

DAS PROJEKT LOSTART.DE

EINE INTERNET-DATENBANK FÜR KULTURGUTVERLUSTE

Kai-Uwe Sattler, Eike Schallehn, Ingo Schmitt, Nadine Schulz

Die Suche nach Kulturgütern, die infolge des Zweiten Weltkrieges und des Nationalsozialismus geraubt wurden oder verloren gingen, ist auch heute noch eine aktuelle Aufgabe – nicht nur für Kunsthistoriker, sondern auch für betroffene Privatpersonen, Institutionen und natürlich die Politik. In diesem Beitrag wird das „Lost Art“-Projekt vorgestellt, in dessen Rahmen eine Web-Datenbank zur Unterstützung dieser Suche entwickelt wurde. Die Datenbank umfasst eine Vielzahl von Informationen zu den registrierten Kulturgütern und erlaubt unterschiedliche Such- und Navigationsmodi in verschiedenen Sprachen. Ausgehend von der Architektur dieses Systems werden Aspekte der Implementierung, der Recherchemöglichkeiten sowie des Datenaustausches zwischen der öffentlichen Web-Datenbank und der eigentlichen internen Datenbank beschrieben.

Die Geschichte von Kulturgütern, die in der Zeit von 1933 bis ca. 1949 enteignet, gestohlen oder als Kriegstrophäen verbracht wurden, wirft bis heute offene Fragen auf. Weltweit existieren unzählige Kunstwerke, zum Teil verschlossen in Depots oder privaten Sammlungen, deren Eigentümerschaft ungeklärt ist. Bis auf den heutigen Tag ist die Rückführung einzelner Objekte oder spezieller Sammlungen Bestandteil internationaler politischer Verhandlungen. Und nach über fünfzig Jahren gibt es immer noch regelmäßig Meldungen, dass Kunstwerke nach langer Suche ihrem rechtmäßigen Eigentümer zurückerstattet werden konnten.

Dabei kann die Geschichte eines einzelnen Objektes während der Zeit des Nationalsozialismus in Deutschland und bedingt durch die Kriegswirren sehr viele Stationen umfassen, die oft nicht dokumentiert wurden. Durch spezielle Sonderkommandos und Einsatzstäbe der Wehrmacht, SS, NSDAP und andere nationalsozialistische Kunstrauborganisationen wurden Kunstgüter von „Juden und anderen Reichsfeinden“ in ganz Europa zusammengeraubt oder erpresst, um sie den Kunstsammlungen der darum miteinander konkurrierenden NS-Größen einzuverleiben. So war z. B. in Linz, wo Hitler seine Jugendzeit verbracht hatte, seit 1938 ein „Führermuseum“ geplant, die dazu vorgesehenen Kunstwerke in einer so genannten „Linzer Liste“ erfasst. Auch der Reichsmarschall Göring tat sich als „Kunstsammler“ ganz besonderer Art hervor. Ein Teil dieser Kunstschätze konnte nach Kriegsende sichergestellt werden und wurde in den von den Alliierten eingerichteten Central Collecting Points (CCP) zusammengetragen. Es existieren zum Beispiel aber auch Gemälde, die durch die Nationalsozialisten aus jüdischem Besitz enteignet wurden, in den Bestand deutscher Museen übergangen und von dort am Ende des Zweiten Weltkrieges durch sowjetische Trophäenkommissionen in die Sowjetunion verbracht wurden. Zum Teil kam es vor, dass diese Gemälde entwendet und später

wieder illegal ohne korrekten Herkunftsnachweis auf dem Kunstmarkt auftauchten. Ein Bestand von über 2 000 Kulturgütern mit ungeklärten Eigentumsverhältnissen, die somit auch den ursprünglichen Eigentümern bisher nicht zurückgegeben werden konnten, befindet sich heute noch im Bundesvermögen. Ähnliche Sammlungen existieren in einigen Ländern, wenn auch mit anderen geschichtlichen Hintergründen.

Kulturgüter mit unklaren Herkunftsverhältnissen stellen nicht nur für die heutigen Besitzer, sondern auch für das kulturelle Leben weltweit ein Problem dar. Sie können nicht auf dem offiziellen Kunstmarkt gehandelt werden, eine Ausstellung ist oft problematisch und somit bleiben zahlreiche einmalige Kunstwerke der Öffentlichkeit unzugänglich. In den betroffenen Ländern arbeiten deshalb verschiedene Institutionen an der Dokumentation und Publikation von Such- und Fundmeldungen, mit dem Ziel, den Verbleib oder die Eigentumsverhältnisse der Objekte zu klären.

In Deutschland wird diese Aufgabe von der Koordinierungsstelle für Kulturgutverluste, welche in Magdeburg ansässig ist, wahrgenommen. Dort werden Informationen zu Such- und Fundmeldungen verwaltet, die zum Beispiel die erwähnten Objekte der „Linzer Liste“, aber auch Verluste aus deutschen Museen oder Privatbesitz umfassen. Insgesamt sind dort summarisch über drei Millionen Objekte erfasst und zum Teil lie-

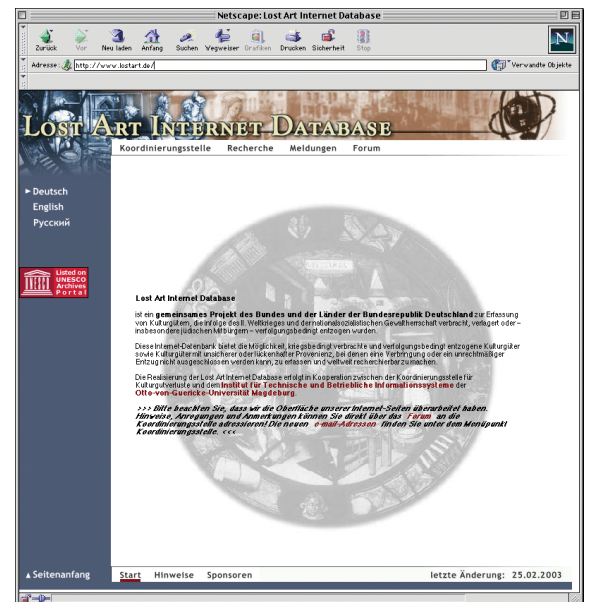


Abb. 1
Die Datenbank ist unter www.lostart.de öffentlich zugänglich.

gen dazu detaillierte Verlustgeschichten und Bildmaterial vor. Um den Verbleib beziehungsweise die Eigentumsverhältnisse klären zu können, ist es jedoch dringend notwendig, diese Informationen einem möglichst großen Kreis von Interessenten zugänglich zu machen. Deshalb entschied sich die Koordinierungsstelle, ihre Daten in Kooperation mit dem Institut für Technische und Betriebliche Informationssysteme der Otto-von-Guericke-Universität Magdeburg über das World Wide Web zugänglich zu machen. Im „Lost Art“-Projekt wurde ein entsprechendes Informationssystem zu kulturhistorischen Daten umgesetzt, das über www.lostart.de zu erreichen ist. Neben umfangreichen Recherchemöglichkeiten werden weitere Aspekte wie Mehrsprachigkeit, Diskussionsforen und die Meldung von Verlusten oder Funden über das Web unterstützt.

weitere Informationen nicht mehr vorliegen. Umgesetzt wurde deshalb eine Mischung aus klassischen Datenbanktechniken, Bildsuche und Information Retrieval, unter besonderer Berücksichtigung der Ähnlichkeit multimedialer als auch beschreibender Informationen.

Zwei weitere interessante Aspekte ergaben sich Ende des Jahres 2001 in Magdeburg auf einer internationalen Konferenz von Institutionen, die wie die Koordinierungsstelle jeweils in ihren Ländern vergleichbare Such- und Fundmeldungen verwalten oder publizieren. Einerseits erschwert die mittlerweile recht große Anzahl dieser Datenbanken die Recherche, da auf Grund der oft recht komplexen Verlustgeschichten meist alle diese Datenbanken durchsucht werden müssen. Hier erscheint es sinnvoll, einen einheitlichen Zugriffspunkt zu schaffen, über den Daten in allen Systemen gesucht werden können. An der Magdeburger Universität wurde ein entsprechender Prototyp einer Meta-Suchmaschine entwickelt, der nun in Kooperation mit anderen Institutionen ausgebaut werden soll. Der zweite interessante Punkt ergibt sich, da Daten zu gesuchten oder vorhandenen Objekten in verschiedenen Datenbanken, d. h. mehrfach enthalten sein können. Es ist sogar anzunehmen, dass Suchmeldungen zu einem Objekt in der einen Datenbank bereits als Fundmeldungen in der anderen Datenbank vorliegen. Ein ähnlichkeitsbasierter Abgleich der Daten könnte dazu dienen, dies zu erkennen und dadurch Kulturgüter wieder ihren rechtmäßigen Eigentümern zuzuführen.

Im Folgenden sollen einige der im Projekt verwendeten und entwickelten Techniken im Anwendungskontext vorgestellt werden.

ARCHITEKTUR UND TECHNISCHE UMSETZUNG

Die Datenbank der Koordinierungsstelle wurde ursprünglich auf Basis von MS Access realisiert. Bei der Entwicklung der Web-Datenbank wurde aber schnell deutlich, dass diese Lösung für die geplanten Recherchemöglichkeiten und die zu erwartende Zugriffslast nicht ausreichend ist. Daher wurde der Datenbestand zunächst in eine Oracle8-Datenbank migriert, wobei auch Probleme des Ausgangsschemas beseitigt werden konnten. Gleichzeitig eröffnete dieser Schritt die Möglichkeit, die von Oracle angebotenen Erweiterungsmodule zur Volltext- und Bild-Ähnlichkeitssuche zu nutzen. Speziell die von Suchmaschinen bekannte Form der Volltextrecherche hat sich inzwischen als sehr sinnvoll erwiesen, da viele Attribute der verwalteten Objekte längere Texte umfassen, beispielsweise zur textuellen Beschreibung der Objekte oder zur Verlustgeschichte.

Neben den öffentlich zugänglichen Informationen enthält die interne Datenbank auch Daten, die vertraulich behandelt werden müssen. Hierzu zählen u. a. nicht gesicherte Annahmen über den Verbringungsort einzelner Objekte, deren Veröffentlichung nicht ohne politische Konsequenzen wäre. Da darüber hinaus die „Datenhoheit“ der Koordinierungsstelle gewahrt bleiben sollte, die technisch-organisatorische Einbindung in eine



Abb. 2
Zurückgeführte Gemälde
„Landschaft mit Hirten, Pferd,
Schaf und Kuh“
(Adriaen van de Velde) und
„Ziegenmelkerin in der
römischen Campagna“
(Karel Du Jardin)

Dass diese Bemühungen von Erfolg gekrönt waren, wird am besten durch eine Reihe von Kunstwerken dokumentiert, die nach der Publikation im Internet durch das „Lost Art“-Projekt wieder den rechtmäßigen Besitzern zugeführt werden konnten. Beispiele hierfür sind das Gemälde „Landschaft mit Hirten, Pferd und Kuh“ von Adriaen van de Velde (1635-1672), das im Jahr 2001 an die Nachkommen der Betroffenen zurückgegeben werden konnte, sowie die Rückkehr des Gemäldes „Ziegenmelkerin in der römischen Campagna“ von Karel Du Jardin (1622-1678) im Mai 2002 in die Staatlichen Kunstsammlungen Dresden.

Neben dem Anwendungsfeld und der Umsetzung der Web-Schnittstelle bietet das Projekt auch aus wissenschaftlicher Sicht interessante Fragestellungen. Dies resultiert einerseits aus der Natur der Daten, die sehr unterschiedlich sind. Andererseits sind bei der Suche nach Kulturgütern oft nur wenige Informationen bekannt, da zum Beispiel das Wissen über Kunstgegenstände aus Familienbesitz über die Jahre verloren ging. Aus diesen Gründen sind möglichst umfassende und flexible Suchmechanismen notwendig, die dem Nutzer auf einfache und verständliche Weise angeboten werden müssen. So kann zum Beispiel eine Bild-Ähnlichkeitssuche verwendet werden, wenn noch Fotos eines Gemäldes existieren, aber

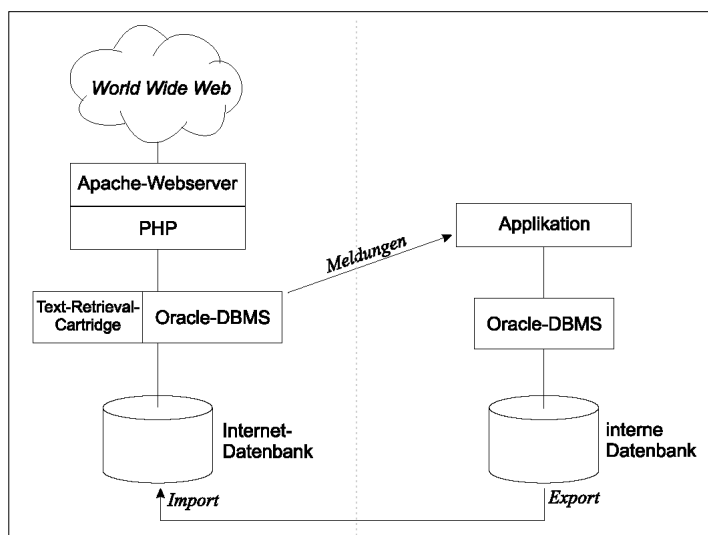
Firewall-gesicherte Umgebung jedoch eine Platzierung der Internet-Datenbank in der Koordinierungsstelle ausschloss, wurde eine replizierte Datenbanklösung gewählt (Abbildung 3).

In der internen Datenbank der Koordinierungsstelle werden alle Daten verwaltet. Ausschließlich hier werden neue Informationen, z. B. zu Verlusten, eingetragen. Damit verbunden sind Operationen, die das Auffinden von redundanten Informationen unterstützen, wobei hier zur Berücksichtigung unterschiedlicher Schreibweisen von Künstlernamen und Werktiteln in Kürze ähnlichkeitsbasierte Vergleiche zum Einsatz kommen sollen.

In der Internet-Datenbank werden nur die öffentlich zugänglichen Informationen verwaltet. Diese Datenbank wird regelmäßig durch Übernahme der Daten aus der internen Datenbank aktualisiert. Beide Datenbanken basieren auf Oracle8i unter Linux (Internet-DB) bzw. Windows NT (interne DB). Die Website www.lostart.de mit Zugang zur Internet-Datenbank ist unter Nutzung des Web-Servers Apache und PHP implementiert. Zur Realisierung eines einheitlichen Layouts der Web-Seiten, das auch eine transparente Umschaltung zwischen mehreren Sprachen unterstützt, wird ein Template-Mechanismus genutzt, der die Programmlogik von der Darstellung der Daten trennt. Diese Methode ermöglicht durch Austauschen von Elementen in den Templates die Anpassung an die gewünschte Sprache. Die Website bietet neben Informationen zum Hintergrund der Datenbank Recherchemöglichkeiten in folgenden Datenbeständen:

- öffentliche Verluste von Museen, Bibliotheken, Sammlungen und Archiven als Folge des Zweiten Weltkrieges („Beutekunst“),
 - private Verluste aus dem gleichen Kontext,
 - insbesondere private Verluste infolge nationalsozialistischer Verfolgung („Raubkunst“),
 - Objekte der so genannten „Linzer Liste“ (entspricht dem so genannten „Restbestand CCP“ für Central Collecting Point München), d. h. Kulturgüter, die insbesondere der jüdischen Bevölkerung durch die Nationalsozialisten mit der Absicht geraubt wurden, in Linz ein „Führermuseum“ aufzubauen,
 - Objekte, die sich u. a. in Besitz öffentlicher Einrichtungen der BRD oder Österreichs befinden und deren Herkunft einen Entzug infolge nationalsozialistischer Verfolgung nicht ausschließt.
- Weiterhin können von den Besuchern der Website sowohl Such- als auch Fundmeldungen mit Hilfe entsprechender Formulare aufgegeben werden. Diese Meldungen werden zunächst in der öffentlichen Datenbank abgelegt und – nach Benachrichtigung – von den Mitarbeitern der Koordinierungsstelle nach Prüfung in den internen Datenbestand eingearbeitet. Erst danach wird die sich daraus ergebende Aktualisierung in der Internet-Datenbank sichtbar.

Der Datenbestand umfasst gegenwärtig summarisch ca. 3,5 Millionen Objekte, von denen mehr als 50 000 Objekte aus über 300 Institutionen bzw. von über 100 Personen mit Detailinfor-



mationen versehen sind. Hinzu kommen über 6 000 Bilder, die sowohl als Vorschaubild als auch im Großformat vorliegen. Täglich werden ca. 10 000 Zugriffe sowie mehr als 500 direkte Rechercheanfragen registriert. Die Zugriffe stammen zu einem großen Teil aus Medien-Unternehmen (z. B. Verlagen oder Zeitungen), Hochschulen und Bibliotheken, aber auch aus Auktionshäusern.

Weiterhin wird ein Forum für Ankündigungen und Diskussionen zum Thema Kulturgutverluste angeboten. Nutzer können hier selbst Beiträge veröffentlichen bzw. auf Anfragen anderer Nutzer antworten.

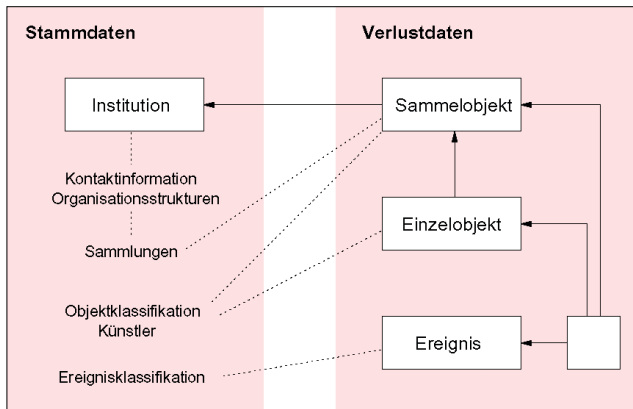
STRUKTUR DER DATENBANK

Wie bereits im vorhergehenden Abschnitt beschrieben, existieren zwei Datenbanken mit separaten Schemata für die interne Verwaltung innerhalb der Koordinierungsstelle sowie für den Zugriff über die Web-Schnittstelle, welche über die im Abschnitt „Synchronisation“ beschriebene Lösung zur Datensynchronisation gekoppelt sind. Beide Schemata sind relativ komplex (ca. 34 bzw. 24 Tabellen), da umfangreiche Informationen zu einzelnen (Einzelobjekte) und zusammengefassten Verlustmeldungen (Sammelobjekte), den verlustmeldenden Institutionen sowie zur Verlustgeschichte verwaltet werden. Das kompaktere Schema der Internet-Datenbank ergibt sich, da einzelne Tabellen für

- eine einfachere Handhabung der Synchronisation,
- bessere Recherchemöglichkeiten sowie
- eine effizientere zukünftige Wartung und Pflege zusammengefasst wurden, wo dies ohne Verletzung grundlegender Anforderungen wie Redundanzfreiheit möglich war. Einige Tabellen wurden nicht übernommen, da sie für die Recherche nicht relevant sind bzw. interne Informationen widerspiegeln.

Im Folgenden sollen nur die Grundzüge des Schemas der Internet-Datenbank dargestellt werden, die aber in dieser Darstellungsform im Wesentlichen mit denen der Datenbank der Koordinierungsstelle übereinstimmen. Die Daten sind unterteilt in Stammdaten und Verlustdaten.

Abb. 3
Architektur der lostart.de-
Datenbank



RECHERCHEMÖGLICHKEITEN

In diesem Abschnitt sollen die Recherchemöglichkeiten in der „Lost Art“-Internet-Datenbank dargestellt werden. Hierbei müssen verschiedene Anwendergruppen beachtet werden: zunächst zufällige Besucher der Internet-Seite, Kunsthändler, Kunsthistoriker sowie auch Betroffene bzw. ihre Erben. Diese Anwender sind zumeist keine Datenbank-Experten und benötigen möglichst intuitive Recherchemöglichkeiten. Weiterhin sind diese Anwendergruppen international zusammengesetzt, d. h. die Website als auch die Recherche müssen in mehreren Sprachen angeboten werden.

Möglichkeiten der Suche

Die Suche bietet verschiedene Möglichkeiten: eine *Textsuche* über die textuellen Beschreibungen der Verlustobjekte, eine *Bildsuche*, die die Abbildungen der Objekte in den Mittelpunkt der Recherche rückt sowie in Kürze eine *Bild-Ähnlichkeitssuche*. Die Textsuche unterteilt sich weiterhin in eine *einfache Textsuche* und eine *detaillierte Textsuche*. Beide Arten der Textsuche nutzen Techniken des Information Retrieval in Texten. Eine Übersicht über die verschiedenen Eingabeformulare der Suchmöglichkeiten gibt Abbildung 5.

Eigenschaften der Textsuche

Jedes der gesuchten oder aufgefundenen Objekte wird durch textuelle Beschreibungen mit verschiedenen Attributen dargestellt. Für die Suche auf diesen Eigenschaften werden unscharfe Anfragen auf Basis von Textähnlichkeiten eingesetzt. Dieses Vorgehen unterstützt den Anwender soweit, dass er auch bei fehlerhaften Schreibweisen relevante Ergebnisse erzielen kann. Die Erstellung einer nach Relevanz geordneten Ergebnismenge ist durch das Ranking von Textähnlichkeiten möglich.

Abb. 4 Vereinfachtes Schema der lostart.de-Datenbank

Stammdaten: Dies umfasst den relativ stabilen und vergleichsweise kleinen Teil des Datenbestands, der z. B. Informationen zu *Institutionen*, zugehörigen Ansprechpartnern und organisatorischen Strukturen enthält. Weiterhin finden sich hier wichtige Informationen zur Klassifizierung von gesuchten oder vorhandenen Objekten sowie von Ereignissen in der Verlustgeschichte.

Verlustdaten: Hier werden die Daten zu den verschollenen oder aufgefundenen Gegenständen verwaltet, was den größeren Anteil des Datenvolumens ausmacht. Einerseits werden Informationen zu *Sammelobjekten* verwaltet, die im Fall von vorliegenden Detailinformationen mit Informationen zu *Einzelobjekten* ergänzt werden. Zu diesen Detailinformationen zählen die teilweise vorhandenen bildlichen Darstellungen der Objekte. Die Verlustgeschichte kann für Sammel- und Einzelobjekte durch zugeordnete *Ereignisse* dargestellt werden.

Diese Grundstruktur des Schemas ist in Abbildung 4 dargestellt.

Abb. 5 Eingabeformulare für die verschiedenen Suchmöglichkeiten

Da die Beschreibung auf mehreren Attributen basiert, existieren die zwei Anfragemodi *einfache Textsuche* (Abbildung 5) und *detaillierte Textsuche* (Abbildung 5). Die einfache Suche erlaubt dabei eine Suche über alle Attribute. Die Suchattribute werden durch ein ODER verknüpft. Dagegen erlaubt die detaillierte Suche die explizite Angabe von Zeichenketten für die jeweiligen Attribute. Die einzelnen Angaben werden konjunktiv verknüpft, wodurch die Suchergebnisse eingeschränkt werden. Eine Kombination von einfacher und detaillierter Suche ermöglicht die Formulierung aller Anfrageformen.

Eine Anfrage an die Datenbank – in diesem Beispiel wird nach „van Gogh“ im deutschen Teil der Datenbank gesucht – lautet für die einfache Textsuche wie folgt:

```
SELECT
  einzelobjekt_id,sammelobjekt_id,
  institution_id,sammlung_id,
  objektart_id,institutiontyp_id
FROM
  verlust_index
WHERE
  www_access='-1'
  AND ( contains(person,'van Gogh',1) > 0
  OR contains (titel_tech,'van Gogh',2) > 0
  OR contains (inst_name,'van Gogh',3) > 0
  OR contains (sammlung_name,'van Gogh',4) > 0 )
```

Hierbei wird eine Volltextsuche auf verschiedenen Attributen der Tabelle `verlust_index` durchgeführt, in der alle für die Suche relevanten Daten gespeichert sind. Dabei greift die `contains`-Funktion auf den Volltextindex zu, der jeweils für die vier angegebenen Attribute und für die entsprechenden Attribute für englische und russische Daten erstellt wurde. Diese Indexe werden wie folgt erstellt:

```
create index EObjekt_Person on
verlust_index(Person)
```

```
indextype is ctxsys.context
parameters ( 'lexer ctxsys.basic_lexer' );
```

Die einzelnen Ranking-Ergebnisse für die verschiedenen Attribute werden standardmäßig vom Oracle-DBMS zusammengefasst und somit ein kombiniertes Ranking erstellt. Allerdings ist hier später der Einsatz von bestimmten, verschiedenen Gewichtungen denkbar. Im Allgemeinen bietet die Volltextsuche einen großen Gewinn für die Nutzbarkeit des Systems. Die Präsentation des Ergebnisses einer Anfrage zeigt die Abbildung 6 links.

Eigenschaften und Anforderungen der Bildsuche

Neben textuellen Daten über gesuchte oder vorhandene Objekte existieren auch Abbildungen bzw. Fotos, mit deren Hilfe die Suche nach verschollenen Gütern für den Anwender erleichtert werden kann. Zur Unterstützung dieser Aufgabe müssen die Abbildungen verwaltet und dem Anwender präsentiert werden. Für die Präsentation der Abbildungen stehen zwei Ansichtsmodi zur Verfügung:

- Ein verkleinertes *Vorschaubild* (Thumbnail) bietet dem Anwender einen schnellen Eindruck über das Objekt. Diese Ansicht zeichnet sich durch eine geringe Auflösung und Qualität sowie geringen Speicherplatzbedarf aus. Diese werden aus den Originalbildern vorberechnet.
- Die *Detailansicht* dagegen präsentiert die Abbildung in einer großen Auflösung und deutlich besserer Qualität. (Die Bildqualität hängt dabei entscheidend von den Ausgangsvorlagen ab, die oft 60 Jahre und älter sind.) Es dient zur genaueren Betrachtung der Kunstwerke.

Das Oracle-DBMS bietet zur Verwaltung der Bilddaten die Möglichkeiten der direkten Speicherung der binären Daten in einem `BLOB`-Feld als auch die Verwaltung von Referenzen auf Verzeichnispfade von Dateien, in denen die Abbildungen als Bilddateien gespeichert sind.

Abbildung 6 zeigt zwei Darstellungen von Suchergebnissen. Links ist eine detaillierte Ansicht von 177 Suchergebnissen dargestellt, die in einer Liste mit sechs Einträgen zu sehen ist. Jeder Eintrag enthält eine Nummer, einen Titel, den Künstler, den Objektart, das Material / die Technik und den Bereich. Rechts ist eine globale Suche dargestellt, die 179 Suchergebnisse in einer Gitteransicht zeigt. Die Gitteransicht ist in zwei Zeilen unterteilt, die jeweils 179 Suchergebnisse zeigen. Die Gitteransicht ist mit 'Globale Suche' und '179 Suchergebnisse' beschriftet.

Abbildung 6
Darstellungen der Suchergebnisse



Abb. 7
Gewähltes Anfragebild und
Eingabeformular für die
Gewichte

Zunächst wurden in der „Lost Art“-Datenbank alle Bilddaten direkt in der Datenbank als BLOB-Felder verwaltet, das hatte vor allem bei der Bereitstellung großer Detailansichten Einbrüche in der Leistung zur Folge. Diese Daten wurden deshalb als Dateien abgelegt, die im Zugriffspfad des Webservers liegen und somit sehr viel schneller geliefert werden können. Die Vorschaubilder verblieben dagegen innerhalb der Datenbank, da sie relativ geringe Speicheranforderungen haben.

Für die Suche über Abbildungen können zwei generelle Ansätze angewandt werden: die direkte Suche auf den Inhalten der Bilder bzw. die Einteilung der Objekte und ihrer Abbildungen in eine Konzepthierarchie. Zunächst wurde der zweite Ansatz gewählt, und die Bilder wurden in eine dreistufige Hierarchie eingeordnet. Die Hierarchie besteht aus den Stufen *Objektart*, *Oberbegriff* und *Unterbegriff*. Die Suchattribute werden über ein Web-Formular eingegeben, wie in Abbildung 5 zu sehen ist.

Das Suchergebnis wird in einer Übersichtsseite (siehe Abbildung 6 rechts) mit allen gefundenen Bildern als Thumbnails dargestellt. Hier kann der Anwender sich einen Überblick über die erfassten Kulturgüter verschaffen. In einem zweiten Schritt ist eine Erweiterung der Suche über Bildeigenschaften denkbar, indem der Benutzer ein für ihn relevantes Beispielbild auswählt und ähnliche Abbildungen zu diesem suchen lässt.

Eigenschaften der Bild-Ähnlichkeitssuche

Bei den bisher beschriebenen Recherchemöglichkeiten über die Textsuche bzw. über eine hierarchisch aufgebaute Bildsuche wird auf manuell erstellte Daten, wie z. B. Künstler, Titel, Objektart und Verluststatus der Kulturgüter, zurückgegriffen.

Ofmals sind die beschreibenden und identifizierenden Daten bzw. die Suchkriterien jedoch aus verschiedenen Gründen entweder fehlerhaft oder unvollständig, so dass eine Datenbankanfrage falsche Ergebnisse liefert. Eine andere Möglichkeit, um insbesondere Kunstwerke zu identifizieren, bietet die Bild-Ähnlichkeitssuche. Damit erschließt sich auch die Möglichkeit der Dublettenerkennung, also die Erkennung, ob unterschiedliche Einträge sich auf das selbe Kulturobjekt beziehen. Die Bild-Ähnlichkeitssuche erlaubt außerdem eine sprachunabhängige Recherche des Datenbestandes.

Bei der Bild-Ähnlichkeitssuche werden zu einem gegebenen Anfragebild alle ähnlichen in der Datenbank befindlichen Objekte gesucht. Um eine Ähnlichkeitssuche zu gewährleisten, werden sowohl für jedes Bild in der Datenbank als auch

für das Anfragebild Features, wie Textur, Farbe und Form, automatisch extrahiert. Diese Features beschreiben den Inhalt eines Bildes. Zur Bestimmung der Ähnlichkeit zwischen zwei Bildern wird eine Ähnlichkeitsfunktion, wie beispielsweise die Euklidische Distanz, verwendet. Der Grad der Ähnlichkeit wird als numerischer Wert im Intervall [0,1] ausgedrückt. Je höher dieser Wert ist, desto höher ist die Ähnlichkeit zwischen den Abbildungen.

In der „Lost Art“-Datenbank wird die Ähnlichkeitssuche unter Zuhilfenahme des Visual-Image-Retrieval-Cartridge von Oracle realisiert. Dies ermöglicht das automatische Generieren einer Signatur für jedes Bild in der Datenbank. Auf Basis dieser Signatur, welche die Featurewerte für die lokale und globale Farbe, Textur und Struktur zusammenfasst, erfolgt die Suche nach ähnlichen Objekten.

Bei der Ähnlichkeitssuche in der „Lost Art“-Datenbank wählt der Anwender ein Anfragebild aus und kann durch die Angabe von Relevanzgewichten für die einzelnen Features seine spezifischen Suchpräferenzen ausdrücken, siehe Abbildung 7. Das Anfragebild kann entweder ein Bild aus dem bestehenden Datenbestand sein oder ein Bild, welches durch den Anwender bereitgestellt wird, indem dieser es von seiner Platte hochlädt. Das System sucht dann auf Basis des vom Anwender gewählten Anfragebildes und der spezifizierten Gewichte nach ähnlichen Abbildungen in der Datenbank.

Das Suchergebnis wird in einer Übersichtsseite (siehe Abbildung 8) mit allen gefundenen Bildern als Thumbnails dargestellt. Dabei sind die Bilder entsprechend ihrer Ähnlichkeit absteigend sortiert. Der Anwender kann sich somit schnell einen Überblick über die vorhandenen Kunstwerke verschaffen, die eine hohe Ähnlichkeit zu seinem gewählten Anfragebild aufweisen. Für dieses Ranking stellt das Visual-Image-Retrieval-Cartridge ebenfalls entsprechende Mechanismen bereit.

Sollte der Anwender nicht mit dem Ergebnis seiner Suche zufrieden sein, ist eine iterative Anfrageverfeinerung (Relevanz-Feedback) denkbar. Dafür markiert der Anwender relevante und nicht relevante Abbildungen in der Ergebnisliste. Auf der Basis dieses Feedbacks werden die Gewichte für die Anfrage neu berechnet und eine neue Suche initiiert. Die aktualisierte Ergebnisliste wird dann dem Anwender präsentiert.

Navigation über den Daten

Der navigierende Zugriff auf die Daten, insbesondere im Bereich der öffentlichen Suchmeldungen, aber auch der öffentlichen Fundmeldungen, stellt einen zweiten Ansatz der Recherche in der „Lost Art“-Datenbank dar. Dieses Zugriffsparadigma richtet sich hauptsächlich an Forscher, allgemein Interessierte und Mitarbeiter öffentlicher Einrichtungen, die so einen schnellen Überblick über die gemeldeten Objekte bekommen.

Abb. 8
Ergebnisübersicht der
Ähnlichkeitssuche



Abbildung 9 zeigt die gewählte Navigationshierarchie. Als Einstiegspunkt dient dazu eine Liste der Staaten, im Falle der BRD im Weiteren der Bundesländer, von der aus es dem Anwender ermöglicht wird, die Hierarchie über die Institutionen, Objektgruppen bis hin zu den Einzelobjekten abzustiegen und detailliertere Informationen zu erhalten. So ist auf der Ebene der Objektgruppen und Einzelobjekten eine grafische Darstellung des Verlustortes von Kulturgütern mittels einer Karte möglich.

Eine Verknüpfung der beiden Rechercheparadigmen – Suche und Navigation – ist durch die Möglichkeit des Aufsteigens in die Navigationshierarchie ausgehend von einem Einzelobjekt, das durch eine der oben genannten Suchmöglichkeiten gefunden wurde, gegeben.

MEHRSPRACHIGKEIT

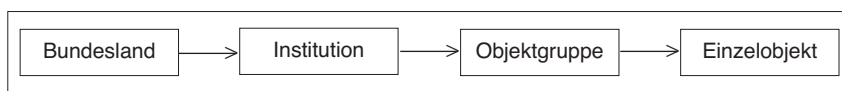
Die Recherche nach kriegsbedingt verbrachten Kunstgütern ist nur durch eine internationale Kooperation erfolgreich durchzuführen, deshalb unterstützt die „Lost Art“-Website die Sprachen Deutsch, Englisch und Russisch. Während die Integration von Deutsch und Englisch technisch gesehen kein großes Problem darstellt, erfordert die kyrillische Schrift den Einsatz eines umfangreicheren Zeichensatzes. Dafür steht der Unicode-Zeichensatz zur Verfügung. Dieser Standard wird sowohl von der Middleware (PHP) als auch durch die Client-Software (HTML-Browser) unterstützt. Allerdings benutzt Oracle den Zeichensatz Unicode nur passiv¹⁾, das heißt, es kann nur abgespeichert werden, aber eine Verwendung für die Volltextsuche ist nicht möglich. Somit müssen die Daten russischer Sprache in einem speziellen kyrillischen Zeichensatz in der Datenbank gespeichert und für die Ausgabe in Unicode umgewandelt werden.

Während des Reengineering des ursprünglichen Datenbankschemas wurden zur Unterstützung der Mehrsprachigkeit zusätzliche Attribute für englische und russische Bezeichnungen für Eigenschaften der Verlustobjekte eingeführt. Diese Art der Unterstützung der Mehrsprachigkeit erfordert allerdings weiteren Aufwand bei der Volltextsuche, da jetzt je nach geforderter Sprache der deutsche, englische bzw. der russische Text durchsucht werden muss, was zu einer Erzeugung und Verwaltung von Volltextindizes für die jeweiligen Sprachen führt.

Die Flexibilität der Anwenderschnittstelle bzgl. der Sprache wird durch eine PHP-Template-Technologie erreicht, die die Programmlogik von der Darstellung der Daten trennt. Diese Methode ermöglicht durch Austauschen von Elementen in den Templates die Anpassung an die gewünschte Sprache. Ein weiteres Problem der Mehrsprachigkeit stellt der Arbeitsaufwand für unseren Kooperationspartner dar, der annähernd alle Informationen in den drei verschiedenen Sprachen erstellen muss.

SYNCHRONISATION

Die Daten der Koordinierungsstelle werden in zwei verschiedenen Datenbanken gehalten – der *Internet-Datenbank* und der *internen Datenbank*.



Die erstere dient der Veröffentlichung im Internet und der initialen Aufnahme von Such- bzw. Fundmeldungen sowie der Änderung von Institutionsdaten, während die zweite innerhalb der Koordinationsstelle Verwendung findet, wobei hier die Daten u. a. auch für die Veröffentlichung eingearbeitet werden. Somit entsteht der Bedarf nach einer Synchronisation der Daten der beiden Systeme. Allerdings ist hier das Zeitkriterium nicht entscheidend, da:

- Eintragungen von Personen bzgl. Such- oder Fundmeldungen oder die Änderungen der Institutionsdaten zunächst auf Plausibilität überprüft und ggf. nachrecherchiert werden müssen, d. h. eine Meldung wird nicht direkt aus der Internet-Datenbank in die interne Datenbank überführt.
- Änderungen in der internen Datenbank nicht immer sofort öffentlich zugänglich sein müssen. Ein weitaus entscheidenderes Kriterium stellt die Datensicherheit dar, da in der internen Datenbank vertrauliche Informationen gespeichert sind. Aus diesen Gründen wurde der Ansatz einer manuellen Übertragung der Daten zwischen den Datenbanken gewählt.

Die Änderungen in der Internet-Datenbank werden in separaten Tabellen gehalten, die eine Kopie zu den zuletzt offiziell freigegebenen Tabellen darstellen. Während des Exports wird ein SQL-Skript generiert, welches alle Änderungen in die interne Datenbank überträgt. Dieses wird nicht automatisch abgearbeitet, sondern ein Mitarbeiter der Koordinierungsstelle muss die einzelnen Änderungen für die Einarbeitung freigeben.

Im umgekehrten Fall, also der Synchronisation von Koordinierungsstelle zur Internet-Datenbank, werden alle neuen Daten zur Veröffentlichung aus der internen Datenbank exportiert, auf einen physischen Datenträger kopiert und in die Internet-Datenbank importiert. Somit wird vor allem die Sicherheit gewahrt, da die interne Datenbank nicht aus dem Internet erreichbar ist.

ZUSAMMENFASSUNG UND AUSBLICK

Die Umsetzung von Web-Datenbanken ist heutzutage gängige Praxis und wird, insbesondere in Standardanwendungsfällen wie Web Shops, Auktionen usw., durch umfangreiche Erfahrungen und Entwicklungswerkzeuge unterstützt. Die dargestellte Datenbank zur Recherche nach verschollenen Kulturgütern bot die Möglichkeit, anhand der vielen außergewöhnlichen Aspekte des Anwendungsszenarios einerseits neue, nutzerorientierte Konzepte für den Datenbankzugriff über eine Web-Schnittstelle zu entwickeln und andererseits Erfahrungen über die Nutzbarkeit von Standardlösungen aus dem Bereich Datenmanagement zu gewinnen.

Abb. 9
Navigationshierarchie für
öffentliche Verluste

1)
In der von uns verwendeten
Version 8i.

Da die in der Datenbank verwalteten Informationen über die Verluste, deren Herkunft und Geschichte relativ komplex strukturiert sind, wurde für den Zugriff eine nutzerfreundliche Mischung aus Information Retrieval, strukturierten Datenbankabfragen, Verzeichnisstrukturen und Navigation umgesetzt. Dies ermöglicht eine schnelle und effiziente Suche mit verschiedenen Paradigmen, die für die Anwender leicht verständlich ist und prinzipiell auf Akzeptanz stieß.

Die auf Grund des Anwendungskontextes notwendige Unterstützung der Mehrsprachigkeit konnte zwar auf Seite der Anwendungsschnittstelle einfach umgesetzt werden, führte jedoch bei der Datenverwaltung zu erheblichem Mehraufwand. Für dieses im Bereich Web-Datenbanken häufig auftretende Problemen wären herstellerseitige Unterstützungen denkbar. Zu Perfor-

mance-Problemen führte auch die Verwendung von Paketen zur Verwaltung von Graphikobjekten in der Datenbank. Eine spezielle Lösung zur Bild-Ähnlichkeitssuche bei Bibliotheksstempeln konnte unter deren Verwendung bisher nicht umgesetzt werden, da für diesen recht spezifischen Anwendungsfall keine sinnvollen Anfrageergebnisse erzeugt wurden.

Neben dem zuletzt genannten Problem der Ähnlichkeitssuche für Bibliotheksstempel, für das eine spezielle Lösung entwickelt werden muss, soll in Zukunft auf der Basis der Standardpakete eine Bild-Ähnlichkeitssuche unterstützt werden. Da diese Anfragen sich in ersten Versuchen bei hohen Zugriffsraten als unpraktikabel berechnungsintensiv erwiesen, wird diesbezüglich und auch hinsichtlich der Gesamtarchitektur nach Optimierungspotenzialen gesucht.



Dr. Kai-Uwe Sattler

ist wissenschaftlicher Assistent in der Arbeitsgruppe Datenbanken der Otto-von-Guericke-Universität Magdeburg. Momentan hat er eine Vertretungsprofessur in Halle sowie zuvor in Dresden und Davis/Kalifornien. Seine Arbeitsgebiete sind Datenbankintegration und -föderation sowie Anfragebearbeitung in heterogenen Datenbanksystemen.



Eike Schallehn

ist seit 1999 wissenschaftlicher Mitarbeiter in der Datenbankgruppe der Fakultät für Informatik der Otto-von-Guericke-Universität Magdeburg. Zuvor schloss er ebenfalls dort sein Studium ab und arbeitete anschließend bei verschiedenen Firmen im Umfeld von Datenbanktechnologien und -anwendungen. Seine aktuellen Forschungsschwerpunkte liegen in den Bereichen Anfragebearbeitung, Datenbankintegration und Digitale Bibliotheken.



Dr. Ingo Schmitt

arbeitet in der Datenbankgruppe der Otto-von-Guericke-Universität Magdeburg. Er promovierte 1998 zum Thema Schemaintegration für den Entwurf föderierter Datenbanken. Im Jahr 1999/2000 war er als Assistant Professor an der University of Windsor in Kanada tätig. Seit 1999 forscht er auf dem Gebiet Multimedia-Datenbanken.



Nadine Schulz

hält einen B. Sc. in Computer Information Systems der University of Wisconsin Stevens Point (USA) und hat ein Diplom der Wirtschaftsinformatik von der Otto-von-Guericke-Universität Magdeburg. Seit Oktober 2000 untersucht sie verschiedene Aspekte zur Gewichtung und Transformation von Ähnlichkeitsanfragen in Multimedia-Datenbanken an der Universität Magdeburg.