welche Benutzer welche Produkte bestellen können.

3.4.3 Anforderungen an die Content-Management Software

Die Content-Management Software sollte mit verschiedenen Katalog-Systemen nutzbar sein, damit das Unternehmen nicht von einer Katalog-Lösung abhängig ist. Ausserdem sollte die Content-Struktur frei wählbar sein, d.h. das Unternehmen kann selbst entscheiden, wie es seine Daten archiviert. Dies stellt keinen Widerspruch zur Benutzung von Standards dar, da ja jede Struktur auf einen enstprechenden Standard abgebildet werden kann. Die CM-Software sollte wennmöglich Schnittstellen zu ERP-Systemen, einem Data Warehouse und zu anderen Informations-Diensten beinhalten. Ein externer Zugriff durch Lieferanten ist wünschenswert, damit diese bequem ihre eigenen Daten einpflegen können.

3.4.4 Anforderungen an ein „gutes“ Katalog-Suchsystem

Eine Suchmöglichkeit, die nie fehlen darf, ist die Schlüsselwortsuche, die in folgenden Variationen existiert:

- Ähnlichkeitssuche, z.B: Suche nach orthographisch ähnlichen Wörtern (interessant bei Tippfehlern).
- Phonetische Suche, also Suche nach ähnlich klingenden Wörtern
- Boolesche Suche
- Iterative Suche, d.h. die schrittweise Verfeinerung der Suche durch eine enstprechende Benutzerführung
- Thesaurussuche bzw. Suche über einer Ontologie (→ Synonyme, Hypernyme u.ä.)

Bei der parametrisierten Suche ist die Semantik und häufig auch die Wertemenge jedes Eingabefeldes fest vorgegeben (z.B. “erweiterte Suche” bei Amazon.de mit enstrechend gekennzeichneten Feldern für Autor, Buchformat, ISBN usw.).

Ausserdem sollten zukünftige Suchsysteme Queries unterstützen, mit denen der Benutzer seine persönlichen Präferenzen ausdrücken kann. Preference SQL, womit wir uns in Kapitel 4 näher beschäftigen, stellt eine Möglichkeit zur Verfügung, genau dies zu erreichen.

4 Preference SQL

4.1 Einführung

Wie schon in Kapitel 3.4.4 angedeutet wird die Fähigkeit, Benutzerpräferenzen abbilden zu können, in der Zukunft ein wichtiges Kriterium für „gute“ (Katalog-)Suchsysteme darstellen. Genauso wie

beim „realen“ Einkauf in einem Laden, hat ein Kunde seine persönlichen Kriterien und Präferenzen, die die Suche nach dem „idealen“ Produkt lenken. Dabei können die Kriterien in Muss- und Wunsch-Kriterien aufgeteilt werden; Muss-Kriterien müssen dabei, wie der Name schon sagt, auf jeden Fall und Wunsch-Kriterien eben so gut wie möglich erfüllt werden. Ein Kunde, der in einen realen Laden einkaufen geht, erwartet dort einen hilfsbereiten Mitarbeiter, der ihm beim Finden des für ihn am besten geeignetsten Produktes unterstützt bzw. berät und genau die gleiche Erwartungshaltung für „gute“ Kundenbetreuung überträgt sich auf e-shops im Internet.

Leider ist die derzeitig vorherrschende Situation im Internet weit entfernt von diesem Idealzustand, weswegen gängige B2C und B2B e-shops Kundenpräferenzen nicht angemessen behandeln können. Als Konsequenz resultieren aus einer „e-shop session“ sehr häufig frustrierte Benutzer, da allzuoft keine bzw. keine vernünftigen Ergebnisse als Resultat auf eine Suchanfrage hin zurückgegeben werden, bei der der Benutzer versucht hat, seine persönlichen Präferenzen möglichst genau abzubilden. Für den e-shop-Betreiber ist dies natürlich eine sehr schädigende Situation, denn im allgemeinen reichen schon wenige missglückte Versuche dazu aus, daß der Benutzer das Portal bzw. den e-shop verlässt und auch für längere Zeit nicht mehr benutzt [13].

Diese Problematik blieb natürlich im Laufe der Zeit nicht unentdeckt, führte bis dato aber lediglich zu einer Reihe von „ad-hoc“ Lösungen, von denen zwei am gängisten sind.

Die erste „Lösung“ zwingt den Benutzer mehr oder weniger radikal, von seinen eigentlichen Präferenzen Abstand zu nehmen, indem er Suchfelder unspezifiziert lassen muß (alle ausgefüllten Felder werden automatisch als „harte“ Kriterien in die Suche miteinbezogen). Die Folge davon ist allerdings, daß der Benutzer meist mit unrelevanten Informationen überhäuft wird, was im Vergleich zu einem leeren Suchergebnis eher eine bescheidene Verbesserung darstellt.

Die zweite „Lösung“ ist die bereits erwähnte iterative Suche, bei der der Suchraum konsekutiv eingeschränkt und der Benutzer interaktiv durch den Suchprozess geführt wird; bei leeren Suchergebnissen ist dabei ein „Backtracking“ möglich. Obwohl dadurch die Anzahl der leeren Suchergebnisse deutlich verringert wird, ist der gesamte Suchprozess sehr zeitaufwendig und verlangt dem Benutzer ständige Aufmerksamkeit ab.

Preference SQL ist nun der (kommerzielle) Versuch, ein semantisch intuitives Modell von Präferenzen zu entwickeln, das eine passende formale Semantik zur problemlosen und effizienten Integration in SQL bietet.

4.2 Preference SQL Sprach-Design

Suchsysteme, die mit „normalem“ SQL implementiert sind haben den Nachteil, daß SQL nicht direkt den Begriff der Präferenz versteht und unterstützt. Deswegen müssen Präferenzen irgendwie in „harte“ SQL-Kriterien der Where-Klausel übersetzt werden. Das Präferenzmodell von PSQL entstand aus der Erkenntniss, daß Präferenzen, so wie wir sie auch im Alltag verwenden (z.B. „ich mag A lieber als B“), relativ einfach auf ein mathematisches Modell, nämlich auf *strikte partielle Ordnungen* abgebildet werden können. Mathematisch gesehen, ist eine Präferenz

$$P = (A, <P)$$

eine irreflexive, transitive und asymmetrische binäre Relation $<P$ auf einer Domäne von Werten einer Attributmenge A. Für eine ausführliche Erläuterung des mathematischen Grundmodells von PSQL wird auf [PSQL1] verwiesen, da wir uns in diesem Papier auf die für uns relevanten Aspekte beschränken wollen, die für das intuitive Verständnis von PSQL ausreichen.

Man kann also sagen, daß PSQL eine Erweiterung von SQL um eine Präferenzsyntax ist, bei der der Benutzer die Möglichkeit hat, seine Präferenzen abzubilden:

$$\boxed{\text{Preference SQL} = \text{Standard SQL} + \text{Preferences}}$$

[13] Quelle: Forrester Research

Die Syntax von PSQL entspricht deswegen auch im wesentlichen der Syntax von SQL, wobei Preferenzen nun durch das Schlüsselwort „PREFERRING“ spezifiziert werden können:

SELECT	<selection>
FROM	<table_references>
WHERE	<where_conditions>
PREFERRING	**<soft_conditions>**
GROUPING	<attribute_list>
BUT ONLY	<but_only_condition>
ORDER BY	<attribute_list>

Die Rückgabesemantik einer PSQL-Query richtet sich nach dem „**B**est **M**atches **O**nly“-Modell, d.h. es werden nur die „besten“ Tupel zurückgegeben, also alle, die von keinem anderen Tupel dominiert werden.

4.2.1 Integrierte Präferenz Basis-Typen

Im folgenden geben wir einen Überblick über die wichtigsten eingebauten Präferenz Basis-Typen:

1. Approximation: AROUND, BETWEEN
   „AROUND“-Präferenzen bevorzugen Werte, die einem numerischen Wert am nächsten kommen, was z.B. sehr nützlich sein kann, wenn es schwer ist, einen genauen Wert anzugeben. Die nachfolgende Beispiel-Query liefert Ausflüge zurück, wenn möglich mit einer genauen Dauer von 14 Tagen und wenn nicht, jene mit einem kürzesten Abstand zu 14 Tagen:

   ```
   SELECT * FROM trips
   PREFERING duration AROUND 14;
   ```

   „BETWEEN[untere Grenze, obere Grenze]“-Präferenzen verhalten sich analog: Werte, die sich innerhalb des Intervals befinden werden als beste Treffer betrachtet und andernfalls Werte, die sich in nächster Nähe zu einer Intervall-Grenze befinden.

2. Minimierung/Maximierung: LOWEST, HIGHEST
   Häufig vorkommende Präferenzen verlangen nach dem höchsten bzw. niedrigsten Wert. Die nachfolgende Query sucht nach Wohnungen, wobei eine möglichst große Wohnfläche präferiert wird:

   ```
   SELECT * FROM apartments
   PREFERRING HIGHEST(area);
   ```

   Diese Query hat ein einfaches Pendant in SQL, wobei allerdings zu bemerken ist, daß anstelle eines einzelnen Attributes wie in diesem Fall „area“, auch arithmetische Ausdrücke und Prozedur-Aufrufe verwendet werden können.

3. Favorites, dislikes: POS, NEG
   Eine POS-Präferenz drückt aus, daß der gewünschte Wert in einer spezifizierten Liste von Werten vorkommen soll:

   ```
   SELECT * FROM programmers
   PREFERRING exp IN ('java', 'C++');
   ```

   Diese Query präferiert Programmierer, die schon Erfahrung in einer der Programmiersprachen Java oder C++ haben, aber falls es keine mit den gewünschten Kenntnissen gibt, werden auch andere betrachtet und zurückgeliefert.
   Durch eine NEG-Präferenz kann man ausdrücken, daß ein Attribut möglichst nicht einen bestimmten Wert haben soll. Durch die nachfolgende Query werden Hotels bevorzugt, die nicht in „Downtown“ liegen, aber falls es nur dort Hotels gibt, die Zimmer frei haben, ist es immer noch besser eines von diesen anzubieten, als gar keins:

```
SELECT * FROM hotels
PREFERRING location <> 'downtown';
```

4.2.2 Zusammenbau von komplexen Präferenzen

Mit Hilfe der eingebauten „einfachen“ Basis-Typen und geeigneten Operatoren können nun induktiv komplexe Präferenzen gebildet werden, und zwar zum einen mit gleicher, zum anderen mit ungleicher Gewichtung zueinander.

Gleiche Gewichtung: „Pareto Akkumulation“ (AND) Eine Pareto Akkumulation von Präferenzen $P_1, ..., P_n$ in eine komplexe Präferenz P ist wie folgt definiert:

> $v = (v_1, ..., v_n)$ ist besser als $w = (w_1, ..., w_n) \Leftrightarrow \exists\ i$, so daß v_i besser als w_i und v ist gleich oder besser als w in allen anderen Komponenten.

Die „maximalen“ Werte von P werden als die **Pareto-optimale Menge** bezeichnet. In der Syntax von PSQL werden die einzelnen Präferenzen mit gleicher Gewichtung einfach mit „AND“ verknüpft. Als Beispiel könnte man sich die Suche nach einem Computer vorstellen, wobei die Größe des Hauptspeichers und die Geschwindigkeit der CPU für den Benutzer gleich wichtig sind:

```
SELECT * FROM computers
PREFERRING HIGHEST(main_memory)
        AND HIGHEST(cpu_speed);
```

Geordnete Gewichtung: Kaskadierung von Präferenzen (CASCADE) Eine Kaskadierung von Präferenzen weist den einzelnen Präferenzen verschiedene Level mit absteigender Wichtigkeit zu, was zur Folge hat, daß die Präferenzen in ihrer vorgegebenen Reihenfolge angewendet werden. Eine geordnete Reihenfolge von Präferenzen wird in PSQL mit dem Schlüsselwort CASCADE eingeleitet. Als einfaches Beispiel betrachten wir folgende Query, bei der es wieder um die Suche nach einem Computer geht, wobei die Farbe eine unwichtigere Rolle als der Hauptspeicher spielt:

```
SELECT * FROM computers
PREFERRING HIGHEST(main_memory)
    CASCADE color IN ('black', 'brown');
```

Die Semantik dieser Query ist also wie folgt: Zunächst wird der PREFERRING-Block ausgewertet. Falls es nun mehrere Computer mit der gleichen Größe an Hauptspeicher gibt, so bleiben nach Anwendung des CASCADE-Blockes entweder nur noch jene mit Farbe schwarz oder braun übrig, weil diese die anderen dominieren; falls es keine mit entsprechender Farbe gibt, werden alle Computer zurückgegeben, da alle „gleich gut“ sind.

Kombination von Pareto Akkumulation und Kaskadierung Durch die Kombination von gleichgewichteten und ungleichgewichteten Präferenzen können in PSQL nun sehr mächtige Queries konstruiert und dadurch sehr spezifische Wünsche ausgedrückt werden. Als Beispiel für die intuitive Ausdruckskraft von PSQL wollen wir zunächst einen Kundenwunsch in natürlicher Sprache formulieren und dann die entsprechende PSQL-Query analysieren:

> „Mein Wunschauto *muss* ein Opel sein. Es *sollte* ein „Roadster“ sein, aber falls kein solcher gefunden werden kann, bitte keinen Kombi. Gleich wichtig ist für mich, daß das Auto um die 20.000 Euro kostet und es so viel PS wie möglich hat. Weniger wichtig für mich ist die Farbe, wobei ich aber ein rotes Auto präferiere. Falls dann noch mehrere Autos zur Auswahl stehen, soll die geringste Kilometeranzahl entscheiden.“

Wie man nun an der nachfolgenden PSQL-Query erkennen kann, ist diese fast eine eins-zu-eins Übersetzung der verbalen Formulierung:

```
SELECT * FROM car
WHERE make = 'Opel'
PREFERRING (category = 'roadster' ELSE
                category <> 'station wagon')
        AND price AROUND 20000
        AND HIGHEST(power)
   CASCADE color = 'red'
   CASCADE LOWEST(mileage);
```

4.2.3 Qualitätsfunktionen

Ob ein Tupel in der Lösungsmenge einer PSQL-Query vorkommt oder nicht, hängt im Gegensatz zu SQL nicht nur von der Qualität des Tupels selbst, sondern auch von der Qualität der „Mitstreiter-Tupel“ ab. Deswegen benötigt man in PSQL sogenannte Qualitätsfunktionen, durch die berechnet werden kann, wie gut ein Tupel eine Präferenz erfüllt:

- TOP(A): 1 $\Leftrightarrow$ A perfekter Treffer (boolesche Funktion)
- LEVEL(A): Wie weit ist der Attributwert A von dem „besten“ Attributwert (Level 1) entfernt (nicht numerischer Wert)?
- DISTANCE(A): Wie weit ist der Attributwert A von dem „besten“ Attributwert (Distanz 0) entfernt (numerischer Wert)?

Betrachten wir dazu folgende Tabelle:

oldtimer	ident	color	age
	Maggie	white	19
	Bart	green	19
	Homer	yellow	35
	Selma	red	40
	Smithers	red	43
	Skinner	yellow	51

Die nachfolgende PSQL-Query Pareto-kombiniert eine POS/POS-Präferenz mit einer AROUND-Präferenz, wobei zur Verdeutlichung der Qualitätsfunktionen zusätzlich noch LEVEL(color) und DISTANCE(age) selektiert werden:

```
SELECT ident, color, age,
       LEVEL(color),
       DISTANCE(age)
FROM   oldtimer
PREFERRING color = 'white' ELSE
               color = 'yellow'
       AND age AROUND 40;
```

Die Pareto-optimale Menge sieht dann wie folgt aus:

ident	color	age	level	distance
Maggie	white	19	1	21
Homer	yellow	35	2	5
Selma	red	40	3	0

Anhand der Qualitätsfunktionen kann man nun sofort erkennen, wie gut jedes der drei Tupel jedes einzelne Kriterium erfüllt, und daß jedes andere Tupel, das nicht zu der Pareto-optimalen Menge gehört von mindestens einem der drei „Best-Matches“ dominiert wird.

4.2.4 Qualitätskontrolle

In PSQL können die Qualitätsfunktionen auch dazu benutzt werden, eine gewisse Mindestqualitätsanforderung an ein Lösungstupel zu stellen. Als Beispiel stellen wir uns einen Kunden vor, der nach einem Ausflug sucht, der um den 3.Juli herum beginnt und ca. 14 Tage dauert, wobei er bei beiden Kriterien jeweils eine Abweichung von höchstens zwei Tagen akzeptiert. Solch eine Qualitätsanforderung kann in PSQL durch die „BUT ONLY“-Klausel ausgedrückt werden:

```
SELECT * FROM trips
PREFFERING start_day AROUND '2003/7/3'
        AND duration AROUND 14
BUT ONLY DISTANCE(start_day) <= 2
      AND DISTANCE(duration) <= 2;
```

Die „BUT ONLY“-Klausel stellt Muss-Kriterien bzgl. der Lösungs-Tupel einer Preferring-Klausel dar, was jetzt natürlich dazu führen kann, daß die Ergebnis-Menge leer ist. In diesem Fall ist dies aber explizit von dem Benutzer berücksichtigt.

4.3 Implementation von PSQL

4.3.1 Plug-and-Go Application Integration

Preference SQL wird als Zwischenschicht zwischen der Applikation und der SQL-Datenbank implementiert, wie man auch auf folgender Grafik erkennen kann:

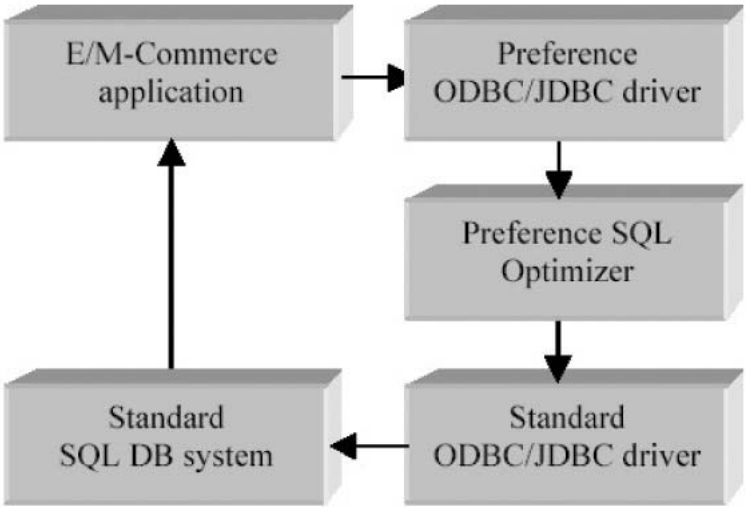

Abbildung 17: Preference SQL Integration

PSQL-Queries werden zunächst vom PSQL-Optimizer in „normalen“ SQL'92-kompatiblen Code übersetzt und anschliessend an die Datenbank gesendet. Queries ohne Präferenzen werden einfach durch den PSQL-Optimizer geschleust, ohne größere Performanceeinbußen zu erleiden. Die Anbindung von PSQL an vorhandene Datenbank-Applikationen gestaltet sich deswegen relativ einfach.

4.3.2 Preference SQL Optimizer

Als Beispiel für die Funktionsweise des PSQL-Optimizers betrachten wir folgende Query

```
SELECT * FROM cars
PREFERRING Make = 'Audi'
        AND Diesel = 'yes';
```

auf einer Tabelle cars:

cars	Identifier	Make	Model	Price	Mileage	Airbag	Diesel
	1	Audi	A6	40000	15000	yes	no
	2	BMW	5 series	35000	30000	yes	yes
	3	Volkswagen	Beetle	20000	10000	yes	no

Sie kann in folgender SQL92-kompatiblen Query ausgedrückt werden:

```
CREATE VIEW Aux AS
SELECT *, CASE WHEN Make = 'Audi'
               THEN 1 ELSE 2 END
               AS Makelevel,
          CASE WHEN Diesel = 'yes'
               THEN 1 ELSE 2 END
               AS Diesellevel
FROM Cars;

INSERT INTO Max
SELECT Identifier, Make, Model,
       Price, Mileage, Airbag,
       Diesel
FROM Aux A1
WHERE NOT EXISTS
   (SELECT 1 FROM Aux A2
    WHERE A2.Makelevel <= A1.Makelevel
      AND A2.Diesellevel <= A1.Diesellevel
      AND (A2.Makelevel < A1.Makelevel
           OR
           A2.Diesellevel < A1.Diesellevel));
```

Wie man sieht, wird zunächst ein Hilfs-View gebildet, durch den zusätzlich noch die Qualitätsfunktionen ausgewertet werden. Danach werden alle Tupel, die von keinem anderen Tupel dominiert werden, in die Relation Max eingefügt. Es wird also jedes Tupel mit jedem anderen Tupel verglichen, wobei aber derzeitige SQL-Optimizer unter Verwendung geeigneter Indizes auch diese komplexe Query relativ effizient auswerten können. Die Durchführung und Auswertung eines Benchmarks über einer Job-Börse, bei dem das Ergebnis festgestellt wurde, daß SQL-Queries im Vergleich zu entsprechenden PSQL-Queries nicht wesentlich schneller sind, kann in [PSQL2] nachgelesen werden.

4.4 Anwendungsbeispiel

Bei dem Design eines Präferenz-Suchportals müssen folgende Kriterien beachtet werden:

- Welche Selektionskriterien sind „hart“ (WHERE-Klausel), welche sind „weich“ (PREFERRING-Klausel)?

- Qualitätskontrolle (BUT ONLY-Klausel)?
- Wichtigkeit der Präferenzen (Pareto Akkumulation vs. Kaskadierung)?
- Woher kommen die Präferenzen: vom Benutzer selbst oder vom e-service provider, d.h. transparent für den Benutzer?

Als Anwendungsbeispiel, betrachten wir uns das in Abbildung 18 dargestellte Suchportal für gebrauchte Autos, bei dem die Präferenz-Logik transparent für den Benutzer ist:

Abbildung 18: PSQL Demo-Portal mit Beispiel-Query

Die darin dargestellte Query kann dann z.B. wie folgt in PSQL abgebildet werden:

```
SELECT *, TOP(manufacturer)
          TOP(model), TOP(price),
          TOP(mileage), TOP(regyear),
          TOP(diesel), TOP(airbag),
          TOP(autotransmission),
          TOP(aircondition)
FROM used_cars
PREFERRING manufacture = 'BMW'
        AND model = '7'
```

```
CASCADE price BETWEEN 0, 75000
        AND mileage BETWEEN 0, 30000
        AND regyear BETWEEN 1997, 1999
        AND diesel = 'yes'
        AND airbag = 'yes'
        AND autotransmission = 'yes'
        AND aircondition = 'yes';
```

In diesem Beispiel wurden der Hersteller und das Modell zueinander gleichgewichtet und gegenüber den anderen Kriterien, die untereinander auch wiederum gleiches Gewicht haben, als wichtiger betrachtet. Zusätzlich wird zu jedem Attribut noch die TOP-Qualitätsfunktion ausgewertet, um in der Ausgabe farblich hinterlegen zu können, ob der Wert zu einem Attribut ein perfekter Treffer ist oder nicht:

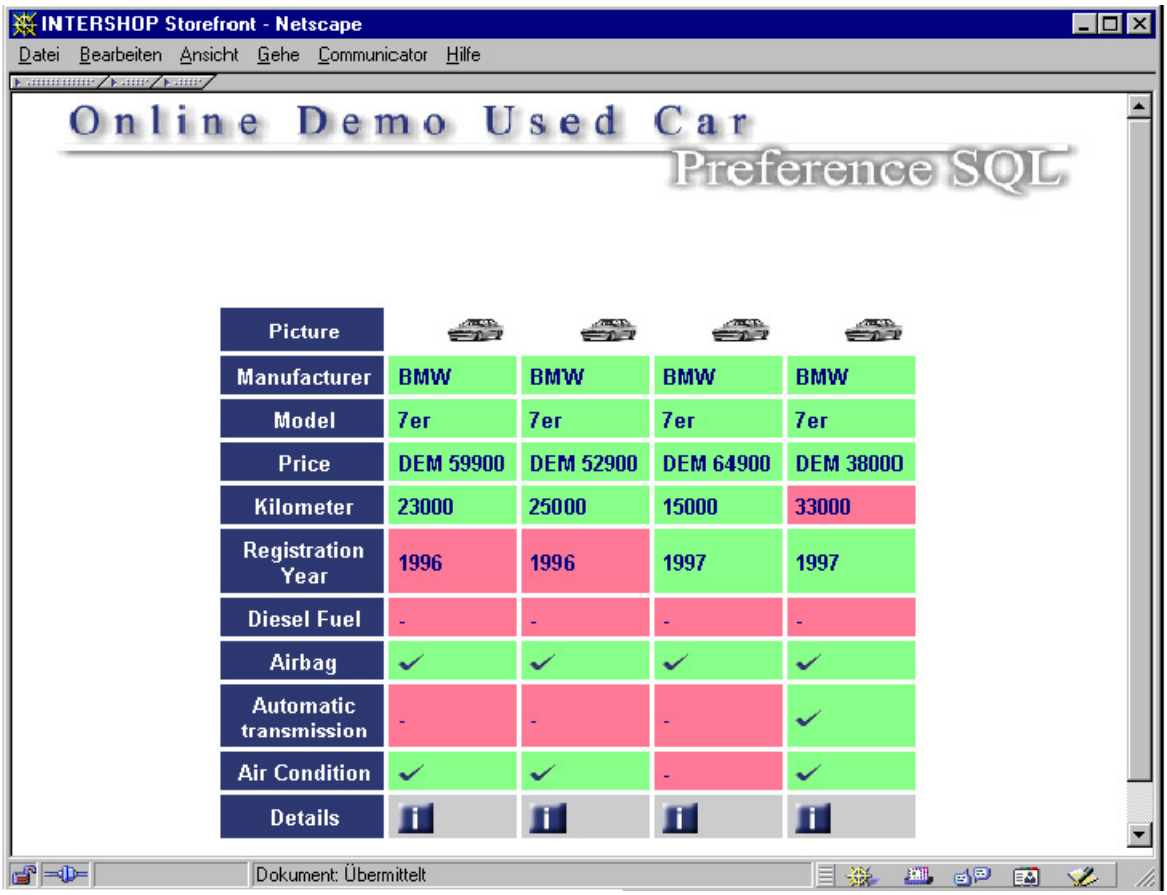

Abbildung 19: Rückgabe der Beispiel-Query

Wie man sehen kann, konnten nicht alle Kriterien gleichzeitig erfüllt werden, aber anstatt dem Benutzer zu sagen „Es konnten keine Ergebnisse gefunden werden!“, können durch Preference SQL interessante Alternativen angeboten werden, wobei die als am wichtigsten angesehenen Attribute „Hersteller“ und „Modell“ sogar noch erfüllt werden konnten.

5 Fazit

Abschliessend möchten wir für jeden großen Teilbereich dieses Papiers noch ein kurzes Fazit angeben:

1. **E-Procurement:** Es hat sich gezeigt, daß es im E-Procurement noch sehr viele Einsparpotentiale, vor allem im Bereich der indirekten Materialbeschaffung gibt. Wie wichtig E-Procurement in der Zukunft sein wird, kann man anhand einer Schätzung der Unternehmensberatung IDC ableiten, die für das Jahr 2003 250 Mio. Nutzer von E-Procurement-Lösungen weltweit prognostizieren.

2. **Catalog Content Management:** In diesem Kapitel sollten die Grundideen elektronischer Kataloge und deren Rolle in einem E-Procurement System klar geworden sein. Dabei liegt die technische Herausforderung ihrer Verwendung in der Integration von mehreren Katalogen, wobei zum einen –wann immer möglich– Standards genutzt werden sollten, zum anderen aber auch (semi-)automatische Integrationsverfahren zur Verfügung stehen, die den Umgang mit nicht standardisierten Daten zumindest erleichtern.

3. **Preference SQL:** Preference SQL stellt unserer Meinung nach einen echten Mehrwert für (Katalog-)Suchsysteme dar, wobei die Integration in vorhandene SQL-Applikationen durch die Implementation einer Zwischenschicht sehr einfach zu bewerkstelligen ist. Der einzige Nachteil von PSQL ist, daß es eine kommerzielle Erweiterung von SQL darstellt und deswegen nicht einfach kostenlos benutzt werden kann.

Abbildungsverzeichnis

1 Kostenstrukturen für direkte und indirekte Güter 5
2 Der traditionelle Beschaffungsprozess 6
3 Der Beschaffungszyklus durch E-Procurement 7
4 Die E-Procurement-System Architektur am Beispiel von SAP EBP 12
5 Papier -vs. Online-Katalog 13
6 Import-Tool mit Staging-Komponente 16
7 XML-Struktur von T_NEW_CATALOG 17
8 Element CATALOG_GROUP_SYSTEM 18
9 Element CATALOG_STRUCTURE 18
10 Element ARTICLE 19
11 Element ARTICLE_TO_CATALOGGROUP_MAP 19
12 XML-Struktur von T_UPDATE_PRODUCTS und T_UPDATE_PRICES 19
13 XML-Struktur des BMEcat-Headers 20
14 BMEcat in Verbindung mit eCl@ss 20
15 Schema der eCl@ss-Hierarchie 21
16 Skizze eines eCl@ss-Sachgebiets 21
17 Preference SQL Integration 30
18 PSQL Demo-Portal mit Beispiel-Query 32
19 Rückgabe der Beispiel-Query 33

Literatur

[PWC02] PricewaterhouseCoopers Unternehmensberatung GmbH: *Wegweiser Katalogmanagement – Wesentlicher Erfolgsfaktor für E-Procurement-Projekte*. 2002.

[ACT01] ACTIO Managementberatung GmbH: *Die E-Procurement-Strategie für Ihr Unternehmen – richtiges Einkaufen unter Nutzung des Internets*. 2001.

[BSB00] John P. Baron, Michael J. Shaw, Andrew D. Bailey, Jr.: *Web-based E-catalog Systems in B2B Procurement*; In: Communications of the ACM, Vol. 43, Nr. 5. Mai 2000.

[SAP02] SAP AG: *SAP Enterprise Buyer Professional, Release 3.0 – Functions in Detail*, Whitepaper. 2002.

[For98] Forrester Research: *Business Catalog Strategies*. Februar 1998.

[BMEDTD] eBusiness Standardization Comittee: *DTDs zu BMEcat V1.2*. Vgl. auch *www.bmecat.org*.

[ECLPRES] *eCl@ss - Standard für Materialklassifikation und Warengruppen*. Vgl. *www.eclass.de*.

[ECLAPP] *Infos für mittelständische Nutzer*. Vgl. *www.eclass.de*.

[Ag01] R. Agrawal, R. Srikant: *On integrating catalogs*. WWW10, Hong Kong, Mai 2001.

[PSQL1] W. Kießling: *Foundations of Preferences in Database Systems*. Proc. 28^{th} Intern. Conf on Very Large Databases (VLDB), Hong Kong, August 2002.

[PSQL2] W. Kießling, G. Köstler: *Preference SQL — Design, Implementation, Experiences*. Proc. 28^{th} Intern. Conf on Very Large Databases (VLDB), Hong Kong, August 2002.