



.hist 2006

**Geschichte im Netz:
Praxis, Chancen, Visionen**

ISSN: 1612-5940

Historisches Forum

10 • 2007 • Teilband I

Vom 22. bis zum 24. Februar 2006 veranstaltete der Kooperationsverbund Clio-online in Berlin die Tagung „.hist 2006: Geschichte im Netz – Praxis, Chancen, Visionen“. Dies war bereits der zweite große Kongress an der Humboldt-Universität zu Berlin und der Berlin-Brandenburgischen Akademie der Wissenschaften, der sich mit „Geschichte im Netz“ und mit den Möglichkeiten und Grenzen der Nutzung von Kommunikationsnetzwerken durch die Geschichtswissenschaft in theoretischer Reflexion wie praxisbezogener Diskussion beschäftigte. Der Austausch zwischen Geisteswissenschaftlern/innen, die mit netzbasierten Medien arbeiten, Archivaren/innen, Bibliothekaren/innen, Museumsmitarbeitern/innen und Informatikern/innen zielte auf die Überwindung disziplinärer und institutioneller Grenzen, was sich auch in den vorliegenden Bänden widerspiegelt.

Die Konferenz bot in zwölf Sektionen und Workshops, zwei Podiumsdiskussionen und an zahlreichen Projektständen mit insgesamt etwa hundert Vorträgen, Diskussionsrunden und Präsentationen mehr Informationen, als die zahlreichen Besucher/innen aufnehmen konnten. Umso mehr freuen wir uns, im vorliegenden Doppelband schriftliche Ausarbeitungen der meisten Referate, bereits über H-Soz-u-Kult veröffentlichte Querschnittsberichte zur Tagung sowie eine Zusammenfassung der Abschlussdiskussion vorlegen zu können. Die Teilbände 1 und 2 des Historischen Forums Nr. 10 bündeln mit insgesamt über 80 Beiträgen die Mehrzahl der auf der Tagung „.hist 2006“ präsentierten Vorträge und machen sie allen Interessierten zugänglich.

ISBN: 978-3-86004-205-2

Beiträge der Tagung .hist 2006

Herausgegeben für Clio-online von
Daniel Burckhardt, Rüdiger Hohls und Claudia Prinz



http://edoc.hu-berlin.de/e_histfor/10_I/
Veröffentlichungen von Clio-online, Nr. 2



von Thomas Burch und Andrea Rapp

Nachschlagewerke aller Art verzeichnen, ordnen und bewahren das ständig wachsende Weltwissen. Sie ermöglichen die Orientierung in verschiedenen Fachgebieten und können so die Wissensbildung bei einer in ihrer Ganzheit oft nicht mehr überschaubaren Informationsfülle entscheidend befördern. Digitale Nachschlagewerke sind auf vielfältige Art aufeinander bezogen und damit in gewisser Weise implizit „vernetzt“. Nachschlagewerke, die durch inhaltlich-strukturelles Markup in standardisierte und damit vergleichbar gemachte Informationseinheiten gegliedert und durch Metadaten angereichert sind, machen die impliziten Vernetzungen explizit. Dadurch kann eine neue Qualität der Informationsgewinnung erreicht werden und die Lücke zwischen der schwerfälligen Benutzbarkeit und eingeschränkten Verfügbarkeit der Buchversionen einerseits und der fehlenden Systematik und Beliebigkeit der Information im Internet andererseits geschlossen werden. Der Aufsatz beschreibt Methoden und Verfahren zum Aufbau eines intelligenten Wörterbuchnetzes für deutschsprachige Wörterbücher. Von einem „Meta-Wörterbuch“ der deutschen Sprache, dem ambitionierten Fernziel dieses Vorhabens, könnten alle an Sprachinformationen interessierten Disziplinen profitieren.

Ausgangslage: Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren

Das seit 1998 an der Universität Trier bestehende „Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswis-

senschaften“ hat mehrere Arbeitsschwerpunkte. Dazu gehören nicht nur eigene (Retro-)Digitalisierungsprojekte, sondern auch Serviceleistungen für Projekte und Verlage sowie Programmierarbeiten für interdisziplinäre Projektverbünde.¹ Im Folgenden soll einer der zentralen Arbeitsschwerpunkte des Zentrums ausführlicher vorgestellt werden, nämlich die Retrodigitalisierung und weitere Aufbereitung bzw. Erschließung und Vernetzung von Wörterbüchern.

Wörterbücher als vernetzte Strukturen

Bereits in den ersten Vorschlägen der „Facharbeitsgruppe Inhalt“ zur Vorbereitung des Programms „Retrospektive Digitalisierung von Bibliotheksbeständen“ der Deutschen Forschungsgemeinschaft (DFG) haben Wörterbücher eine dominierende Rolle gespielt, nicht nur weil sie besonders intensiv benutzt werden, sondern auch weil elektronische Nachschlagewerke Auswertungsmöglichkeiten bieten, die in ihrer gedruckten Form kaum jemals ausgeschöpft werden können.²

Nachschlagewerke aller Art verzeichnen, ordnen und bewahren das ständig wachsende Weltwissen. Sie ermöglichen die Orientierung in verschiedenen Fachgebieten und können so die Wissensbildung bei einer in ihrer Ganzheit oft nicht mehr überschaubaren Informationsfülle entscheidend befördern. Schwierigkeiten bereitet jedoch die zunehmende Diversifizierung der Disziplinen und die daraus resultierende Isoliertheit der Einzelinformationen.

1 Die Tätigkeitsfelder reichen von der Organisation der Erfassung bereits vorhandener Grundlagenwerke, von der Anreicherung der Daten mit metasprachlichen Informationen über die Beratung bei der Konzeption von elektronischen Publikationen neuer Werke bis zur Entwicklung und Implementierung netzbasierter Arbeitsumgebungen in dezentralen Arbeitsstellen. Die Zusammenarbeit von Informatik, Informationswissenschaft, philologischer Datenverarbeitung sowie verschiedenen Fachwissenschaften und die daraus resultierende Methoden- und Perspektivenvielfalt kann als Markenzeichen des Trierer Kompetenzzentrums bezeichnet werden: <<http://www.kompetenzzentrum.uni-trier.de>>. Alle im Folgenden angeführten Websites zuletzt eingesehen am 21.2.2006.

2 <http://www.sub.uni-goettingen.de/index_menu.html?http://www.sub.uni-goettingen.de/ebene2/vdf/empfehl.htm>.

Zugleich entspricht ihre Anlage als alphabetisch oder systematisch aufbereitete Informationsquelle einer verbreiteten, von ihnen selbst begründeten Kulturtechnik der Informationssuche, ganz unabhängig von den Inhalten, die durch ihre jeweilige Struktur erschlossen werden. Das macht sie zu zentralen Hilfsmitteln von Forschung und Lehre, denn die von Internet-Suchmaschinen gesammelten Daten können weder die Verlässlichkeit noch den Grad an Systematisierung und Tiefenerschließung der in Nachschlagewerken aufbereiteten Informationen erreichen und bleiben derzeit weit hinter deren Präzision und Umfang zurück.

Digitale Nachschlagewerke sind, wie auch ihre gedruckten Entsprechungen, auf vielfältige Art aufeinander bezogen und damit in gewisser Weise implizit „vernetzt“; eine übergreifende, integrierte Recherche ist jedoch wegen der Unterschiede in der Anlage, Anordnung und Struktur der einzelnen Werke nicht ohne Weiteres möglich.³ Nachschlagewerke, die durch inhaltlich-strukturelles Markup in standardisierte und damit vergleichbar gemachte Informationseinheiten gegliedert und durch Metadaten angereichert sind, machen die impliziten Vernetzungen explizit. Dadurch kann eine neue Qualität der Informationsgewinnung erreicht werden und die Lücke zwischen der schwerfälligen Benutzbarkeit und eingeschränkter Verfügbarkeit der Buchversionen einerseits und der fehlenden Systematik und Beliebigkeit der Information im Internet andererseits geschlossen werden.

Konzepte, die solche Strukturen abbilden, sollten auf der Grundlage einer repräsentativen Auswahl aus den verschiedenen Kategorien von Nachschlagewerken bzw. Wörterbuchtypen entwickelt werden, da diese Typen unterschiedliche und je spezifische wissensorganisierende Strukturen aufweisen. Der Trierer Arbeitsgruppe stehen unterschiedliche Typen tief annotierter Wörterbücher in digitaler Form zur Verfügung, die bereits in unterschiedlicher Dichte untereinander vernetzt, durch Symptomwertangaben und Metadaten intellektuell erschlossen und recherchierbar oder

3 Diese unterschiedlichen Strukturen erklären sich zum einen wissenschaftshistorisch aus der Entwicklung der lexikografisch-lexikologischen Methoden, zum anderen aus den spezifischen Zielsetzungen der verschiedenen Wörterbuchtypen.

auch mit Primärquellen vernetzt sind:

- Typ Sprachstadienwörterbuch: Mittelhochdeutsches Wörterbuch, Mittelhochdeutsches Handwörterbuch (einschlägige Nachträge), Findebuch zum mittelhochdeutschen Wortschatz (als Verbund mit multidirektionalen Verweisen)⁴, Neues Mittelhochdeutsches Wörterbuch einschließlich digitalem Quellen- und Belegarchiv⁵
- Typ sprachstadienübergreifendes Wörterbuch: Deutsches Wörterbuch von Jacob und Wilhelm Grimm⁶ (DWB)
- Typ Dialektwörterbuch: Pfälzisches Wörterbuch, Rheinisches Wörterbuch, Wörterbuch der elsässischen Mundarten, Wörterbuch der

4 Lexner, Matthias, Mittelhochdeutsches Handwörterbuch. 3 Bde. Nachdruck der Ausgabe Leipzig 1872-1878. Mit einer Einleitung von Kurt Gärtner, Stuttgart 1992; Mittelhochdeutsches Wörterbuch. Mit Benutzung des Nachlasses von Georg Friedrich Benecke ausgearbeitet von Wilhelm Müller und Friedrich Zarncke. 4 Bde. Nachdruck der Ausgabe Leipzig 1854-1866. Mit einem Vorwort und einem zusammengefaßten Quellenverzeichnis von Eberhard Nellmann, Stuttgart 1990 (=BMZ); Gärtner, Kurt u.a., Datenverarbeitung Hanrieder, Gerhard, Findebuch zum mittelhochdeutschen Wortschatz. Mit einem rückläufigen Index, Stuttgart 1992; Burch, Thomas u.a. (Hgg.), Mittelhochdeutsche Wörterbücher im Verbund. CD-ROM und Begleitbuch, Stuttgart 2002. CD-ROM-Version auf dem CD-ROM-Server der UB Trier; Internetversion unter <<http://www.mwv.uni-trier.de>>.

5 Gärtner, Kurt u.a. (Hgg.), Mittelhochdeutsches Wörterbuch. Erster Band, Doppellieferung 1/2, Lieferung 1: a-amurschaft bearbeitet in der Arbeitsstelle der Akademie der Wissenschaften und der Literatur Mainz an der Universität Trier von Ralf Plate und Jingning Tao, Lieferung 2: an-balsieren bearbeitet von der Arbeitsstelle der Akademie der Wissenschaften zu Göttingen von Susanne Baumgarte, Gerhard Diehl und Bernhard Schnell. Mit einer CD-ROM, Stuttgart 2006 (vgl. <<http://www.mhdwb.uni-trier.de>>); Digitales Mittelhochdeutsches Textarchiv: <<http://www.mhgt.uni-trier.de>>.

6 Deutsches Wörterbuch von Jacob und Wilhelm Grimm. Nachdr. der Erstbearbeitung München 1984; Der Digitale Grimm. Deutsches Wörterbuch von Jacob und Wilhelm Grimm. Elektronische Ausgabe der Erstbearbeitung, bearbeitet von Hans-Werner Bartz, Thomas Burch, Ruth Christmann, Kurt Gärtner, Vera Hildenbrandt, Thomas Schares, Klaudia Wegge. Herausgegeben vom Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier in Verbindung mit der Berlin-Brandenburgischen Akademie der Wissenschaften. 2 CD-ROMs, Benutzerhandbuch, Begleitbuch. 1. Aufl., Frankfurt a.M. 2004; vgl. <<http://www.dwb.uni-trier.de>>.

- deutsch-lothringischen Mundarten, Luxemburgische Wörterbücher⁷
- Typ Autorenwörterbuch: Goethe-Wörterbuch⁸ (GWB)
- Typ Sachwörterbuch: Ökonomische Enzyklopädie von J.G. Krünitz (Projekt der UB Trier)⁹

Ziel der Überlegungen ist, anhand der verschiedenen Wörterbücher zum ‚Deutschen‘ einen standardisierten Einstieg zu schaffen, von dem aus sich die nicht standardisierten, heterogen organisierten und verstreuten Informationen verzweigen – heterogen im Hinblick auf synchrone wie diachrone Varianz. Dieses Konzept lässt sich nicht nur auf Wörterbücher anwenden, sondern auch auf die Erschließung großer Korpora übertragen. Das skizzierte Ziel soll an einigen Beispielen verdeutlicht werden.

7 Fournier, Johannes, Vorüberlegungen zum Aufbau eines Verbundes von Dialektwörterbüchern, in: Zeitschrift für Dialektologie und Linguistik 70 (2003), S. 155-176; Wörterbuch der elsässischen Mundarten. Bearbeitet von Ernst Martin und Hans Lienhart. 2 Bände, Strassburg 1899-1907. [Nachdruck Berlin/New York 1974]; Wörterbuch der deutsch-lothringischen Mundarten. Bearbeitet von Ferdinand Follmann, Leipzig 1909. [Nachdruck Hildesheim/New York 1971]; Rheinisches Wörterbuch. Im Auftrag der Preußischen Akademie der Wissenschaften, der Gesellschaft für Rheinische Geschichtskunde und des Provinzialverbandes der Rheinprovinz auf Grund der von Johannes Franck begonnenen, von allen Kreisen des Rheinischen Volkes unterstützten Sammlung bearbeitet und herausgegeben von Josef Müller, Heinrich Dittmaier, Rudolf Schützeichel und Mattias Zender. 9 Bände, Bonn/Berlin 1928-1971; Pfälzisches Wörterbuch. Begründet von Ernst Christmann, fortgeführt von Julius Krämer, bearbeitet von Rudolf Post unter Mitarbeit von Josef Schwing und Sigrind Bingenheimer. 6 Bände, Wiesbaden/Stuttgart 1965-1997; Gangler, J. F., Lexicon der Luxemburger Umgangssprache, Luxemburg 1847. [Unveränderter Nachdruck, Vaduz 2002]; Wörterbuch der luxemburgischen Mundart, Luxemburg 1906; Luxemburger Wörterbuch. Bd. 1-4 und Nachtragsband, Luxemburg 1950 – 1977; <<http://www.dvw.uni-trier.de>>.

8 Goethe-Wörterbuch. Hg. v. der Berlin-Brandenburgischen Akademie der Wissenschaften [bis Bd. 1, 6. Lfg.: Deutsche Akademie der Wissenschaften zu Berlin; bis Bd. 3, 4. Lfg.: Akademie der Wissenschaften der DDR], der Akademie der Wissenschaften in Göttingen und der Heidelberger Akademie der Wissenschaften. Bd. 1 (A-azurn), 1978; Bd. 2 (B-einweisen), 1989; Bd. 3 (einwenden-Gesäusel), 1989; Bd. 4, Lieferung IV/1-10 (Geschäft-hinzutreten); <<http://www.gwb.uni-trier.de>>.

9 Oeconomische Encyclopädie oder allgemeines System der Land-, Haus- und Staats-Wirtschaft : in alphabetischer Ordnung. Bd. 1 - 242. Berlin : Pauli, 1773-1858; <<http://www.kruenitz.uni-trier.de>>.

Konkrete Vernetzungen

Im Gegensatz zu einer bloßen, unverbundenen Bereitstellung verschiedener Nachschlagewerke bieten Wörterbuchverbände mit multidirektionalen Verlinkungen komplexe und gezielt spezifizierbare Zugänge zum Material. Diese Verlinkungen gehen weit über eine rein ausdrucksseitige Verknüpfung hinaus, indem sie philologische und informationswissenschaftliche Methoden verbinden. Die Entwicklung von Verfahren, Methoden und Technologien, die eine intelligente, dynamische Vernetzung von Nachschlagewerken unterschiedlichen Typs einschließlich ihrer Quellenverzeichnisse und Primärquellen ermöglichen und dadurch eine neue Qualität des Informations- und Wissensmanagements erreichen, erfordert neben informationstechnologischen Konzepten auch eine stärkere Einbeziehung philologischer und lexikografischer Verfahrensweisen.

Neben der Digitalisierung der Basiskomponenten wird ein Mehrwert durch zusätzliche Inhalte in Form von Meta-Daten geschaffen: Einfachstes Beispiel ist die Abbildung explizit genannter Verweise in andere Wörterbücher auf zunächst monodirektionale Hyperlinks, die jedoch mit Hilfe von automatischen Verfahren auch multidirektional erzeugt werden können. Eine weitere Möglichkeit ist beispielsweise die Auswertung von Sprachangaben im Innern des Artikels, zum Beispiel im Pfälzischen Wörterbuch die Abbildung der Angaben ‚*mhd. Lemma*‘ auf die mittelhochdeutschen Wörterbücher.

Die Stichwörter verschiedener Nachschlagewerke lassen sich beispielsweise über eine ‚Hyper-Lemmaliste‘ zusammenführen, um einen integrierten Zugang zu den unterschiedlichen Lemma-Ansätzen (zum Beispiel dialektal versus hochsprachlich, sprachstufenbezogen versus neuhochdeutsch) zu schaffen. Die Kategorisierungen von Quellen nach so genannten Symptomwerten bieten weitere Möglichkeiten (Raum, Zeit, Textsorte, Stil etc.). Auch die inhaltliche Klassifizierung von Nachschlagewerken und Enzyklopädien (zum Beispiel durch die DDC oder ähnliches)¹⁰ ist ein solcher Mehrwert, der neue Erschließungs- und damit Suchmöglichkeiten schafft.

¹⁰ DDC = Dewey Decimal Classification <<http://www.oclc.org/dewey/>>.

Im Falle der ‚Mittelhochdeutschen Wörterbücher im Verbund‘ beispielsweise wurde die multidirektionale Vernetzung anhand einer standardisierten ‚Hyper-Lemmaliste‘ aller Stichwörter aus den vier eingebundenen Werken realisiert. Hierzu mussten etwa die Stichwortansätze eines der Wörterbücher, das für Verben jeweils die erste Person Singular Indikativ Präsens ansetzt, mit den infinitivischen Lemmaansätzen der übrigen mittelhochdeutschen Wörterbücher verbunden werden.¹¹ Eine rein ausdrucksseitige, alphabetische Kumulierung der jeweiligen Lemmata hätte im Falle der Mittelhochdeutschen Wörterbücher also nicht zu einem brauchbaren Ergebnis geführt. Ferner galt es, Homographen voneinander zu unterscheiden, um auf das semantisch entsprechende Lemma verweisen zu können. Dies wurde anhand halbautomatischer Verfahren realisiert, indem Konkordanzen mit automatisch erzeugten Zuordnungsvorschlägen von philologisch geschulten Bearbeitern/innen geprüft wurden. Bei den automatischen Zuordnungen können zum Beispiel auch grammatische Angaben aus den Wörterbuchartikeln mit herangezogen werden, was die Trefferquote erhöht.¹²

Die Quellenverzeichnisse der Mittelhochdeutschen Wörterbücher wurden darüber hinaus mit Metadaten angereichert, die die Primärquellen kategorisieren, so dass eine spezifizierte Abfrage nach den Kategorien Raum, Zeit und Textsorte möglich ist. Zusammen mit der differenzierten inhaltlich-strukturellen Annotierung der Artikel können im Mittelhoch-

¹¹ Die Stichwortansetzung bei Verben nach der Form der 1. Person Singular Präsens Indikativ ist auch für lateinische Wörterbücher üblich. Die strikt alphabetische Anordnung aller Stichwörter gilt im ‚Mittelhochdeutschen Wörterbuch‘ (BMZ) nur für diejenigen Wörter, die die Wortstämme repräsentieren, das sind in der Regel Simplizia. Unter dem Stichwort für den Wortstamm sind dann alle Glieder der Wortfamilie systematisch (nicht alphabetisch) aufgeführt. Die über eine Hyper-Lemmaliste erfolgte Zuordnung aller Lemmata zu der strikt alphabetischen Anordnung der anderen Komponenten des Verbundes (Lexer, Findebuch) war eine erhebliche Herausforderung, die im Projekt zu bewältigen war.

¹² Thomas Burch, Johannes Fournier, Lexikografische Information per Mausclick: Die wichtigsten Wörterbücher zum Mittelhochdeutschen auf einer CD-ROM, in: ZfdA 130 (2001), S. 306-318; dies., Zur Anwendung der TEI-Richtlinien auf die Retrodigitalisierung Mittelhochdeutscher Wörterbücher, in: Lemberg, Ingrid u.a. (Hgg.), Probleme und Perspektiven computergestützter Lexikografie, Tübingen 2001, S. 133-153.

deutschen Wörterbuchverbund komplexe Recherchen wie beispielsweise die nach allen Ableitungen mit dem Suffix *-unge* in religiösen Texten des 13. Jahrhunderts aus dem Bairischen Raum durchgeführt werden.

Mit der digitalen Erstbearbeitung des „Deutschen Wörterbuchs von Jacob und Wilhelm Grimm“ (DWB) steht ein weiteres, sprachstadienübergreifendes Wörterbuch zur Verfügung, das die zu schaffende Hyperlemmaliste des Deutschen um das Frühneuhochdeutsche und Neuhochdeutsche erweitern wird. Mit seinem reichen Sprachschatz kann das DWB ein zentraler Bestandteil eines Informationsnetzes zur deutschen Sprache sein, von dem aus zahlreiche weitere Verzweigungen, vor allem auch zu den in den Belegzitate repräsentierten Primärquellen, realisierbar sind.

Eine Vernetzung mit den für die deutsche Sprache wichtigen Autorenwörterbüchern liegt also auf der Hand. Zum Individualwortschatz Goethes, des im DWB am häufigsten zitierten Autors, wird seit 1946 ein vollständiges Wörterbuch erstellt, dessen elektronische Aufbereitung derzeit mit DFG-Förderung in Trier erfolgt und das mit dem DWB verknüpft wird. Als Typus des ‚Autorenwörterbuchs‘ wird es eine Verbundkomponente bilden, die beispielsweise die Möglichkeit bietet, über die Belegzitate eine Verfeinerung der Auszeichnung des DWB vorzunehmen und damit weitere Recherchen zu spezifizieren. Auch das Goethewörterbuch selbst wird durch das Einbringen von Metadaten weiter erschlossen: Im GWB werden Belege nach Band und Seite der Sophienausgabe zitiert, nicht nach Werken Goethes. Ideal wäre sicherlich eine direkte und multidirektionale Verknüpfung mit dieser Ausgabe selbst, die jedoch derzeit nicht in einer frei zugänglichen digitalen Version zur Verfügung steht. Als Behelf bietet sich folgendes Verfahren an: Verknüpft man eine Auswertung der Sophienausgabe, die die Seitenangaben mit Werktiteln hinterlegt, mit dem Wörterbuch, lassen sich werkbezogene Recherchen im Goethewörterbuch durchführen. Auch eine Klassifizierung dieser Auswertung nach Textsorten bzw. Gattungen ist geplant, so dass zum Beispiel Recherchen nach Adjektivbildungen in der Lyrik Goethes möglich werden.

Im Falle des derzeit im Aufbau befindlichen „Digitalen Verbundes von Dialektwörterbüchern“ muss neben der genauen Feststellung der Etymologie und Bedeutung der Stichwörter auch eine Auswertung des

Inhalts der einzelnen Artikel erfolgen, um am hochdeutschen Standard orientierte („Pfälzisches Wörterbuch“, „Rheinisches Wörterbuch“) und dialektale Lemmaansätze („Wörterbuch der elsässischen Mundarten“, „Wörterbuch der deutsch-lothringischen Mundarten“, „Luxemburgische Wörterbücher“) korrekt aufeinander beziehen zu können. Überdies müssen die Dialektwörterbücher auch über ein Sachstichwort aufeinander beziehbar gemacht werden, um die reiche Synonymie und Heteronymie der Dialekte recherchierbar zu machen. Damit werden Vorarbeiten nicht nur für eine semasiologisch, sondern auch für eine onomasiologisch orientierte Verbindung von einzelnen Informationspositionen geleistet, die auch die automatische Erstellung von Umkehrwörterbüchern und eine entsprechende Nutzung der angewandten Methoden vorantreiben können.¹³

Bei den Dialektwörterbüchern ließe sich darüber hinaus anhand der Belegorte eine punktgenaue Verortung recherchieren, aber auch kartografisch visualisieren bzw. eine Vernetzung mit Sprachatlanten erreichen¹⁴, so dass die großräumigeren Angaben der Sprachstadienwörterbücher durch kleinräumige Kontrastierungen ergänzt werden könnten.

Eine andere Art der semantischen Erschließung wird von der Universitätsbibliothek Trier bei der Digitalisierung der 242 Bände umfassenden „Ökonomischen Enzyklopädie“ von Johann Georg Krünitz angewendet. Die Klassifizierung der Artikel mit Hilfe der Dewey Decimal Classification (DDC) ermöglicht auch international standardisierbare inhaltliche Recherchen.¹⁵ Als Perspektive ist eine Verbindung der onomasiologischen

13 Eine besondere Schwierigkeit bei der semantischen Auswertung stellen Phraseologismen dar, die sowohl für historische Sprachstufen als auch für Dialekte nur unzureichend erforscht sind. In Trier werden derzeit zwei Datenbanken zur historischen (Dr. Natalia Filatkina, Prof. Dr. Claudine Moulin) und zur dialektalen Phraseologie (Kerstin Knop, M.A.) erstellt und ausgewertet. Die Ergebnisse dieser Arbeiten fließen in die Erschließung der Wörterbuchbibliothek ein.

14 Die Erprobung einer solchen Vernetzung anhand des Verbunds von Dialektwörterbüchern und dem Digitalen Wenkeratlas ist in Arbeit <<http://www.diwa.info>>; vgl. dazu auch Fournier (wie Anm. 7), S. 174.

15 Darüber hinaus stellt dieses Sachwörterbuch eine hervorragende Ausgangsbasis für eine Einbeziehung weiterer Enzyklopädien und Nachschlagewerke dar, indem sie etwa das „Grammatisch-kritische Wörterbuch“ von Johann Christoph Adelung (<<http://mdz.bib-bvb.de/digbib/lexika/adelung>>) oder die „Systema Naturae“ Carl

Erschließung mit der Sacherschließung gut vorstellbar, denn gerade solche Arten der Vernetzung ermöglichen die Überwindung von Disziplinergrenzen.

Das Prinzip der Interoperabilität ermöglicht außerdem die Verknüpfung mit externen Ressourcen. Neben der Vernetzung der durch das Kompetenzzentrum bereitgestellten Wörterbuchressourcen bietet das Wörterbuch-Netz auch die Möglichkeit, aus externen elektronischen Dokumenten angesprochen zu werden. Über die vor den Wortartikeln eingerichteten Icons kann zum einen die persistente Adresse des betreffenden Artikels abgefragt werden, zum anderen stellt das Kompetenzzentrum auf Anfrage auch die vollständigen Lemmalisten mit den zugehörigen Adressen zur Verfügung, um auf diese Weise eine umfassende Vernetzung zwischen einer externen Wörterbuchinstallation und den Ressourcen im Wörterbuch-Netz realisieren zu können. So existieren bereits Verknüpfungen aus dem *Deutschen Rechtswörterbuch*¹⁶ und dem *Wörterbuch der Deutschen Winzersprache*¹⁷ in das Wörterbuchnetz, die demonstrieren, wie Vernetzungen dezentral realisiert werden können.

Schließlich ergibt sich im Kontext einer vernetzten ‚Wörterbuchbibliothek‘ die Möglichkeit der weiterführenden Vernetzung der Wörterbuchbelege mit Primärquellen. Eine solche Verknüpfung von großen Textkorpora und Wörterbüchern bietet eine neue Qualität wissenschaftlichen Arbeitens, da die Vernetzung in beide Richtungen erfolgen kann, von den Wörterbüchern zu den Quellen und umgekehrt. Elektronische Texteditionen, aber auch alle anderen Formen von Primärquellen (Audio, Video, Karten, Musiknoten, Diagramme etc.) können damit auf sehr einfache Weise mit Lemmata oder einzelnen Textstrecken verknüpft werden. Dies macht je nach Maßgabe des Auszeichnungsgrades kritische Apparate unmittelbar transparent und überprüfbar. Modellhaft wurde dies bereits

von Linnés (<http://www-gdz.sub.uni-goettingen.de/cgi-bin/digbib.cgi?PPN371257700>) zitiert, so dass die digitalen Versionen dieser beiden und weiterer Quellen mit der Krünitzschen Enzyklopädie verlinkt werden können.

16 <http://www.deutsches-rechtswörterbuch.de>; Akademie der Wissenschaften Heidelberg.

17 <http://www.winzersprache.de/onlinewb>; Akademie der Wissenschaften und der Literatur Mainz.

durch die Verknüpfung des ‚Findebuchs zum mittelhochdeutschen Wortschatz‘ und des ‚Mittelhochdeutschen Textarchivs‘ realisiert.

Auswertungsbeispiel

Ein konkretes Beispiel soll im Folgenden das Vernetzungs- und Auswertungspotenzial der entsprechend annotierten Nachschlagewerke und Quellen veranschaulichen.¹⁸

Bezeichnungen für Realien, insbesondere für Früchte oder landwirtschaftliche Erzeugnisse, sind zum einen für die Sprachwissenschaft, zum anderen aber auch für kulturwissenschaftliche wie naturwissenschaftliche Disziplinen besonders interessant, da diese Bereiche zumeist eine reiche und häufig hochspezialisierte regional gegliederte Synonymik aufweisen. Die ‚Brombeere‘ (*Rubus fruticosus*) als eher selten domestizierte typische Wildfrucht bietet ein vergleichsweise gut überschaubares Beispiel. Neben dem durchsichtigen Grundwort *-beere* geht das heute nicht mehr verständliche Bestimmungswort *Brom-* vermutlich auf die indogermanische Wurzel **bher* für ‚etwas Spitzes, Kantiges, Eckiges‘ zurück.

Eine semasiologische Gliederung und Auswertung anhand der vorhandenen Wörterbücher zeigt, dass ‚Brombeere‘ regional gegliedert durchaus verschiedene Realien bezeichnet (Brombeere, Brombeerstrauch, dornige Ranke, Heidelbeere, Preiselbeere etc.) und auch im übertragenen Sinn gebraucht werden kann (Gewitterwolke, schwarzhaariges Mädchen etc.). Eine onomasiologische Gliederung des Materials fördert ferner die reiche Synonymik des Begriffs mit unterschiedlichen semantischen Konnotationen zu Tage (Schmerzbeere, Heckenbeere, Bachbeere, Schwarzbeere, Pferdsbeere, Hundsbeere, More etc.). Als Motivation für die Benennung der Brombeere beschreiben die Wörterbücher verschiedene Aspekte, zu denen Eigenschaften, Aussehen und Verwendung der Pflanze gehören können.

Eine sachbezogene Erschließung macht deutlich, dass die volksspra-

18 Die Auswertungsmöglichkeiten können hier nur kurz angerissen werden, vgl. daher ausführlicher unter <http://www.dvw.uni-trier.de/onomasiologie/start.html>.

chigen Bezeichnungen für die Beerenfrüchte biologisch nicht eindeutig sind, das heißt, dass Bezeichnungen wie ‚Heidelbeere, Himbeere, Preiselbeere, Brombeere‘ nach ihrer biologisch korrekten Bezeichnung disambiguiert werden müssen. Die Klassifikation der Artikel des Krünitz gemäß DDC ermöglicht weitergehende Fragestellungen und Vergleichsmöglichkeiten, zum Beispiel ob für Pflanzen der Klassifikation ‚Rosales‘ (im Krünitz unter anderem die Lemmata *Eberäsche*, *Erd=Beere*, *Geiß=Bart*, *Odermennige*) bei ähnlichem Aussehen und Eigenschaften ähnliche Benennungen verwendet werden. Abgesehen von dieser eng an die im Wörterbuch angelegte Thematik eröffnet die DDC-Klassifikation vielfältige Anknüpfungspunkte; für ‚Brombeere‘ sind dies neben den schon genannten ‚Rosales‘ beispielsweise die Themenkreise „Ökologie“, „Schäden, Krankheiten, Schädlinge an Pflanzen“, „Kochen“, „Wein“, „Fütterung (Viehzucht)“, „Pharmakologie“ etc. Nicht nur die Sprachwissenschaft, sondern auch die Wissenschaftsgeschichte bis hin zur modernen Botanik oder Biodiversitätsforschung kann also von einer solch differenzierten Aufschlüsselung entsprechender Wortfelder profitieren.

Die Vorgehensweise für die Realisierung solcher Konzepte kann folgendermaßen aussehen: Die Mikrostruktur der Wörterbuchartikel wird ausgezeichnet, zum Beispiel nach den Informationseinheiten ‚Stichwortansatz (abgebildet auf ein Hyper-Lemma)‘, ‚Etymologie‘, ‚Bedeutungsangabe‘ und ‚Ortsangabe‘, zudem erfolgt eine Annotierung über eine standardisierte Klassifizierung (DDC, Ontologien).

Die Kombination der zusätzlich eingebrachten Metadaten ermöglicht dann spezifizierte Recherchen: Eine semasiologische Recherche funktioniert ausdrucksseitig und greift auf den Wortlaut der Stichwörter zurück. Durch die Abbildung auf das Hyper-Lemma führt die Recherche auch bei grafischer bzw. diachroner Varianz zu etymologisch gleichen Stichwörtern (zum Beispiel mhd. *brâm-ber*, nhd. *Brombeere*). Dabei werden alle Artikel zur Zeichenfolge ‚Brombeere‘ gefunden, auch wenn zum Beispiel semantisch die ‚Maulbeere‘ dahinter steht. Eine onomasiologische Recherche geht semantisch-begrifflich vor, es wird der Begriff, die „Sache“ Brombeere gefunden, auch wenn sie mit *Schwolzbel*, *Mouze* oder *Schmierbat* bezeichnet wird. Eine sachbezogene Recherche findet auch die Artikel, in

denen es im weiteren bzw. impliziten Sinne um die ‚Brombeere‘ geht, auch dann, wenn in dem entsprechenden Artikel die Brombeere gar nicht selbst namentlich, sondern nur implizit – zum Beispiel durch die Nennung der biologischen Klasse der ‚Rosales‘ – genannt ist.

Methoden und Verfahren zur rechnerunterstützten Vernetzung

Aus der Sicht der Informatik stellt sich die Frage, mit welchen Methoden sich der Aufbau eines Wörterbuchnetzes algorithmisch unterstützen lässt. Eine erste Ansatzmöglichkeit, die im Folgenden näher vorgestellt werden soll, bietet hier der Bereich des Information Retrieval. In einem ersten Schritt wird von der Realität abstrahiert und ein geeignetes Modell für ein Wörterbuchnetz erstellt, welches dann als Ausgangspunkt zur Implementierung der Algorithmen dient. Ganz allgemein entspricht einem Wörterbuchnetz dabei die informationstheoretische Datenstruktur eines aus Knoten und Kanten bestehenden Graphen.¹⁹ Die Knoten des Graphen repräsentieren zunächst die Wörterbücher und die Kanten beschreiben die Verbindungen, das heißt die Verweise zwischen den Wörterbüchern. Diese grobe Struktur lässt sich weiter verfeinern und genauer modellieren, indem nicht das gesamte Wörterbuch als Knoten aufgefasst wird, sondern die einzelnen Wortartikel. Die Kanten verlaufen dann zwischen den Artikeln verschiedener Wörterbücher bzw. auch innerhalb eines einzelnen Wörterbuchs. Die Kanten sind gerichtet, das heißt Ausgangs- und Zielpunkt sind eindeutig bestimmt.

Ausgehend von den gedruckten Wörterbüchern lassen sich die Verweise zunächst in drei Klassen einteilen: 1. Verweise, die explizit im Wörterbuch genannt sind, beispielsweise durch Angabe einer Seiten-, Spalten-, Zeilen- oder Artikelreferenz; 2. Verweise, die sich aufgrund von statistischen Berechnungen der Wörter in den Artikeln identifizieren lassen; 3. Verweise, die sich durch Graphalgorithmen aus den Klassen 1 und 2 berechnen lassen, beispielsweise durch transitiven Abschluss. Mit

¹⁹ Vgl. Mehlhorn, Kurt, *Data Structures and Algorithms 2: Graph-Algorithms and NP-Completeness*, in: *EATCS Monographs on Theoretical Computer Science*, Berlin 1984.

den im Folgenden vorgestellten Verfahren sollen zunächst Verweise der Klasse 2 bestimmt werden.

Ausgangsbasis zur Implementierung der Algorithmen bildet ein Datenbanksystem (siehe Abbildung 1), in welchem die Wörterbücher in strukturierter Form abgespeichert sind. Die Architektur des Gesamtsystems basiert auf einer standardisierten SGML/XML-Kodierung²⁰ der Wörterbuchdaten, die sich nach unterschiedlichen Document-Type-Definitionen (DTDs) richten kann. Zu jeder der vorgegebenen Kodierungen existiert ein zugehöriger Importfilter, über den die Wörterbuchdaten in ein Datenbankmanagementsystem (DBMS)²¹ übernommen werden können. In umgekehrter Richtung wird ein Exportfilter entwickelt, über den die durch die Bearbeitung innerhalb des Systems neu eingefügten Informationen in Form von zusätzlichen Metadaten zu den grundlegenden Wörterbuchdaten ebenfalls in standardisiertem XML-Format aus dem Datenbanksystem herausgezogen werden können. Auf diese Weise stehen sowohl die Wörterbuchdaten als auch die Beziehungen zwischen den Wörterbüchern in einer plattformunabhängigen und damit langfristig nutzbaren Form zur Verfügung.

Kernstück des gesamten Systems bildet ein relationales Datenbankmanagementsystem, in dem jeweils ein Wörterbuch durch eine zugehörige Datenbank verwaltet wird. Der Aufbau der einzelnen Datenbanken richtet sich dabei nach der Granularität der Auszeichnung. Als kleinste gemeinsame Schnittmenge wird nur festgelegt, dass separate Tabellen zur Abfrage der Stichwörter existieren. Weitere Konstituenten der Wörterbuchartikel wie beispielsweise Angaben zur Wortart, Bedeutungserläuterungen, Belege, Angaben zur Etymologie usw. werden dann in eigenen Datentabellen verwaltet, wenn sie durch die XML-Auszeichnung innerhalb des Gesamtartikels markiert wurden. Je feiner die Auszeichnung ist, desto differenzierter sind die Suchmöglichkeiten und desto genauer die Suchergebnisse.²²

20 Vgl. Goldfarb, Charles, The SGML Handbook, Oxford 1990.

21 Vgl. Rob, Peter; Coronel, Carlos, Database Systems. Design, Implementation and Management, in: The Wadsworth Series in Management Information Systems, Belmont, California 1993.

22 Bei Wörterbüchern, bei denen die Artikelmikrostrukturen kaum differenziert

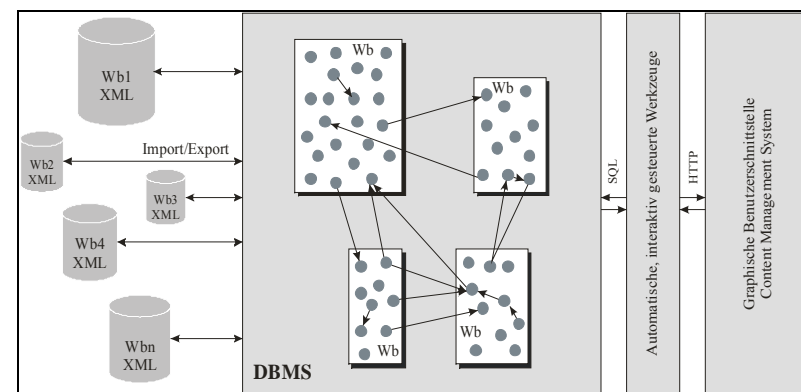


Abbildung 1: Architektur des Datenbanksystems zur Speicherung der Wörterbuchdaten

Ausgehend von den in den Datenbanken abgespeicherten Wörterbuchartikeln können nun automatische Verfahren entwickelt werden, die zunächst zur Aufgabe haben, sämtliche Wortformen der Wörterbücher zu gewichten. Diese Gewichtung erfolgt durch quantitative Angaben über die Häufigkeit einer Wortform (=Term) innerhalb eines Wortartikels und innerhalb des gesamten Wörterbuchs. Berechnet werden die so genannte relative Häufigkeit und die inverse Dokumentfrequenz für jeden Term. Aus beiden Werten ergibt sich dann durch Multiplikation das so genannte Termgewicht. Ein Beispiel für diese Berechnungsvorlage zeigt Abbildung 2.

werden, bliebe im ungünstigsten Fall nur der Zugriff über eine herkömmliche Volltextrecherche.

TextID	Wort	Normierte Form	ArtikelID	Häufigkeit der Wortform	Wortanzahl im Artikel	Dokumenten-häufigkeit	Inverse Dokumentenfrequenz	Gewicht
1507749	butterblume	butterblume	GB13619	1	31	21	14.791	-58.486
1507750	f	f	GB13619	1	31	93291	2.674	-10.572
1507751	nnl	nnl	GB13619	1	31	3629	7.280	-26.788
1507752	boterbloem,	boterbloem	GB13619	1	31	1	19.183	-75.854
1507753	gilt	gilt	GB13619	1	31	2896	7.683	-30.361
1507754	für	fuer	GB13619	4	31	40359	3.882	-10.220
1507755	mehreere	mehere	GB13619	1	31	1408	8.724	-34.495
1507756	kräuter,	kraeuter	GB13619	1	31	467	10.316	-40.791
1507757	für	fuer	GB13619	4	31	40359	3.882	-10.220
1507758	caltha	caltha	GB13619	1	31	31	14.229	-56.264
1507759	palustris,	palustris	GB13619	1	31	163	11.834	-46.795
1507760	dotterblume,	dotterblume	GB13619	1	31	10	15.861	-62.718
1507761	für	fuer	GB13619	4	31	40359	3.882	-10.220
1507762	leontodon	leontodon	GB13619	1	31	42	13.791	-54.531
1507763	taraxacum,	taraxacum	GB13619	1	31	53	13.455	-53.204
1507764	für	fuer	GB13619	4	31	40359	3.882	-10.220
1507765	ranunculus	ranunculus	GB13619	1	31	126	12.183	-48.174
1507766	aunicomus	aunicomus	GB13619	1	31	8	16.183	-63.991
1507767	das	das	GB13619	1	31	94491	2.817	-11.137
1507768	volk	volk	GB13619	1	31	4258	7.127	-28.182
1507769	wähnt,	waeht	GB13619	1	31	74	12.974	-51.300
1507770	davon,	davon	GB13619	1	31	5680	8.711	-26.538
1507771	dass	dass	GB13619	1	31	35412	4.071	-16.096
1507772	sie	sie	GB13619	1	31	45488	3.710	-14.870
1507773	die	die	GB13619	2	31	123292	2.271	-7.653
1507774	kuehe	kuehe	GB13619	1	31	374	10.636	-42.058
1507775	fressen,	fressen	GB13619	1	31	967	8.266	-36.636

Abbildung 2: Berechnung der inversen Dokumentenfrequenz für einen Wortartikel aus dem Deutschen Wörterbuch

Die Tabelle zeigt die durch den Algorithmus berechneten Werte für den Artikel *Butterblume* aus dem Deutschen Wörterbuch von Jacob und Wilhelm Grimm. Entscheidend zur Berechnung des Gewichtes (letzte Spalte) sind die Angaben über die Häufigkeit der Wortform, die Länge des Artikels (gemessen in Wortformen) sowie die Dokumentenhäufigkeit, die die Anzahl der Dokumente (=Wortartikel) repräsentiert, in denen der betreffende Term auftritt. Jedes Dokument wird dabei nur einmal gezählt, auch wenn der Term in ihm mehrfach vorkommt. Aus diesen Angaben lässt sich das Gewicht eines Terms bestimmen: Der Bezug zwischen Häufigkeit der Wortform und Artikellänge ergibt seine relative Häufigkeit im betrachteten Wortartikel; der Bezug zwischen Dokumentenhäufigkeit und der Gesamtanzahl an Dokumenten²³ liefert die inverse Dokumentenfrequenz; das Produkt aus beiden Werten bildet das Gewicht des Terms im vorliegenden

23 Beim Deutschen Wörterbuch sind dies 297.613 Dokumente.

Artikel. Sortiert man nun die Terme eines jeden Artikels nach den berechneten Gewichten, so erhält man einen sogenannten Relevanzvektor der Terme. Für den betrachteten Beispielartikel ist diese Sortierung in Abbildung 3 gezeigt.

TextID	Wort	Normierte Form	ArtikelID	Häufigkeit der Wortform	Wortanzahl im Artikel	Dokumenten-häufigkeit	Inverse Dokumentenfrequenz	Gewicht
1507752	boterbloem,	boterbloem	GB13619	1	31	1	19.183	-75.854
1507766	aunicomus,	aunicomus	GB13619	1	31	8	16.183	-63.991
1507760	dotterblume,	dotterblume	GB13619	1	31	10	15.861	-62.718
1507749	butterblume	butterblume	GB13619	1	31	21	14.791	-58.486
1507758	caltha	caltha	GB13619	1	31	31	14.229	-56.264
1507762	leontodon	leontodon	GB13619	1	31	42	13.791	-54.531
1507763	taraxacum,	taraxacum	GB13619	1	31	53	13.455	-53.204
1507764	für	fuer	GB13619	4	31	40359	3.882	-10.220
1507765	ranunculus	ranunculus	GB13619	1	31	126	12.183	-48.174
1507766	aunicomus	aunicomus	GB13619	1	31	8	16.183	-63.991
1507767	das	das	GB13619	1	31	94491	2.817	-11.137
1507768	volk	volk	GB13619	1	31	4258	7.127	-28.182
1507774	kühe	kuehe	GB13619	1	31	374	10.636	-42.058
1507756	kräuter,	kraeuter	GB13619	1	31	467	10.316	-40.791
1507779	gelb,	gelb	GB13619	1	31	522	10.155	-40.156
1507776	butter	butter	GB13619	1	31	666	8.804	-36.766
1507775	fressen,	fressen	GB13619	1	31	967	8.266	-36.636
1507755	mehreere	mehere	GB13619	1	31	1408	8.724	-34.495
1507753	gilt	gilt	GB13619	1	31	2896	7.683	-30.361
1507751	nnl	nnl	GB13619	1	31	3629	7.280	-26.788
1507768	volk	volk	GB13619	1	31	4258	7.127	-28.182
1507776	verde	verde	GB13619	1	31	5146	6.854	-27.101
1507770	davon,	davon	GB13619	1	31	5680	8.711	-26.538
1507771	dass	dass	GB13619	1	31	35412	4.071	-16.096
1507772	sie	sie	GB13619	1	31	45488	3.710	-14.870
1507767	das	das	GB13619	1	31	94491	2.817	-11.137
1507750	f	f	GB13619	1	31	93291	2.674	-10.572
1507754	für	fuer	GB13619	4	31	40359	3.882	-10.220
1507757	für	fuer	GB13619	4	31	40359	3.882	-10.220

Abbildung 3: Sortierung der Wortformen nach den berechneten Termgewichten (=Relevanzvektor)

Für jeden Wortartikel eines Wörterbuchs wird der Relevanzvektor nach einer festen Maximalanzahl von Elementen abgebrochen.²⁴ Auf diesem Weg entsteht also für jedes in das Netz zu integrierende Wörterbuch eine Liste von Relevanzvektoren.

Im darauffolgenden Schritt werden nun je zwei dieser Vektorlisten elementweise miteinander verglichen, das heißt pro Vektorvergleich werden die in ihnen enthaltenen Wortformen gegeneinander geprüft und die An-

24 Auf diese Weise werden die für die Vergleiche irrelevanten hochfrequenten Wortformen ausgeschlossen, da sie sich durch sehr niedrige Gewichtung auszeichnen.

zahl der Übereinstimmungen ermittelt. Hier wird zunächst auf exakte Gleichheit getestet, womit eine Basis für die möglichen Verweise zwischen den Wortartikeln geschaffen wird. In weiteren Verfeinerungen dieses Algorithmusschrittes könnten auch Informationen über die Wortformen berücksichtigt werden (zum Beispiel durch Lemmatisierung, Stemming, approximative Vergleiche etc.), um zusätzliche Übereinstimmungen zu ermitteln, die dann auch entsprechend bewertet werden können. Einen Vergleichsvektor zu obigem Beispiel zeigt Abbildung 4. Hierbei handelt es sich um den Artikel *Gackel(s)-blume* aus dem Pfälzischen Wörterbuch.

TextID	Wort	Normierte Form	ArtikelID	Häufigkeit der Wortform	Wortanzahl im Artikel	Dokumentenhäufigkeit	Inverse Dokumentfrequenz	Gewicht
1117999	Gackelblumm	gackelblumm	PG00045	1	43	1	17,251	-76,359
1117918	Gackelblümm	gackelblümm	PG00045	1	43	1	17,251	-76,359
1117927	Gackelblümm	gackelblümm	PG00045	1	43	1	17,251	-76,359
1117930	Kuhschelle	kuhschelle	PG00045	1	43	3	15,666	-69,343
1117936	Gackel(s)-e	gackel(e)s	PG00045	1	43	3	15,666	-69,343
1117993	Gackel(s)-blume	gackel(s)blume	PG00045	1	43	4	15,251	-67,506
1117903	Gackel1	gackel1	PG00045	1	43	4	15,251	-67,506
1117901	Godramst	godramst	PG00045	1	43	5	14,929	-66,081
1117911	[Ranunculus]	ranunculus	PG00045	1	43	7	14,444	-63,933
1117897	[Caltha]	caltha	PG00045	1	43	8	14,251	-63,080
1117910	Hahnenfuß	hahnenfuss	PG00045	1	43	8	14,081	-62,328
1117928	putzstail	putzstail	PG00045	1	43	9	14,081	-62,328
1117924	Küchenschelle	kuechenschelle	PG00045	1	43	10	13,929	-61,855
1117914	Butterblume	butterblume	PG00045	1	43	11	13,792	-61,048
1117896	palustris1	palustris	PG00045	1	43	12	13,666	-60,491
1117917	Akelei	akelei	PG00045	1	43	16	13,251	-58,854
1117922	Glockenblume	glockenblume	PG00045	1	43	25	12,807	-55,904
1117925	[Anemone]	anemone	PG00045	1	43	27	12,486	-55,312
1117934	1049	1049	PG00045	1	43	31	12,297	-54,430
1117896	Dotterblume	dotterblume	PG00045	2	43	10	13,929	-61,855
1117906	Dotterblume	dotterblume	PG00045	2	43	10	13,929	-61,855
1117928	[LA-Rohrb]	larohrb	PG00045	1	43	80	11,344	-50,213
1117900	[LA-Wey]	lawey	PG00045	1	43	170	9,842	-43,563
1117912	[Gemm]	gemmh	PG00045	1	43	334	8,868	-39,250
1117919	[GH-Neubg]	ghneubg	PG00045	1	43	564	8,112	-35,905
1117905	c	c	PG00045	1	43	998	7,298	-32,280
1117916	2	2	PG00045	1	43	9211	4,248	-18,802

Abbildung 4: Relevanzvektor zum Artikel *Gackel(s)-blume* aus dem Pfälzischen Wörterbuch

Vergleicht man die normierten Wortformen der Vektoren aus Abbildung 3 und Abbildung 4 miteinander, so findet man drei exakte Übereinstimmungen in den Formen *caltha*, *ranunculus* und *butterblume* innerhalb der ersten 20 gewichteten Terme. Über einen bestimmten Schwellenwert wird für den Gesamtalgorithmus vorgegeben, wann ein Vergleich positiv zu

bewerten ist, das heißt wann ein Verweis zwischen den verglichenen Wortartikeln etabliert werden soll. In obigem Beispiel bedeutet die Anzahl von drei Übereinstimmungen, dass eine Kante im Wortartikelgraphen vom Artikel *BUTTERBLUME* (Deutsches Wörterbuch) zum Artikel *Gackel(s)blume* (Pfälzisches Wörterbuch) eingerichtet werden soll.

Jeder derart ermittelte Verweis wird in einer zusätzlichen Datenbank als Metainformationen über den eigentlichen Wörterbuchdaten abgespeichert, die Basisdaten bleiben also unverändert. Beim Zugriff auf ein Wörterbuch kann aus dieser Datenbank die Vernetzungsinformation ausgelesen und ebenfalls in der grafischen Benutzeroberfläche visualisiert werden (siehe Abbildung 5).

The screenshot shows the online interface of the German Dictionary (Deutsches Wörterbuch) for the entry 'BUTTERBLUME'. The main text area displays the definition: 'BUTTERBLUME, f. nsl. boterbloem, gilt für mehrere kräuter, für caltha palustris, dotterblume, für leontodon taraxacum, für ranunculus auricomus. das voik walde, davon, dass sie die kühe freasen, werde die butter gelb.' To the right, a 'Anpassung / Arbeitsetzung' panel lists related terms with their respective parts of speech and frequency indicators, such as 'BACHBLUME, f.', 'DRAHTBLUME, f.', 'KÜHBLUME, f.', 'Dotterblume', 'Wiesenpappel', 'Weibelenkrut', 'ANENBLUME, f.', 'DOTTERBLUME, f.', 'KASEBLUME, f.', 'MASLIEBE, f.', 'MATTEBLUME, f.', 'MOOSBLUME, f.', 'SCHMALZBLUME', 'SCHMALZKOLLE, f.', 'SCHMERBEL', 'SCHMERBLUME, f.', 'WALDHÄHNLEN, n.', 'WALDSCHMERGEL, m.', 'Butterblume', 'Butterblume', 'Wasserpappel', 'Butterblume', 'Butterblumegäsch'. The bottom of the interface shows search controls and the date 'Besucht seit 15.09.2004 15:07'.

Abbildung 5: Visualisierung der Vernetzungsinformation zur Artikelstrecke *BUTTERBLUME* im Deutschen Wörterbuch von Jacob und Wilhelm Grimm

Abbildung 5 zeigt die Vernetzung des Artikels *BUTTERBLUME* im Deutschen Wörterbuch mit weiteren Artikeln des DWB (zum Beispiel *BACH-*

BLUME, *DRAHTBLUME*, *KUHBLUME*, *KÄSEBLUME* etc.), aber auch mit Artikeln aus anderen Wörterbüchern (zum Beispiel *Dotterblume*, *Schmalzblume* im Pfälzischen Wörterbuch, *Weibelenkrut* im Elsässischen Wörterbuch oder *Butterblume* im Rheinischen Wörterbuch). Die Symbole vor den Verweisgruppen in Abbildung 5 kennzeichnen die Zahl der Übereinstimmungen während der Vektorenvergleiche, höhere Anzahl von Übereinstimmungen führt zu einer Höhergruppierung des Verweises und bedeutet in der Regel eine höhere Zuverlässigkeit der Verknüpfung.

Perspektiven

Das hier skizzierte ambitionierte „Ideal-Ziel“, ein vielfältig verknüpftes, intelligentes ‚Meta-Wörterbuch‘ der ‚deutschen Sprache‘, steht noch in seinen Anfängen und ist gewiss nur in einer gemeinsamen, „koordinierten“ Forschungs- und Umsetzungs-Anstrengung realisierbar. Die dafür notwendigen Konzepte, Methoden und Verfahren müssen in einem interdisziplinären Methodenbündel aus den folgenden Bereichen zusammenwirken:

- Informatik (Ontologien, Graphalgorithmen, Information Retrieval etc.),
- Computerphilologie/Computerlinguistik (automatisches Markup, Pattern Matching, Konkordanzen, automatische Lemmatisierung etc.),
- Systematische Linguistik (Abbildung von Lautgesetzen synchron und diachron, Erarbeitung von Hyper-Lemmalisten, Phraseologie etc.),
- Lexikografie/Lexikologie (Semasiologie, Onomasiologie, Umkehrlexikografie).

Von einem solchen Meta-Zugriff können alle an Sprachinformationen interessierten Disziplinen in höchstem Maße profitieren, darüber hinaus sind weitreichende Impulse für neue Fragestellungen und Lösungsansätze auf verschiedensten Forschungsfeldern zu erwarten.

Dr. Thomas Burch ist seit 1998 Geschäftsführer des Kompetenzzentrums für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier. Die Konzeption und Organisation von Projekten im Bereich des elektronischen Publizierens nach neuesten informationstechnologischen Standards bildet den Schwerpunkt seiner Tätigkeiten. E-Mail: burch@uni-trier.de

Dr. Andrea Rapp war bis September 2004 Leiterin des Göttinger Digitalisierungszentrums an der Niedersächsischen Staats- und Universitätsbibliothek Göttingen; sie ist derzeit Akademische Rätin im Fach Germanistik/Ältere deutsche Philologie und Geschäftsführerin des Kompetenzzentrums für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier. E-Mail: rappand@uni-trier.de

VOM NACHSCHLAGEWERK ZUM VIRTUELLEN
INFORMATIONSSYSTEM –
PERSPEKTIVEN HISTORISCHER LEXIKOGRAFIE IM INTERNET

von Robert Charlier

Der Beitrag widmet sich den Perspektiven historischer Wörterbücher im weltweiten Wissensnetz aus der Sicht des Wörterbuchmachers. Nach einem Überblick über die traditionellen Techniken enzyklopädischer und lexikografischer Wissensordnung und einem Strukturvergleich zwischen historischer und virtueller „Wörterbuchwelt“ rückt das Goethe-Wörterbuch als wichtigstes Klassikerwörterbuch deutscher Sprache in den Mittelpunkt. Der Einblick in die logisch-semantische Datenstrukturierung dieses paradigmatischen historischen Autorenwörterbuchs ermöglicht einen kurzen Ausblick auf aktuelle Entwicklungstendenzen lexikografischer Infrastrukturen im Internet. Ein vorläufiges Fazit resümiert Vernetzung, Visualisierung und Miniaturisierung als wichtigste Tendenzen im Prozess der virtuellen Optimierung von Wörterbüchern.

Von der Systematik zur Alphabetik

Als bedeutendster Autor und repräsentativer Sprecher seines Zeitalters verwendete Johann Wolfgang von Goethe (1749-1832) den Begriff „Diktionär“ als Eindeutschung des französischen Maskulinums „dictionnaire“. Das eingedeutschte Neutrum gebrauchte er zugleich im Sinne von Enzyklopädie, (Real-)Lexikon und von Wörterbuch (Dictionarium, Thesaurus).¹ Erscheint das Wort zunächst unscheinbar, so markiert es eine historische

Wende in der Evolution der abendländischen Wissensordnung. Im Titel von Diderot und d’Alemberts Jahrhundertwerk *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers* (1751-80) kündigt die Vorstellung eines umfassend durchdachten, komplexen und universalen ‚Wörterbuchs‘ von der wohl einschneidendsten Zäsur in der Geschichte der Sammlung von Menschheitswissen seit der monumentalen *Naturalis Historia* des Plinius Secundus des Älteren (23-79 n. Chr.). Es handelt sich dabei um die Ablösung des Prinzips der *hierarchischen* Ordnung von Inhalten nach systematischen Kriterien, die von der Antike bis zur Frühen Neuzeit galten. Zwar ist die alphabetische Anordnung bereits Bestandteil des kulturellen Vermächtnisses der Antike, und auch das Mittelalter hat eine Fülle alphabetischer Texte (wie zum Beispiel Glossare) hervorgebracht. Doch die Enzyklopädisten und ihre Epigonen ersetzten die thematisch sortierte, jeweils mythisch, theologisch, politisch oder naturkundlich auf- oder absteigende Ordnung der Wörter und Dinge in nie zuvor dagewesener Konsequenz und Universalität durch eine Lemmatisierung der Schlag- oder Stichwörter von A bis Z.² Es handelte sich dabei vermutlich um einen Akt der Übertragung sprachlich arbiträrer Ordnungskriterien, wie sie sich in der Abfolge der Buchstaben des lateinischen Alphabets manifestiert, auf die neue Erfahrungswelt, die sich um die Mitte des 18. Jahrhunderts rasant ausdifferenzierte. Die Rede ist von einer exponentiell wachsenden Welt der exotischen Waren, modernisierten Transportmittel, technischen Erfindungen und neuen Entdeckungen, vor allem im transatlantischen und pazifischen Raum. Die Modernisierungsschübe in Wissenschaft, Wirtschaft und Technik brachten nicht nur eine Fülle von Waren und Dingen, zum Beispiel aus der außereuropäischen Welt, sondern auch eine Flut von Wörtern und Begriffen aus der Abstraktionshöhe neuer Theorien und Terminologien. Dieser Wandel der Wissensstrukturierung hatte wesentlichen Anteil an der Wirkmächtigkeit der *Encyclopédie* als

¹ Vgl. Goethe-Wörterbuch, hg. von der Berlin-Brandenburgischen Akademie der Wissenschaften, der Akademie der Wissenschaften in Göttingen und der Heidelberger Akademie der Wissenschaften. Bd. 2, Stuttgart 1989, Sp. 1205, 24-42.

² Die formalen und sachlichen Ordnungskriterien haben sich innerhalb der Realenzyklopädie bzw. der (Fremd-)Sprachenlexikografie je unterschiedlich entwickelt (vgl. Hass-Zumkehr, Ulrike, Die Sachlexikografie und ihr Verhältnis zur Sprachlexikografie, in: Dies., Deutsche Wörterbücher – Brennpunkt von Sprach- und Kulturgeschichte. Berlin u.a.O. 2001, S. 291-332, bes. S. 297-327).

Instrument der europäischen Aufklärung, die durch eine singuläre Säkularisierung, Rationalisierung und Relativierung des Weltwissens und seiner Repräsentation in die Geschichte eingehen sollte. Was zunächst aus der Notwendigkeit der Bewältigung neuer, ungeahnter Komplexität geboren war – die Instrumentalisierung des unter sachlichen Aspekten weitgehend willkürlichen Alphabets als Ordnungskriterium für die Abfolge innerhalb eines universalen Lexikons – gewann mit der produktiven Rezeption der *Encyclopédie* an Eigendynamik. So relativierte die enthierarchisierte, allein von der alphabetischen Buchstabenabfolge bestimmte Zusammenschau der Inhalte im historisch wohl bedeutendsten Universallexikon der Frühmoderne bestehende Hierarchien in nie dagewesener Weise. So gelangten beispielsweise religionskritische Artikel in die unmittelbare Nachbarschaft zu zentralen Leit- und Wesensbegriffen des christlichen Abendlandes oder der katholischen Kirche.

Historische und virtuelle Wörterbuchwelt

Substituiert man für die ältere Wissensordnung die Kategorie der semantischen Systematik (zum Beispiel die Darstellung von Sachgebieten als Aufstieg von der unbelebten über die organische Natur und die Tiere bis zu den Menschen)³ und für die Lemmatisierung den Aspekt der logischen Serialität (durch alphabetische oder numerische Darbietung, etwa in einer Folge von Datensätzen), so finden sich beide Wissensordnungen in der Welt der Datenbanken und Online-Nachschlagewerke verschränkt. Der Bereich von Daten-Ablage/Speicher (memory) ist dabei logisch (alphabetisch, numerisch und ähnliches) aufgebaut, wogegen die Ebene der Abfrage/Suche (retrieval) nach semantischer bzw. mehr oder weniger

³ Eines der ältesten deutschen Wörterbücher ist der *Vocabularius Sancti Galli* in der Stiftsbibliothek des Klosters St. Gallen. Es handelt sich dabei um eine gut 200 Seiten starke Miniaturhandschrift eines lateinisch-althochdeutschen Wörterverzeichnisses. Das Wörterbüchlein ist nicht alphabetisch organisiert, sondern nach sinnvollen Sachgruppen sortiert. Beginnend mit der Welt der Pflanzen und Gewässer behandelt es den Menschen und dessen Kulturbereiche. Darauf folgen Wortfelder aus den Themenbereichen Tierwelt, Ackerbau, ländliche und dörfliche Kultivation usw. Den Abschluss bildet die Benennung des Himmels, der Wetterphänomene und der Jahreszeiten (siehe Abbildung 1).

systematisierter Vernetzung verlangt. Logisch-semantische Abfrage- und Suchmöglichkeiten zählen gegenwärtig zu den unverzichtbaren Errungenschaften eines ‚durchdigitalisierten‘ Alltags.

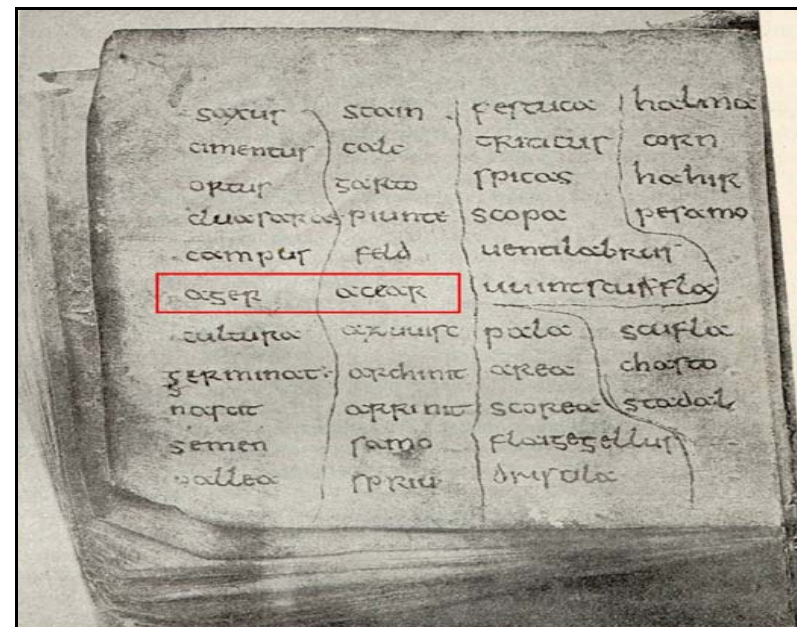


Abbildung 1: *Vocabularius Sancti Galli*, um 790 (althochdt. Systematik lat. Wörter, Beispielwort: *ager*)

Lokale Online-Hilfen für unterschiedlichste Standard-Software, die mehr oder weniger streng nach stichwortbezogenen Kriterien wörterbuchartig bzw. quasi enzyklopädisch aufgebaut sind, bilden dafür nur die naheliegendsten Beispiele. Diese in fast jedem Anwender- und Nutzerbereich der digitalen Medien usuell und ubiquitär gewordenen lexikografischen Infrastrukturen haben den mit der Schriftkultur altüberkommenen Akt der aktiven Wissensaneignung, das „Nachschlagen“, im Prozess des „Durchsuchens“ kraft einer virtuellen Datenabfrage technisch rationalisiert und – semantisch betrachtet – säkularisiert. Ein Informationsvorgang im Internet

oder Digitalmedium, der auf menschliche Empfänger/innen abzielt, basiert auf dem Zusammenspiel eines technisch-physikalischen, logischen Datentransfers mit einer semantischen Schnittstelle. Damit ein Datentransport auch in eine Extraktion von Wissen mündet, also zu einer Information von oder für jemanden werden kann, benötigt diese Schnittstelle einen natürlichsprachlichen Ausgang. Diese anthropomorphe Seite der Schnittstelle zwischen den objektiven Daten und der subjektiven Information aktualisiert zwangsläufig die Ordnungsmuster der althergebrachten (Wörter-)Buch- und Wissenswelt. Damit ist der herkömmliche epistemische Strukturbereich gemeint, der bildhaft auch als „Gutenberg-Galaxis“ firmiert.⁴ So spiegeln die Treffereinträge aus einer Datenabfrage nicht nur Zahl und Inhalt der Felder der zugrunde liegenden Datenbank, sondern auch die semantischen Module eines virtuellen Wörterbucheintrages (Stich-, Schlag- oder Suchwort = Lemma; Fließ- oder Ergebnistext = Belegzitat usw.). Die Akkumulation von Daten im weltweiten Wissensnetz folgt damit mehr oder weniger konsequent den historischen Wissenssystemen, wie sie seit Jahrhunderten durch Institutionen und Techniken wie Archiv, Bibliothek, Index, Kartei, Katalog, Register oder Thesaurus vorgegeben sind.⁵

Die natürlichsprachliche Seite des wichtigsten semantischen Kommunikationsvorgangs in der technischen Medienwelt zum Zwecke der informationellen Mehrwertschöpfung – die Information eines Menschen als Netz- oder Mediennutzer/in – gewinnt ständig an Bedeutung. Im Rahmen des vorliegenden Beitrages kann unmöglich die Problematik semantischer Schnittstellen für (retro)digitale Wörterbücher in ihrer gesamten Breite und

4 Vgl. McLuhan, Marshall, *The Gutenberg Galaxy*, Toronto 1962 [Die Gutenberg-Galaxis. Das Ende des Buchzeitalters, Düsseldorf u.a.O. 1968]. Die Thesen McLuhans über das Ende des Buchzeitalters und die elektronische Vernetzung der Weltgesellschaft zu einem „globalen Dorf“ haben zwar viele Aspekte des Internetzeitalters vorausgesehen. Der sich abzeichnenden technischen und semantischen Synthese der Buch- und Internetwelt wurden sie aber nicht gerecht.

5 Vgl. Charlier, Robert, Synergie und Konvergenz. Tradition und Zukunft historischer Semantik am Beispiel des Goethe-Wörterbuchs, in: Müller, Ernst (Hg.), *Begriffsgeschichte im Umbruch?* (Archiv für Begriffsgeschichte, Sonderheft 2004), Hamburg 2005, S. 167-183, bes. S. 175-177.

Tiefe ausgeleuchtet werden. Ich beschränke mich daher im Folgenden auf einige wenige exemplarische Aspekte.

Logisch-semantische Datenstrukturierung

Die elektronische Redaktion des *Goethe-Wörterbuchs* basiert auf der Standard-Software Microsoft (MS) Word. Die SGML/XML-kompatible Datenauszeichnung erfolgt unmittelbar bei der Texteingabe für die Artikel-erstellung. Dabei dienen die standardmäßig vorgegebenen Zeichen- und Absatz-Formatvorlagen als Parameter bzw. Attributräger. Dieses kombinierte Mark-up folgt etwas vereinfacht im Wesentlichen den Kriterien einer Document Type Definition (DTD) des Typs ‚Wörterbuch‘ (*dictionary.dtd*) nach den Richtlinien der Text Encoding Initiative (TEI) und garantiert die logische und semantische Strukturierung⁶ der Text- und Sinnelemente eines Wörterbuchartikels (Textdaten). Die Standard-Software MS Word fungiert durch makrobasierte Modifikationen der Entwicklungsumgebung von MS Office, Visual Basic for Applications, quasi als *Textdatenverarbeitung*. Zwar bleibt die Funktionsmächtigkeit eines klassischen Spezialprogramms der geisteswissenschaftlichen Informationsverarbeitung wie TUSTEP damit unerreicht, dennoch lassen sich für die Erfordernisse der elektronischen Redaktion eines historischen Wörterbuchs erstaunliche Rationalisierungen erzielen.⁷

Die logisch-semantische Datenauszeichnung beschreibt genotypisch, was als Phänotyp auf der sichtbaren Oberfläche der Software-Anwendung vom virtuellen Wörterbuch erscheint. Die Tiefe der Auszeichnung entscheidet über die Breite der Information, die man über ein Wort, einen Begriff oder eine Sache abfragen kann. Die grafische Erscheinungsform rekuriert dabei auf den zugrundeliegenden Datenschatz, eine Textdaten-

6 Vgl. dazu etwa Münz, Stefan; Nefzger, Wolfgang, Logische und physische Auszeichnung, in: Dies., *HTML Referenz: Grundlagen, Alternativen und Erweiterungen*. Feldkirchen 1997, S. 81 f. und 18* f. (Systematische Kurzreferenz) sowie im Folgenden öfter ebd.

7 Vgl. Charlier, Robert, *Goethe-Lexikographie im Internet – Aspekte prospektiver Digitalisierung*, in: Reeg, Gottfried; Schubert, Martin J. (Hgg.), *Edieren in der elektronischen Ära (Informationstechnologie und Geisteswissenschaften)*, 1), Berlin 2004, S. 17-37, bes. S. 22-35.

struktur, die wie ein *Palimpsest* aufgebaut ist. Je nach Auszeichnungstiefe enthält ein strukturierter Text mehrere Textdatenschichten, so etwa die Nutzeransicht im Dokumentenformat (*what you see is what you get*), die versionsunabhängige Fassung im Rich Text Format (RTF) oder die ansichtsunabhängige, gleichsam ‚grammatische‘ SGML- bzw. XML-Fassung im reinen ASCII-Format (*what you mark is what you get*). Die Berührung von Informationstechnologie und Philologie führt allenthalben zu bildhaften Übertragungen, das belegen zum Beispiel die Metaphern von *Buch*, *Buchrolle* („Scrolling“) oder *Kartei/Karte* im Hypertextbereich. Durch die digitale Revolution sind traditionelle Konzepte des kulturellen Gedächtnisses aber auch auf ihre ursprüngliche Bedeutung zurückgeführt worden. So erweisen sich moderne Bibliotheken aufgrund ihrer Anstrengungen um die elektronische Erschließung von Büchern und Texten in gewisser Weise als digitale *Skriptorien*. Dies ganz im Sinne einer älteren Bedeutung von (Kloster-)Bibliothek, die sich vor allem auch als Ort der Herstellung und Vervielfältigung von (Hand-)Schriften verstand.⁸

Grafische Oberflächen

Die visuelle Anmutung und Struktur der Benutzeroberfläche eines Online-Wörterbuchs ist zwar datenlogisch eher der Peripherie zuzuordnen. Im Zeitalter des Bildes und des Primats der visuellen Kommunikation ist sie jedoch von kaum zu unterschätzender Bedeutung. Daher sei im Folgenden eine Typologie der Erscheinungsbilder virtueller Wörterbücher entworfen. Zu unterscheiden wäre zunächst der Wörterbuchtypus mit windows-kompatibler Vollanzeige, basierend auf einer komplexen Schachtelung von Fenstern und Spalten, wie sie unter anderen das Trierer Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften optimiert hat (1). Ein weiterer grafischer Oberflächentyp stellt die integrierte Teil- oder Kompaktanzeige einer (externen) Wörterbuchabfrage als Toolbar oder Add-In dar (2). Neben der Lösung

8 Vgl. Charlier, Robert; Habermehl, Peter, Die Bibliothek von Babel. Impressionen einer Kölner Tagung, in: Circular der Berlin-Brandenburgischen Akademie der Wissenschaften 8/29 (2004), S. 36f. sowie <<http://www.goethe-wortschatz.de/index.php?c=Tagungen#ad4>> (06.09.2006).

von Google-Toolbar bietet das Leipziger *Wortschatz-Lexikon* dafür einen technisch erwähnenswerten Ansatz. Schließlich sind die virtuellen Kleinst-Wörterbücher und miniaturisierten Indizes zu nennen, die eine kontextsensitive Minimalanzeige lexikografischer Angaben in Form von Bubble-Helps oder Pop-Ups gewährleisten (3). Exemplarisch sind hierfür die Thesaurus-, Synonym- und Übersetzungsfunktionen in den neueren Versionen von MS Word zu nennen.

(1) Buchform und virtuelle Vollanzeige: Fenster und Spalten

Das Kompetenzzentrum Trier hat mit einschlägigen Referenzprojekten im Bereich der geisteswissenschaftlichen (Retro-)Digitalisierung zweifellos auch im Bereich der Formensprache digitaler historischer Lexikografie Maßstäbe gesetzt. Das gilt für den Bereich der retrospektiven bzw. prospektiven Datenstrukturierung ebenso wie für die Gestaltung der Benutzeroberflächen historischer Standardwerke, wie sie das *Deutsche Wörterbuch von Jacob und Wilhelm Grimm*⁹ oder das *Goethe-Wörterbuch* darstellen. Die in Trier digitalisierten Wörterbücher dokumentieren, wie weitgehend die Doppelstruktur von „Fenstern/Eingabezeilen“ einerseits und „Karten und Spalten“ andererseits zum grafischen Quasi-Standard in der Web- und Windows-Welt geworden sind. Das virtuelle Formschema, das sich mit der Formel ‚Fenster und Spalten‘ charakterisieren läßt, vereint synoptisch-hypertextuelle und traditionelle hierarchische Elemente (zum Beispiel Sucheingabezeile versus Lemmalaufleiste; Gliederungsansicht versus Artikelfließtext – siehe Abbildung 2).

9 Der Digitale Grimm. Deutsches Wörterbuch. Elektronische Ausgabe der Erstbearbeitung von Jacob Grimm und Wilhelm Grimm, hg. vom Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier in Verbindung mit der Berlin-Brandenburgischen Akademie der Wissenschaften. 2 CD-ROMs sowie Begleit- und Benutzerhandbuch, Frankfurt am Main 2004, vgl. <<http://www.Zweitausendeins.de>>.

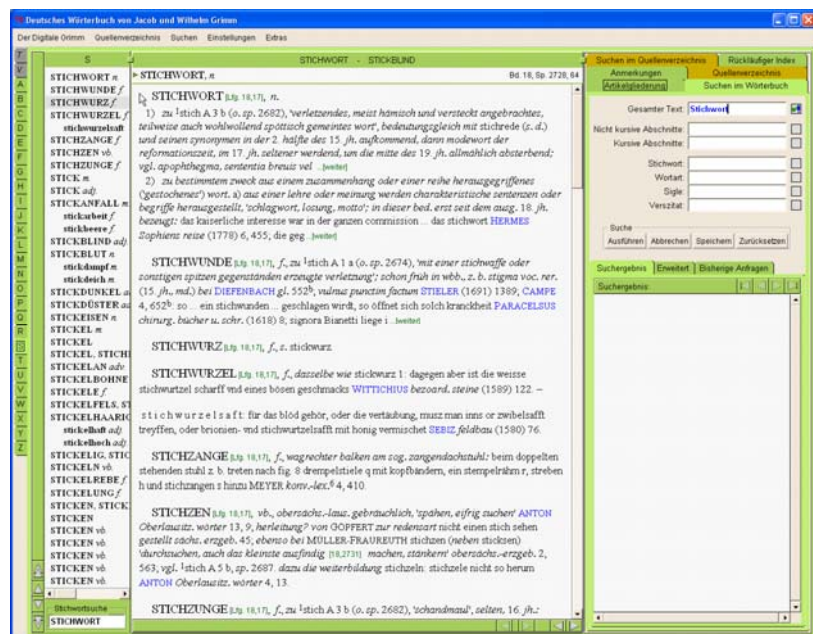


Abbildung 2: Der Digitale Grimm, lokal installierte Vollversion (Stand der Edition von 2004)

Ohne der Feinstruktur dieser virtuellen Bildschirm-Repräsentation eines monumentalen historischen Wörterbuchs bis in alle Verästelungen nachgehen zu können, sei in diesem Zusammenhang auf die einheitliche visuelle Morphologie verwiesen, wie sie sich in der Selbstdarstellung des Trierer Zentrums offenbart.¹⁰ Kulturell aus der Wörterbuch- und Wissenschaftswelt übernommene Formvorgaben wie Buch- oder Seitenmetapher, Karteikarte und Kartenreiter, stehen rein technisch bedingten Modulen gegenüber, wie sie die Eingabezeile für die Stichwortsuche oder das An-

¹⁰ Vgl. Mittelhochdeutsche Wörterbücher auf CD-ROM und im Internet (vgl. <<http://www.mwv.uni-trier.de>>) sowie die Informationsbroschüre: Elektronische Erschließungs- und Publikationsverfahren: Volltextdigitalisierung – Textstrukturierung – SGML/XML-Markup – Graphische Oberflächen – Typographie. Trier 2003, S. 3; 7-11 (vgl. <<http://www.kompetenzzentrum.uni-trier.de>>).

zeigefeld für die erweiterten Suchfunktionen rechts im Bildschirm repräsentieren. Morphologisch und semantisch lassen sich auf diese Weise die Archiv-, Buch- und Drucksatz-Elemente (Seite, Karte, Reiter, Kolumne) von den computerspezifischen Technischelementen separieren (Anzeigefeld, Hilfefenster, Pfeilnavigation). Da es sich bei allen Elementen lediglich um eine virtuelle Repräsentation handelt, kann man bestimmte Strukturen auch dichotomisch deuten. So entspricht zum Beispiel die Anzeige einer Wörterbuchseite einem (oder einer bestimmten Anzahl von) Datenbankfeld(ern) für den Artikelfließtext oder die Synopse mehrerer Buchseiten der Schachtelung einer bestimmten Anzahl von Frames. Statische Lesetextzeilen stehen dynamischen Zeilenformen gegenüber, die durch einen Hyperlink oder als Eingabezeile für die Stichwortabfrage über die Funktionalität einer gleichsam starr gedruckten Zeile hinausgehen.



Abbildung 3: Retrodigitalisierung des Goethe-Wörterbuchs, Band I-III (Stand bei Drucklegung)

Ein Blick auf den Trierer Prototyp *Das Goethe-Wörterbuch im Internet* veranschaulicht, inwieweit sich eine derartig gestaltete Oberfläche als

gelungene Synthese traditioneller und technischer Strukturelemente für die Wiedergabe eines historischen Autorenwörterbuches erweist (siehe Abbildung 3).¹¹

Die Beispiele veranschaulichen, in welchem hohem Maße Traditionen aus dem Buchdruck und der Typografie die Gestaltung digitaler Medien geprägt haben. So belegt das Beispielprojekt ‚Goethezeitportal‘ anschaulich, wie der Kolumnensatz des klassischen Zeitungs- und Zeitschriften-drucks das visuelle Informationsmanagement im Rahmen eines modernen Content-Management-Systems (CMS) dominieren kann.¹²

(2) Integrierte Kompaktanzeige externer Wörterbücher

In der expandierenden Welt der lexikografischen Informationssysteme haben sich neue Wörterbuchtypen etabliert. Bis auf Ausnahmen sind diese Online-Ressourcen auf korpuslinguistische Wissensextraktion spezialisiert. Sie lassen daher Kulturwissenschaftler/innen, Historiker/innen oder Philologen/innen oft eher ratlos zurück. Zudem erscheint ihr lexikografischer und wissenschaftlicher Anspruch im Einzelfall einer populären Breitenwirksamkeit untergeordnet.¹³ Mit Blick auf ihre technische Interaktivität und offene Nutzbarkeit bieten derartige Projekte allerdings durchaus interessante Anregungen. So können die Nutzer/innen auf der Website des Leipziger *Wortschatz-Lexikons* durch Kopieren des bereitgestellten HTML-Codes die externe Abfrage der zugrunde liegenden Korpusdatenbank in ihre lokale oder netzgestützte Anwendung integrieren. Dies geschieht über ein HTML-Formular (Applet), das auf ein entsprechend eingerichtetes Common Gateway Interface (CGI) des Leipziger Wortschatz-Servers zugreift. So ergibt sich per Skript-Übernahme nach dem Open-Source-Prinzip die Möglichkeit einer Wörterbuchabfrage, die voll-

11 Vgl. <<http://germazope.uni-trier.de/Projects/GWB>> (17.01.2006).

12 Vgl. <<http://www.goethezeitportal.de>> (Stand bei Drucklegung).

13 Vgl. den Pfad: Wortschatz-Downloads-Suchabfrage-Homepage unter <<http://wortschatz.uni-leipzig.de/>> (Stand bei Drucklegung). Eine Fülle sprechender Beispiele für die fachwissenschaftliche Nützlichkeit solcher Wörterbuchschnittstellen präsentiert die in der Germanistik einschlägige ‚Erlanger Liste‘, vgl. <<http://www.erlangerliste.de/ressourc/liste.html>> (Stand bei Drucklegung).

ständig in das Erscheinungsbild der jeweiligen Benutzeroberfläche eingehen kann (siehe Abbildung 4).

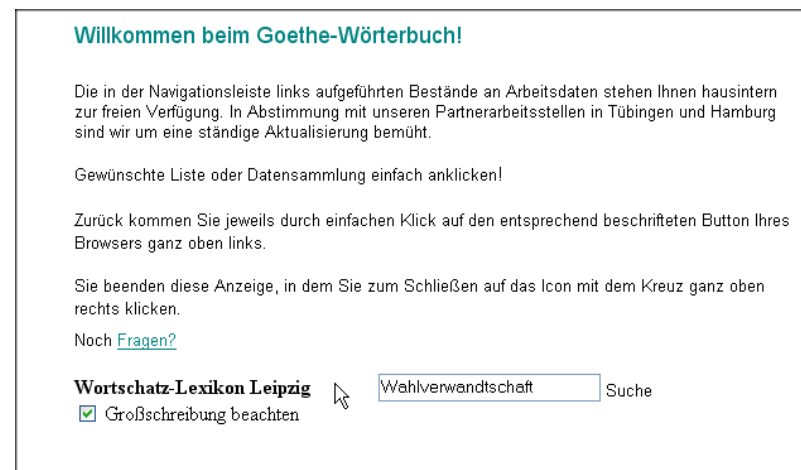


Abbildung 4: Abfrage Wortschatz-Lexikon Leipzig in Datensammlung *GWb intern*

Erst mit der Ergebnisanzeige wird die externe Wortschatz-Datenbank sichtbar (siehe Abbildung 5).

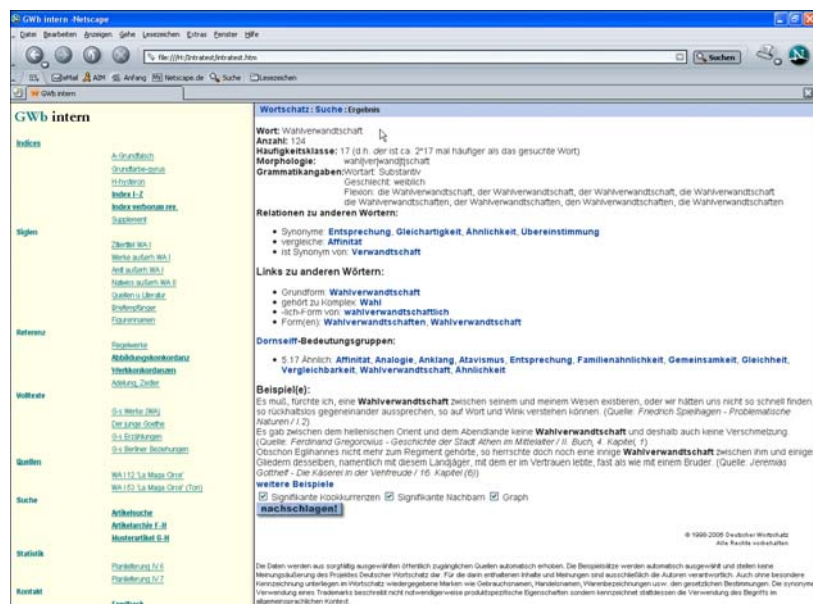


Abbildung 5: Ergebnis-Anzeige Wortschatz-Lexikon, Stichwort *Wahlverwandtschaft*

(3) Kontextsensitive Minimalanzeige lexikografischer Angaben

Die technischen Entwicklungen seit der Jahrtausendwende haben ihre immensen Beschleunigungstendenzen rigoros vom Hardware- auf den Software-Bereich übertragen. So verlangt das Wireless Access Protocol (WAP) der schnurlosen Telekommunikation nicht nur eine noch ökonomischere Rationalisierung der Datenpakete im Internet, sondern auch eine grafische und typografische Verkleinerung der semantischen Schnittstellen. Ein Index oder Thesaurus, der per Mobiltelefon abgefragt werden soll, stellt ganz neue Anforderungen an die Wörterbuchmacher/innen. Auch im zumeist auf größere Displays optimierten Bereich der Standard-Software für Personal Computer spiegeln integrierte Kleinst- oder Minimalwörterbücher diesen Trend. So sind die Wörterbuchfunktionen im Falle von MS Word in der Version 2003 zu kontextsensitiven Hilfenfenstern ausge-

baut. Mehrsprachige Thesauri, Synonymenverweise oder Übersetzungsvorschläge können so per Markierung oder rechter Maustaste eingeblendet werden (siehe Abbildung 6). Neuere Versionen des Internet-Explorers bieten Wörterbucheinträge sogar automatisch per Mouse-Over als Pop-Up bzw. Bubble-Help.

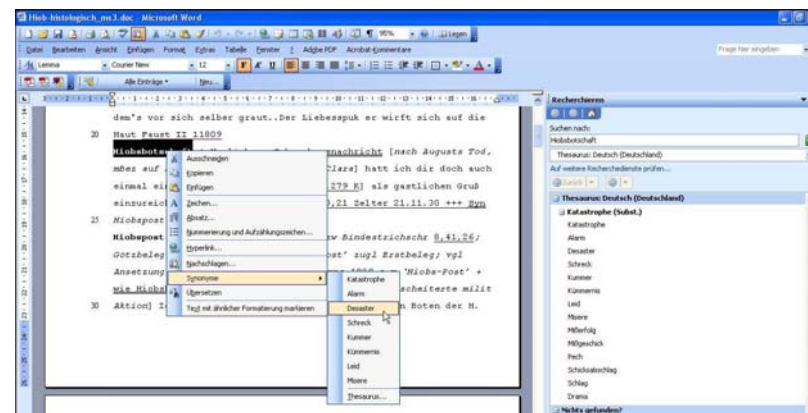


Abbildung 6: Wörterbuch-Hilfen in MS Word 2003, Beispielwort *Hiobsbotenschaft*

Das Wörterbuch der Zukunft: Visualisierung und Miniaturisierung

Folgt man der geschichtlichen Dynamik von *Lexikon* und *Wörterbuch* als Medium der Repräsentation und Anordnung von ‚Geschichte im Netz‘, so scheinen drei Tendenzen die Zukunft der Lexikografie im Internet und in den neuen Medien zu bestimmen: die Tendenzen zu *Vernetzung*, *Visualisierung* und *Miniaturisierung* lexikografischer Informationsstrukturen. Für die Wörterbuch*vernetzung*, das heißt die integrative Entwicklung mächtiger virtueller Wörterbuchplattformen, die ganze Bündel oder sogar Verbünde von historischen Fachwörterbüchern lemmabezogen miteinander verbinden, haben das Berlin-Heidelberger Wörterbuchportal¹⁴ mit dem retrodigitalisierten *Wörterbuch der Deutschen Gegenwartssprache* bzw.

14 Vgl. <<http://www.woerterbuch-portal.de>> (Stand bei Drucklegung).

dem digitalen *Wörterbuch der deutschen Sprache des 20. Jahrhunderts* (DWDS) sowie das Trierer Wörterbuchnetz¹⁵ mit dem *Deutschen* und dem *Goethe-Wörterbuch* wissenschaftliche und technische Maßstäbe gesetzt. Für Detailinformationen zu diesem Aspekt sei auf die entsprechenden Beiträge im Rahmen dieses Tagungsbandes verwiesen.

Abschließend bleibt zu resümieren, dass sich die Entwicklungen der Visualisierung und Miniaturisierung wiederum in der Geschichte bedeutender lexikografischer Werke im Geiste der europäischen Aufklärung antizipiert finden. So war es nicht zuletzt die Diderot/d'Alembertsche *Encyclopédie*, die mit ihren monumentalen, separat edierten Tafelbänden die Fülle der schönen neuen Alltags- und Technikwelt veranschaulichte und damit das abendländische Bildgedächtnis revolutionierte. Und hat Voltaire mit seinem *Dictionnaire philosophique, portatif* (1764) nicht in gewisser Weise die usuelle und ubiquitäre Mobilisierung des Wissens vorweggenommen, die mit der heutigen technischen Allverfügbarkeit von Informationen Realität geworden ist? Als virtuelle Informationssysteme – werk- und korpusübergreifend vernetzt, miniaturisiert und multimedial abrufbar, werden wissenschaftliche Wörterbücher in ihrer digitalen Derivation einmal mehr an ihre Ursprünge zurückkehren: als digitales *Encheiridion* (Handbüchlein) oder elektronisches „Allbuch“, als allgegenwärtiges und unverzichtbares informationelles Gebrauchsmedium der Zukunft.

Abbildungsnachweise

1. Bildzitat: Vogt, Friedrich; Koch, Max, *Geschichte der deutschen Literatur von den ältesten Zeiten bis zur Gegenwart*, 2. Auflage, Bd. 1 (Leipzig 1904), S. 29 – Wiedergabe einer Seite aus dem *Vocabularius Sancti Galli*. Nach dem Original in der Stiftsbibliothek zu St. Gallen; Hervorhebung R. C.
2. Bildzitat [Screenshot]: *Der Digitale Grimm. Deutsches Wörterbuch. Elektronische Ausgabe der Erstbearbeitung von Jacob Grimm und Wilhelm Grimm*, hg. vom Kompetenzzentrum für elektronische

15 Vgl. <<http://germazope.uni-trier.de/Projects/WBB/woerterbuecher/>> (Stand bei Drucklegung).

- Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier in Verbindung mit der Berlin-Brandenburgischen Akademie der Wissenschaften. 2 CD-ROMs sowie Begleit- und Benutzerhandbuch, Frankfurt am Main 2004, Ergebnisanzeige der Abfrage nach ‚Stichwort‘.
3. Bildzitat [Screenshot]: <<http://germazope.uni-trier.de/Projects/GWB/>> [Stand bei Drucklegung, September 2006].
 4. Bildzitat [Screenshot]: BBAW intern, lokale Datensammlung; vgl. dazu Demo-Ansicht unter <<http://www.goethe-wortschatz.de/>> | Link ‚Intern‘ (Passwort) [17.01.2006].
 5. Bildzitat [Screenshot]: Lokale Datensammlung „GWb intern“ [siehe zu Abbildung 4].
 6. Bildzitat [Screenshot]: Arbeitsbildschirm Microsoft Word für Windows XP Professional, Version 2003 [19.01.2006].

Dr. Robert Charlier ist Wissenschaftlicher Mitarbeiter der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW), Akademienvorhaben Goethe-Wörterbuch, und lehrt an der Technischen Universität Berlin sowie an der Universität Potsdam. E-Mail: charlier@bbaw.de

WÖRTERBÜCHER ALS DIGITALE RESSOURCEN FÜR MENSCH UND
MASCHINE – DIE WÖRTERBUCHPROJEKTE DER BERLIN-
BRANDENBURGISCHEN AKADEMIE DER WISSENSCHAFTEN

von *Gerald Neumann*

Der Vortrag berichtet über die Wörterbuchprojekte der Berlin-Brandenburgischen Akademie der Wissenschaften. Die Bearbeitung der akademischen Wörterbücher im Rahmen der Telota-Initiative wird beschrieben. Erläutert wird ein neuer, dynamischer Ansatz in der Aufbereitung von digitalen Wörterbuchressourcen. Dieser Ansatz wird an vier bearbeiteten Projekten vorgestellt.

Einführung

Der überwiegende Teil der hier vorgestellten digitalen Wörterbuchressourcen wurde im Rahmen der Telota¹-Initiative der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) erstellt. Im Jahre 2002 ins Leben gerufen, soll Telota Instrumente entwickeln, mit denen die BBAW ihre Forschungsergebnisse digital erarbeitet, dokumentiert und präsentiert. Sie unterstützt und befördert die Standards der digitalen geisteswissenschaftlichen Arbeit an der Akademie.

Die aktuellen Aufgaben von Telota sind die Evaluation des digitalen Zustands der Forschungsvorhaben der BBAW, die Entwicklung von Konzepten und das Setzen von Zielen der Entwicklung digitaler Arbeitsmethoden sowie die Initiierung von Projekten zur Förderung der digitalen Forschungsarbeit. Das aktuelle Projekt (2005-2006) ist das „Telota-Projekt

des Monats“ (TPM). Es präsentiert jeden Monat ein neues, digitales Projekt aus den Akademienvorhaben. Es soll das Informationsangebot für die geisteswissenschaftliche Forschung bereichern und einer interessierten Öffentlichkeit Einblicke in die aktuelle Forschung geben. Das TPM arbeitet seit Februar 2005.

Die Schwerpunkte der Projekte liegen, induziert durch die an der Akademie angesiedelten Vorhaben, im Bereich der digitalen Editionen, Wörterbücher und Prosopografien. Darüber hinaus gibt es verschiedene andere Vorhaben, für die im Rahmen des TPM die unterschiedlichsten Einzelfalllösungen erarbeitet wurden, wie zum Beispiel einen georeferenzierten Zugang zu einer kunsthistorischen Datenbank per GoogleMaps (Census), die wechselseitige Verknüpfung von Datenbank und Scan-Galerie (Ritter-Katalog der Leibnizforschung) oder die Entwicklung des Archiv-Editors für das Vorhaben *Preußen als Kulturstaat*, der aber allen „Archivgängern“ nützlich sein wird.² Die bisher vorliegenden Ergebnisse vom TPM lassen sich nur unter den zeitlichen und personellen Bedingungen angemessen einschätzen, unter denen sie entstanden sind. Der Koordinator, die beiden Mitarbeiter und die zwei Studenten sind jeweils mit einer halben Stelle für die Aufgaben vom TPM angestellt. Das Projekt trägt den Namen „Projekt des Monats“ mit voller Berechtigung, da es de facto nur jeweils einen Monat für die Arbeit in einem Vorhaben zur Verfügung hatte. Da die Monate zu Beginn der Arbeit von TPM den Vorhaben zugeordnet wurden, gab es auch keine Möglichkeit, an diesem Zustand etwas zu ändern. Zu Beginn jedes neuen Monats stand das aktuelle Vorhaben mit seinen Forderungen und Aufgaben schon bereit. Um nicht eine Vielzahl von zusammenhangslosen Einzellösungen zu produzieren, wurde von Beginn an versucht, eine einheitliche technische Grundlage zu schaffen und weiterzuentwickeln, die mit möglichst geringem technischen Aufwand die spezifischen Anpassungsaufgaben, die sich aus den Einzelprojekten ergaben, ermöglichen sollte. Wie diese technische Grundlage für die

¹ Telota = The electronic life of the academy.

² Eine vollständige Liste aller Projekte findet sich unter <<http://www.bbaw.de/pom/pom.html>> (09.11.2006). Zum Projekt *Preußen als Kulturstaat* vgl. Czmiel, Alexander; Holtz, Bärbel, Quellenarbeit im Projekt „Preußen als Kulturstaat“, in diesem Band.

Wörterbuchprojekte aussieht und welche Anwendungen sie ermöglicht, soll im Folgenden vorgestellt werden.

Überblick

Akademische Wörterbücher in eine angemessene digitale Form zu verwandeln, zählt zu den derzeit größten Herausforderungen der elektronischen Textaufbereitung. Neben historisch-kritischen Editionen sind Wörterbücher aus elektronischer Sicht diejenige Textsorte mit der höchsten Informationsdichte. Die adäquate Darstellung der über Generationen entstandenen Informationskodierung in einem Wörterbuchartikel und die gleichzeitige Umsetzung der neuen digitalen Möglichkeiten verlangen eine sorgfältige Analyse des gedruckt Bestehenden und eine umfassende Kenntnis des digital Möglichen. Dieser Aufgabe hat sich das TPM in den vergangenen Monaten verstärkt gewidmet. Nachdem schon im April 2005 das *Wörterbuch der deutschen Gegenwartssprache* (WDG) bearbeitet wurde, waren im Juni und August 2005 das *Deutsche Wörterbuch* (DWB) und das *Goethe-Wörterbuch* (GWB) an der Reihe. Von den beiden Letzteren wurde jeweils die aktuelle Lieferung digital aufbereitet und im Internet präsentiert. Die Einbeziehung dieser drei Projekte folgt dem Plan, schrittweise ein experimentelles, digitales System für die akademischen Wörterbücher der BBAW aufzubauen. Dieses System soll nicht nur dem unterschiedlichen Aufbau des jeweiligen Wörterbuches gerecht werden, sondern auch konsequent die Vorteile einer strukturierten digitalen Aufarbeitung deutlich machen.

Die Aufnahme der drei Wörterbücher in das Wörterbuch-Portal³ der Akademie verdeutlicht darüber hinaus die Vernetzungsmöglichkeiten, die sich aus einer strukturierten Aufarbeitung der Wörterbücher ergeben.

Alle vorgestellten digitalen Wörterbuch-Projekte basieren auf der gleichen Technologie. Wie inzwischen üblich, lagen die Wörterbücher als XML-Dokumente vor, bzw. wurden in eine XML-Struktur überführt. Die Vorteile einer solchen strukturierten Erfassung von lexikografischen Daten

³ Das Wörterbuch-Portal ist unter <http://www.woerterbuch-portal.de> (22.11.2006) zu erreichen.

hinsichtlich Langzeitarchivierbarkeit und Softwareunabhängigkeit liegen auf der Hand. Das vom TPM entwickelte System geht aber noch einen Schritt weiter. Auf der Grundlage der nativen XML-Datenbank eXist⁴ wurde die Möglichkeit geschaffen, auf die zugrunde liegenden XML-Strukturen während des Recherche- und Anzeigeprozesses zuzugreifen. Dies ist dadurch möglich, dass das für die Anzeige benötigte Format, im Normalfall sind dies HTML-Dokumente, erst auf eine Nutzeranfrage hin direkt aus den XML-Dokumenten generiert wird. Die XML-Dokumente werden unter Benutzung des W3C-Standards XQuery⁵ abgefragt. Das Ergebnis dieser Abfrage, das ein XML-Fragment ist, wird dann unter Benutzung des W3C-Standards XSLT⁶ in ein HTML-Dokument transformiert und an den anfragenden Web-Browser ausgeliefert. Dieser bringt es dann den Benutzern/innen zur Ansicht.

Hier kann sich natürlich die Frage stellen, warum so kompliziert, wenn es auch einfacher geht? Auf diese Frage will der Vortrag zwei Antworten geben. Erstens, das beschriebene Verfahren ist nicht so kompliziert, wie es auf den ersten Blick scheint. Zweitens, das gewählte Verfahren eröffnet Möglichkeiten, die anders nur schwer und mit erheblichem Mehraufwand oder aber gar nicht zu erreichen sind.

Die Aufbereitung von XML-Daten zur Anzeige für menschliche Benutzer/innen ist ein Vorgang, der nicht nur dann anfällt, wenn die Anzeige, wie in unserem Fall, direkt aus den XML-Dokumenten generiert wird. Das heißt, die Konvertierung in ein lesbares Format, wie zum Beispiel HTML oder PDF findet bei jedem elektronisch verfügbaren Wörterbuch statt, sei es nun in einer relationalen Datenbank gespeichert oder als statisches HTML angelegt. Die Generierung des Anzeigeformats zur Laufzeit erzeugt also keinen zusätzlichen Arbeitsaufwand, sondern hat allenfalls positive Auswirkungen auf die direkten Korrekturmöglichkeiten der Quellen und

⁴ Der Terminus „native XML-Datenbank“ bezeichnet die besondere Eigenschaft einer Datenbank, spezielle technische Voraussetzungen für die genuine Verarbeitung von XML-Dokumenten zu besitzen. Vgl. <http://www.exist-db.org> (22.11.2006).

⁵ Vgl. <http://www.w3.org/XML/Query/> (22.11.2006).

⁶ Extensible Style Language Transformation, vgl. <http://www.w3.org/TR/xslt20/> (22.11.2006).

eventuell negative Auswirkungen auf die Performance der Anzeige. Die Trennung von Daten-, Anfrage- und Anzeigeschicht bringt zwar einen erhöhten Implementierungsaufwand mit sich, lässt dafür aber im Gegenzug auch den gezielten Eingriff in nur einen Bestandteil des Gesamtsystems zu. So wird es möglich, die XSL-Transformation der HTML-Seiten zu ersetzen durch eine XSL-FO-Transformation, die aus den Suchergebnissen ein PDF-Dokument erzeugt, ohne dass an der Daten- oder Anfrageschicht etwas geändert werden muss. Dadurch wird es für Nutzer/innen möglich, Wörterbuchauszüge bzw. ganze Wörterbücher auf der Grundlage von speziellen Nutzerinteressen als PDF-Dokumente zu generieren. Ein auf der .hist 2006 vorgestelltes Beispiel ist die Generierung eines suffix-geleiteten Teilwörterbuchs auf der Basis des GWB für die Suffixe *-heit* und *-keit*.

Auch an den anderen Schichten lässt sich ansetzen. So kann der Output der Anfrageschicht, ein XML-Fragment, auch direkt an einen Nutzer/eine Nutzerin weitergereicht werden, ohne vorherige Transformation in ein Anzeigeformat. Dadurch entsteht eine Schnittstelle für eine mögliche maschinelle Nutzung der Wörterbuchressource. Die Anfrageschicht lässt über XQuery zudem die Formatierung des ausgegebenen XML-Fragmentes zu. Das heißt, die weiterverarbeitende Maschine kann ihr XML-Fragment in genau der Form bekommen, in der sie es braucht. Die dafür notwendigen Arbeitsschritte können und sollen von den Nachnutzern/innen des Wörterbuchs ausgeführt werden und nicht von denjenigen, die es bereitstellen.

Aber das gewählte technische Verfahren eröffnet nicht nur interessante technische Möglichkeiten. Auch die inhaltliche Arbeit und Nutzung des Wörterbuchs gewinnt durch die Teilung des Verfahrens in die unterschiedlichen Schichten. So ermöglicht der Nutzerzugriff auf die Anfrageschicht das Erstellen von Anfragen, die nur durch die vorhandenen Strukturen der XML-Dokumente beschränkt sind und nicht durch eine vorgegebene Suchmaske mit zwangsläufig vorgegebenen Recherchemöglichkeiten. Als ein sprechendes Beispiel sei hier die Neubearbeitung des *Deutschen Wörterbuchs* von Jacob Grimm und Wilhelm Grimm beschrieben. Die Tatsache, dass in diesem Wörterbuch jeder Beleg eine strukturierte zeitliche Einordnung besitzt und dass die Belege innerhalb der Artikelstruktur

eindeutig der Bedeutungsvariante eines Stichworts zugeordnet sind, lässt eine Wörterbuchbenutzung zu, wie sie bisher nicht realisiert wurde. So ist es möglich, durch eine XQuery-Anfrage alle Stichwörter des Wörterbuchs zu bekommen, die über eine Bedeutungsvariante verfügen, deren erster Beleg nicht vor dem Jahre 1832 liegt. Ebenso ist es möglich, alle die Stichwörter auszugeben, die nach 1832 nicht mehr belegt sind.

Wozu ist das gut? Mit diesen beiden Anfragen lässt sich eine Liste von Stichwörtern zusammenstellen, die die Lektüre historischer Texte, in diesem Fall von Goethetexten, unterstützt. Die zweite Anfrage generiert offensichtlich eine Liste von Wörtern, die heute nicht mehr benutzt werden. Diese Wörter würden natürlich bei der Lektüre eines Textes ohnehin auffallen, so dass mit der vorangehenden Textaufbereitung nur der Nachschlageprozess überflüssig wird. Die erste Anfrage hingegen listet Wörter auf, die aus der heutigen Perspektive eine andere, zusätzliche Bedeutungsvariante haben als zur Goethezeit. Das heißt, die Möglichkeit einer Fehlinterpretation der Wortbedeutung ist gegeben. Wollte man die Anforderungen an eine digitale Wörterbuchressource auf die Spitze treiben, so könnte man von ihr verlangen, dass sich das Wörterbuch von selbst zu Wort meldet, wenn eine solche Möglichkeit der Fehlinterpretation beim Lesen historischer (elektronischer) Texte besteht.

Die Trennung der drei Aufbereitungsschichten ermöglicht also den getrennten, dynamischen Zugriff und auch eine schichtenabhängige Nutzerinteraktion mit dem System. Dies erweitert die Möglichkeiten der menschlichen und der maschinellen Benutzung der gesamten Wörterbuchressource beträchtlich. Außerdem enthebt es den Betreiber eines elektronisch verfügbaren Wörterbuchs von der unlösbaren Aufgabe, allen Nutzerinteressen zu genügen, bzw. selbst festlegen zu müssen, welche Nutzungsmöglichkeiten die ‚richtigen‘ sind. Was nicht heißen soll, dass er für die Nutzbarkeit seiner Ressource keine Verantwortung trägt. Im Gegenteil, mehr als bisher sollten die verschiedenen Nutzungsmöglichkeiten einer digitalen Wörterbuchressource Kriterium für die qualitative Bewertung ihrer elektronischen Präsentation sein.

Die auf dem Vortrag vorgestellten digitalen Wörterbuchressourcen sind folgende:

Wörterbuch der deutschen Gegenwartssprache (WDG)

Beim ersten Wörterbuchprojekt handelte es sich um die Erweiterung der Abfragemöglichkeiten des *Wörterbuchs der deutschen Gegenwartssprache* (WDG)⁷, dessen Online-Ausgabe seit März 2003 vom Projekt „Digitales Wörterbuch der deutschen Sprache“ bereitgestellt wird. Die Aufgabe von TPM bestand in der so genannten ‚Auszeichnung‘ von besonderen inhaltlichen Strukturen in den Wörterbuchartikeln. Es wurden sämtliche Belegstellen und Siglen ausgezeichnet und mit dem speziell dazu digitalisierten Quellenverzeichnis von ungefähr 1.000 Werken verknüpft. Dadurch ergaben sich völlig neue Funktionalitäten in der Wörterbuchabfrage. So ist es jetzt nicht nur möglich von den einzelnen Siglen an die zutreffende Stelle im Quellenverzeichnis zu springen, sondern auch vom Quellenverzeichnis zu allen Stichwörtern, die sich aus einer bestimmten Quelle speisen.

Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm (DWB)

Beim *Deutschen Wörterbuch*⁸ wurde erstmals eine Schnittstelle für den freien Zugriff auf die Daten nach einem vom W3C verabschiedeten Standard XQuery eingefügt, womit das Wörterbuchsystem über die Begriffsabfrage hinaus erweitert wird. Die Nutzer/innen können jede mögliche Anfrage stellen, die sich aus der Strukturierung der Daten ergibt, niemand gibt ihnen mehr vor, welche Fragen sie haben dürfen. Ein Beispiel für eine solche freie Anfrage ist etwa: Welches sind die am häufigsten zitierten Autoren in der Lieferung des DWB? Als Antwort erscheinen die ersten sechs Autoren des Rankings: Luther (53), Schiller (31), Franck (30), Goethe (29), Lessing (28), Herder (27), wobei die Zahlen in den Klammern die Häufigkeit angeben. Diese zusätzlichen Funktionen machen das Wörterbuchsystem zu einer echten digitalen Ressource, die in ihrer Komplexität nicht nur für die Mitarbeiter/innen bereitsteht, sondern auch von jedem/jeder anderen genutzt werden kann.

Darüber hinaus wurde beim DWB ein halb-automatisches Verfahren entwickelt, das die originalen Tustep-Daten in XML-Daten umwandelt.

7 Vgl. <<http://www.dwds.de>> (22.11.2006).

8 Vgl. <<http://pom.bbaw.de/dwb>> (22.11.2006).

Goethe-Wörterbuch (GWB)

Auch die Aufarbeitung der Lieferung des *Goethe-Wörterbuchs*⁹ passt sich nahtlos in das Wörterbuchsystem ein. Die strukturierten Daten werden direkt aus dem lexikografischen Workflow gewonnen. Auch die XQuery-Schnittstelle des Wörterbuchsystems kann genutzt werden.

In der Digitalversion des GWB neu hinzugekommen ist die Möglichkeit der Online-Bearbeitung von XML-Dokumenten. Damit ist der Anfang eines Online-Artikel-Redaktionssystems gemacht. Bisher gestattet das Redaktionssystem das Anlegen neuer, strukturierter Artikel und das Editieren bestehender Artikel auf der Basis und in den Grenzen der vorgegebenen XML-Strukturen.

Als eine weitere Erweiterung gibt es beim GWB die Option der PDF-Ausgabe nach selbst bestimmbareren inhaltlichen und gestalterischen Merkmalen. Dabei werden die PDF-Dokumente unter Benutzung des XML-FO-Standards direkt aus den XML-Dokumenten generiert.

Zusatz: Wörterbuch der manichäischen Texte (WMT)

Eine weitere Möglichkeit des Wörterbuchsystems wurde durch das „Telota – Projekt des Monats Januar 2006“ eindrucksvoll demonstriert. Durch die strukturierte Aufbereitung und Bereitstellung des *Wörterbuchs der manichäischen Texte*¹⁰ wurde eine automatische Verknüpfung von digitaler Edition und digitalem Wörterbuch möglich. Für jede Wortform aus den digital vorliegenden mittelpersischen Texten konnte so eine Verknüpfung zum entsprechenden Lemma im Wörterbuch ohne großen Aufwand realisiert werden. Am digitalen WMT zeigen sich die bisher weitgehend ungenutzten Perspektiven des gleichermaßen maschinen- und menschenlesbaren Formats XML/TEI.

Schlussfolgerungen und Konsequenzen

Die vier oben beschriebenen, konkreten Projekte haben einiges über die praktische Seite der Implementierung eines digitalen Wörterbuchsystems gezeigt. Die Erfahrungen seien hier kurz zusammengefasst.

9 Vgl. <<http://pom.bbaw.de/gwb>> (22.11.2006).

10 Vgl. <<http://pom.bbaw.de/turfan>> (22.11.2006).

Die verfügbaren Datenbanken, Webserver, andere Software und Hilfsmittel aus dem Open-Source-Bereich erfüllen die notwendigen Voraussetzungen an Funktionalität und Performance überraschend gut. Die Standards für Datenauszeichnung (XML/TEI), Datenkonvertierung (XSL, FO), und Datenabfrage (XQuery) sind für grundlegende Anliegen hinreichend weit entwickelt. Die Unicode-Unterstützung in den verwendeten Applikationen deckt fast alle Bedürfnisse der Zeichendarstellung und Suchfunktionalität ab. Die Datenmodellierung der XML-Dokumente und die konsequente Nutzung von Open Source-Technologien sichert ein hohes Maß an Unabhängigkeit und Langzeitverfügbarkeit der erschlossenen Quellen.

Die XML-Strukturierung der Daten ist nur der erste wichtige Schritt. Die XML-Strukturierungen müssen auch für Recherche und Weiterverarbeitung zur Verfügung stehen. Elektronische Ressourcen von Wörterbüchern und anderen komplexen Texten, wie zum Beispiel Digitale Editionen, müssen maschinenlesbar und –auswertbar sein. Die Anwendung von internationalen Standards ist Voraussetzung für eine Vernetzung von Informations- und Wissensbeständen.

Gerald Neumann ist Mitarbeiter beim Akademievorhaben „Digitales Wörterbuch der Deutschen Sprache“ und Koordinator des „Telota – Projekt des Monats“ an der Berlin-Brandenburgischen Akademie der Wissenschaften. Wichtigste Arbeitsgebiete: Digitale Wörterbücher, Strukturierte Erfassung der Ergebnisse geisteswissenschaftlicher Grundlagenforschung. E-Mail: gneumann@bbaw.de

BAYERISCHES MUSIKER-LEXIKON ONLINE

von Josef Focht

Das Projekt der Deutschen Forschungsgemeinschaft (DFG) Bayerisches Musiker-Lexikon Online (BMLO) wird als virtuelles biografisches Lexikon zur bayerischen Musikgeschichte erarbeitet. Es soll eine auf Biografien gestützte grundlegende Dokumentation der bayerischen Musikgeschichte in zeitgemäßer und allgemein zugänglicher Form bereitstellen. Derzeit (Februar 2006) informiert es über seine Strukturen und den Fortgang seiner Vorbereitung und Umsetzung. Zukünftig wird es als Lexikon Benutzern/innen aller Fächer und Bildungshorizonte unbeschränkt und kostenfrei zugänglich sein. Der Editionsbeginn des BMLO ist für das zweite Quartal 2006 vorgesehen; bis 2008 soll es bis zur vollständigen Editionsreife geführt werden, sodass dann Datensätze von deutlich über 20.000 Personen medial angereichert abgerufen werden können. Der folgende Aufsatz gibt einen Überblick über die fächerübergreifende Struktur des Projekts und beschreibt Verfahren zur Umsetzung des Lexikons sowie Prioritäten in der Auswahl der Quellen.

Einleitung

Der Bedarf an wissenschaftlich qualifizierten und kontrollierten Informationsangeboten im Internet ist in den vergangenen Jahren für alle Fächer sprunghaft angestiegen. Deshalb wurden in jüngster Zeit in verschiedenen Wissenschaftsbereichen digitale Lexika konzipiert und realisiert. Das Bayerische Musiker-Lexikon Online (BMLO)¹ will für die historische Musikwissenschaft ein vergleichbares Pilotprojekt entwickeln. Es soll eine auf Biografien gestützte grundlegende Dokumentation der bayerischen

¹ Vgl. <<http://www.bml0.uni-muenchen.de>>.

Musikgeschichte in zeitgemäßer und allgemein zugänglicher Form bereitstellen, die modular aufgebaut, vertiefungs- und erweiterungsfähig ist.

Das BMLO wird derzeit (März 2006) als virtuelles Lexikon vorbereitet; es informiert über seine Strukturen und den Fortgang seiner Bearbeitung. Zukünftig wird es Benutzern/innen als biografisches Lexikon zugänglich sein.

Trägerschaft, Finanzierung und Förderung

Das BMLO ist ein Projekt zur Erfassung, Digitalisierung, Erschließung und Bereitstellung biografischer Daten und personenbezogener Medien. Es wird seit Anfang 2004 vom Institut für Musikwissenschaft an der Universität München in Kooperation mit der Bayerischen Staatsbibliothek (BSB) und der Gesellschaft für Bayerische Musikgeschichte durchgeführt. Als Partner im wissenschaftlichen Austausch fungieren in- und ausländische Universitäten, Akademien, Forschungseinrichtungen und Lexikonprojekte sowie einzelne Autoren/innen. Bei der Durchführung des Projekts stehen neben den Eigenmitteln der Träger – also der Universität München, der BSB und der Gesellschaft für Bayerische Musikgeschichte – auch Drittmittel der Edith-Haberland-Wagner-Stiftung und der DFG zur Verfügung.

Die BMLO-Struktur im Überblick

Das BMLO ist als Online-Lexikon zu Personen der bayerischen Musikgeschichte (Komponisten/innen und Dirigenten/innen, Sänger/innen und Instrumentalisten/innen, Musikverleger/innen, Instrumentenbauern/innen und Musikforscher/innen aus allen Epochen) angelegt. Seine Daten- und Editionsstruktur bietet eine recherchefähige Datenbank, darüber hinaus Lexikonartikel und multimediale Ergänzungen.

Infolge seiner interdisziplinären Verortung – zwischen der Musikwissenschaft und der Landesgeschichte und mit wesentlichen Beiträgen des Bibliothekswesens sowie der Informatik – erfordert das BMLO ebenso dringend eine bedarfsgerechte und aufwändige Methoden-Entwicklung wie infolge seines Pilotprojekt-Charakters und seines großen Volumens.

Das BMLO führt die biografische Forschungsarbeiten aus der musikwissenschaftlichen und landeskundlichen Literatur zusammen, retrokon-

vertiert lexikalische Quellen, erschließt diese sowie weitere Bibliotheks- und Archivbestände; schließlich vernetzt es hochwertige biografische Internet-Ressourcen. Die Datenfülle des BMLO zielt mit über 20.000 Lemmata auf eine Dokumentationsbreite ab, die den Themenkomplex weiträumig abdeckt; ferner bemüht es sich um den Nachweis relevanter Quellen in möglichst großer Erschließungstiefe.

Warum gerade Bayern?

Aus methodischen und pragmatischen Gründen ist das BMLO zunächst auf Bayern beschränkt:

Die Lexikografie zur bayerischen Musikgeschichte setzte schon 1811 mit dem „Baierischen Musiklexikon“ von Felix Joseph von Lipowsky ein. Dies war gleichzeitig die erste auf einen „Nationalstaat“ konzentrierte Sammlung von Musikerbiografien, die im 19. und 20. Jahrhundert dann lebhaft nachgeahmt wurde.

Im 20. Jahrhundert fanden viele Biografien bayerischer Musiker/innen Eingang in allgemeine musikalische Enzyklopädien², ihr überwiegender Anteil wurde jedoch in Monografien und Einzelstudien (etwa über einzelne Städte, Klöster, Hofkapellen oder den Instrumentenbau) bzw. in Periodika veröffentlicht – sie sind also weit verstreut und oft schwer auffindbar. Die relevanten Periodika sind auf die Fächer Musikwissenschaft und Landesgeschichte verteilt sowie in dem bibliografisch nur unvollständig erfassten Bereich der Grauen Literatur zu suchen.³

Bayern ist wie kaum ein anderes Bundesland seit zwei Jahrhunderten in seinen Grenzen stabil und klar definiert, was günstige Konsequenzen namentlich für das Bibliotheks- und Archivwesen hat. Es ist groß genug und soziologisch wie konfessionell differenziert genug, um innerhalb seiner Grenzen eine große Vielfalt musikhistorischer Phänomene exemplarisch abzubilden: katholisch-klösterliche ebenso wie protestantisch-reichsstädtische Musikkultur, höfisches Musikleben an kleinen, weltlichen

2 Unter anderem Die Musik in Geschichte und Gegenwart, (New) Grove Dictionary of Music and Musicians, Riemann Musiklexikon usw.

3 Wie z. B. Programmhefte, Chroniken und Festschriften von Verbänden, Vereinen, Theatern, Orchestern, Instrumentenbau-Firmen.

und geistlichen Residenzen wie an großen von europäischem Rang, ländliche ebenso wie großstädtische Popularkultur.

Bayern ist eine europäische Kernregion, die nach allen Seiten hin offen und durchlässig für Einflüsse aus allen Nachbarregionen ist. Es ist ein paradigmatisches Durchgangsland, in dem sich die wichtigsten europäischen Migrationsrouten kreuzen, und ringförmig umgeben von Musikmetropolen wie Salzburg, Wien, Prag, Dresden, Mannheim, Paris, Mailand und Venedig – was viele Musiker/innen durch Bayern führen musste. Bayern stand stets in intensiver Wechselwirkung mit musikhistorisch führenden oder besonders innovativen Nachbarregionen wie Italien, Österreich, Thüringen/Sachsen oder Böhmen.

Die BSB stellt nicht nur wegen ihres Sondersammelgebiets Musikwissenschaft einen unschätzbaren Standortvorteil dar. Hinzu kommt in Bayern eine beispiellose Dichte von historisch oder musikwissenschaftlich ausgerichteten Institutionen und Vereinigungen, vom engmaschigen Netz der Universitäten bis zu den circa 200 historischen Vereinen, den Komponisten-Instituten, Museen, Archiven und Bibliotheken. Im kultur- und musikhistorischen Bereich verfügt Bayern für ein Projekt wie das BMLO wegen der bereits seit zwei Jahrhunderten betriebenen landeskundlichen Forschung also über günstige Ausgangsbedingungen.

Auch wenn das BMLO aus pragmatischen Gründen nur „bayerische“ Musiker/innen berücksichtigt, also diejenigen, die historisch nennenswert in Bayern gewirkt haben und/oder genealogisch in Bayern verwurzelt sind, deckt es damit doch – wie oben erwähnt – einen Personenbestand ab, der vielfach auch international tätig war. Stichprobenartige Vergleiche mit Nachbarregionen (wie zum Beispiel Tschechien oder Baden-Württemberg) haben bereits Überschneidungen der relevanten Personenkreise von etwa 40 Prozent ergeben. Die enge Kooperation mit Wissenschaftlern/innen und Projekten aus allen Nachbarländern wird diesen bereits in der Sache selbst angelegten überregionalen und internationalen Charakter deutlich herausarbeiten.

Nach Abschluss der (ersten) „bavaristischen“ Phase ist das BMLO strukturell offen für Ausweitungen in alle Richtungen und Kooperationen jeder Art. Es versteht sich dann als Grundstock eines überregionalen

europäischen Musiker-Lexikons.

Interdisziplinäre Struktur

Gemäß der bereits beschriebenen fächerübergreifenden Struktur wird das BMLO in mehreren spezialisierten und differenziert profilierten Arbeitsgruppen vorbereitet: in der BMLO-Redaktion am Institut für Musikwissenschaft der Universität München (fachwissenschaftliche Redaktion), in der IT-Gruppe Geisteswissenschaften der Universität München (EDV- und Datenbankpflege), in der Digitalen Bibliothek der BSB (Retrodigitalisierung, Abgleich mit bibliothekarischen Normdaten, Langzeitarchivierung) sowie von Kooperationspartnern (musikwissenschaftliche Forschung), die auf relevante Themenbereiche spezialisiert sind.

Prioritäten in der Bearbeitung

Die Lexikondaten des BMLO müssen aus Gründen der Quantität und der Methoden in mehreren Phasen bearbeitet und ediert werden. Die höchste Priorität genießt die Entwicklung der Datenbank mit ihrer differenzierten Recherche-Funktionalität. In der nachfolgenden Ausbaustufe werden verschiedenartige Lexikonartikel hinzugefügt; sie entstehen wahlweise durch Erschließung bereits publizierter Beiträge und/oder durch Neuformulierung. In der abschließenden Ausbaustufe werden Medien hinzugefügt: Bilder, Noten, Audios, Videos.

Bearbeitungsphasen im Detail

In der Vorbereitung des BMLO sind mehrere Arbeitsphasen in der folgenden Reihenfolge nötig, die sich jedoch in beachtlichem Maß überlappen:

- Definition von Strukturen und Normen (für die Dateneingabe und -ausgabe, Datenmigration und -konversion, Redaktion und Erschließung, die Langzeitarchivierung und Verwaltung)
- Entwurf einer Namensliste, Verdichtung zur Lemmaliste, die als Grundlage personenbezogener Datensätze dient
- Quellenerschließung und Datensammlung

- Kumulation⁴ bzw. Individualisierung⁵ der Datensätze, das heißt: Zusammenführung verschiedener Namen zu einer Person bzw. bedarfsgerechte Trennung eines Namens in mehrere Personen
- Entwicklung und Optimierung von Methoden vom Versuchsstadium bis zur Serienreife
- Abgleich mit bibliothekarischen und amtlichen Normdateien (Personennamendatei und Normansetzung von Namen, Schlagwortnormdatei, Ortskennzahlen usw.)⁶
- redaktionelle Vervollständigung der Datensätze
- Transformation der plattformabhängigen Eingabedaten und Edition in spezifischen Ausgabe-Oberflächen⁷, die dem virtuellen Medium angemessen sind und den divergierenden Anforderungen des breiten Nutzerkreises gerecht werden
- Erschließung (Retrokonversion)⁸ und Bearbeitung von Lexikonartikeln

4 Ein Beispiel: Der Oberpfälzer Komponist Johann Georg Holler (* 1751 – † nach 1823) trat 1775 als Benediktiner in das Kloster Attl am Inn ein, wo er später Prior wurde. Dort führte er den Namen Pater Aegidius. Ab 1787 hielt er sich längere Zeit in Siena auf, wo seine Werke mit der Autorenangabe Diego Orelli überliefert sind. Die Zusammenführung der Namensformen „Johann Georg Holler“, „Pater Aegidius“ und „Diego Orelli“ in einen Datensatz ist als Kumulation zu verstehen.

5 Ein Beispiel: Gegenwärtig enthält das BMLO etwa 70 Datensätze zu Personen mit dem Nachnamen „Fischer“. Bei einigen liegen sehr wenige biografische Daten vor, gleichzeitig erstrecken sich ihre Wirkungsdaten auf so große Zeiträume, dass sie sich wohl auf verschiedene gleichnamige Personen – etwa Vater und Sohn – beziehen. Diese Datensätze bedürfen dann einer Individualisierung.

6 Die bibliothekarischen Normdateien werden von der Deutschen Bibliothek Frankfurt am Main, vgl. <<http://www.ddb.de>>, als zentrale bibliografische Dienstleistungen betreut. Die Ortsdaten werden von den Statistischen Ämtern der Länder, des Bundes und der Europäischen Gemeinschaft geführt, vgl. <<http://www.statistikportal.de>>.

7 Aus methodischen Gründen hat sich die Trennung von Dateneingabe, -transformation und -ausgabe als zweckmäßig erwiesen. Die Pflege des sehr großen BMLO-Datenbestandes erfordert damit bei geringer Fehlerdichte eine vergleichsweise kurze Bearbeitungszeit. Voraussetzung für diese Trennung der Datenebenen ist eine höchstmögliche Konsistenz in den Ansetzungsformen.

8 Zwei Bestandsgruppen der gedruckten biografischen Literatur kommen für die Retrokonversion in Betracht: erstens überregionale Lexika mit einem nennenswerten Anteil für das BMLO relevanter Personen, zweitens Monografien zu Themen der bayerischen Musikgeschichte mit einem hohen biografischen Informationsgehalt.

- Erschließung weiterer Medien durch Digitalisierung
- fortlaufende fachliche, formale und EDV-technische Kontrolle.

Die Qualität lexikalischer Daten hängt wesentlich vom Grad der Kumulation bzw. Individualisierung der Datensätze ab. Im Gegensatz zu einer Vielzahl von anderen gegenwärtigen Retrokonversionsprojekten leistet die BMLO-Redaktion die unverzichtbare fachwissenschaftliche Beurteilung und Bewertung der höchst inhomogenen Quelldaten, womit die Qualität der BMLO-Daten wesentlich steigt und ihre Benutzer/innen von deren fachlicher Kontrolle entlastet werden, die sie in der Regel weniger leicht leisten können als die BMLO-Redaktion.⁹

Auswahl und Priorität der Quellen

Um gleichzeitig die große Anzahl der Lemmata möglichst effektiv bearbeiten und trotzdem dem Einzelfall gerecht werden zu können, verfolgt das BMLO zwei parallele Strategien bei der Auswahl und Erschließung von Quellen.

- Primäre Strategie der seriellen Erschließung: In der BMLO-Redaktion werden serielle Quellen erschlossen, das heißt Lexika, Monografien, Archivbestände, Internet-Ressourcen vor allem der Musikforschung und der Landesgeschichte.
- In der Digitalen Bibliothek der BSB wird die Retrokonversion historischer musikwissenschaftlicher Lexika und bavaristischer Standardliteratur vorgenommen.
- Sekundäre Strategie der individuellen Vertiefung: Der Ertrag aus seriellen Quellen wird durch individuelle Vertiefung ergänzt, die jedoch von der BMLO-Redaktion allein aus Gründen der Ökonomie nicht zu leisten ist. Sie geschieht in der Regel in Kooperation mit Partner-Institutionen und spezialisierten Autoren/innen.

9 Die Datenkontrolle innerhalb der BMLO-Redaktion geht weit über die Beurteilung musikhistorischer Sachverhalte und die Diskussion von Forschungskontroversen hinaus; sie korreliert etwa auch historische Ortsnamen oder ergänzt Daten, die nicht expressis verbis zu lesen, aber doch aus dem Kontext zu erschließen sind.

BMLO-Datenmodell

Die BMLO-Daten sind plattformunabhängig strukturiert, womit ihre Langlebigkeit besonders wahrscheinlich wird. Die Datensätze sind beliebig aktualisierbar, erweiterungs- und ergänzungsfähig.

Das BMLO-Datenmodell weist in der Eingabemaske insgesamt 48 Datenfelder auf, die qualitativ zu unterscheiden sind. Sie sind in folgende Gruppen gebündelt:

- Gerüstdaten: Namen und Varianten (2 Felder), Lebensdaten (6), Zugehörigkeiten (6), Wirkungsdaten (6)
- Nachweise (6)
- Werke und Verzeichnisse (5)
- BMLO-Lexikonartikel (2)
- Bild- und Audiovisuelle Medien (7)
- Redaktionelle Vermerke (8).

Interdisziplinäre Edition und breite Rezeption

Das BMLO richtet sich an Interessierte verschiedener Fachrichtungen und wird deshalb an verschiedenen Stellen der Öffentlichkeit kostenfrei zugänglich gemacht: als selbständiges Lexikon der Universität München¹⁰ sowie durch die Einbindung in zwei Fachportale der Bayerischen Staatsbibliothek zur Landesgeschichte und zur Musikwissenschaft, nämlich die Bayerische Landesbibliothek Online¹¹ und die Virtuelle Fachbibliothek Musik der BSB.¹² Um das BMLO bekannt zu machen, wird ein umfangreiches Programm zur Öffentlichkeitsarbeit vor allem bei Multiplikatoren/innen, in Mailinglisten und in Internet-Fachportalen durchgeführt.

Ausblick

Beim Projektstart des BMLO im Januar 2004 sah die Planung eine Bearbeitungszeit von fünf Jahren vor. Um die langfristige Weiterentwicklung und nachhaltige Betreuung des BMLO sicherzustellen, ist für die Folgezeit

eine kooperative Fortführung auf drei Schultern vereinbart: Die Gesellschaft für Bayerische Musikgeschichte übernimmt die BMLO-Redaktion und leistet damit die redaktionelle Betreuung, die IT-Gruppe Geisteswissenschaften der Universität München die Pflege der Hardware und der Datenbank sowie die BSB die Langzeitarchivierung der Daten. Die gegenwärtig noch unvollständigen Datensätze von 15.600 Personen sollen bis 2008 bis zur Editionsreife geführt, medial angereichert und auf deutlich über 20.000 vermehrt werden.

Dr. Josef Focht leitet seit Anfang 2004 das DFG-Projekt „Bayerisches Musiker-Lexikon Online“ am Institut für Musikwissenschaft der Universität München. Arbeitsschwerpunkte: Themen der Instrumentenkunde, der historischen Aufführungspraxis, der Regionalmusikgeschichte und der Biographieforschung. E-Mail: focht@lrz.uni-muenchen.de

¹⁰ Vgl. <<http://www.bmlo.uni-muenchen.de>>.

¹¹ Vgl. <<http://www.bayerische-landesbibliothek-online.de>>.

¹² Vgl. <<http://www.vifamusik.de>>.