

Der Stammbaum aus dem Computer / Ahnendatenbank Niederbayern

Warum eine neue Datenbank?

Kennen Sie das auch? Sie sitzen im Archiv und schreiben kaum leserliche Daten aus einem Kirchenbuch ab, und plötzlich überkommt Sie der Gedanke: „Wie viele Leute werden vor mir schon die gleichen Daten aus dem Buch herausgeschrieben haben?“ Anfang 2005 beschlossen Wolfgang Gerauer, Bezirksvorsitzender von Niederbayern, und Siegfried Nyssen, diesem Gedanken Taten folgen zu lassen. Die Ahnenlisten der Bezirksgruppe sollten in einer neuen Datenbank erfasst und vernetzt werden. Neben einem schnellen Zugriff auf die Ahnendaten würden außerdem viele „tote Punkte“ überwunden, die den einzelnen Ahnenforschern zu schaffen machen. Zuvor waren aber noch eine Reihe technischer Voraussetzungen zu klären.

Technische Voraussetzungen

Der Anforderungskatalog für die Datenbank wurde schnell sehr umfangreich: Die Eingabe und die Auswertung der Daten sollten benutzerfreundlich organisiert sein. Das Programm sollte bei mehr als 500.000 Personen noch stabil und schnell laufen. Das einfache Auffinden und Verschmelzen von doppelten Datensätzen sollte ebenso möglich sein wie geographische und statistische Auswertungen. Da die finanziellen Mittel für die Entwicklung einer entsprechenden Software nicht zur Verfügung standen, musste eine Auswahl aus bereits erhältlichen Programmen getroffen werden. Die Entscheidung fiel zugunsten des Programms „Personal Ancestry Files 5.2“ (PAF) der Mormonen. Obwohl auch Programme mit deutlich besserer Funktionalität geprüft wurden, überzeugte dieses Programm durch seine Stabilität bei großen Datenmengen, seiner Verarbeitungsgeschwindigkeit und der übersichtlichen Verschmelzungsfunktion. Als nächstes war die Arbeitsmethodik zu klären und festzulegen.

Arbeitsmethodik für die Datenbank

Grundsätzlich ist es leichter, Datensätze zu verschmelzen als sinnvoll zu trennen. Deshalb wurde die Datenbank in neun geographische Bereiche eingeteilt, entsprechend den neun niederbayerischen Landkreisen. Die Benennung der Personen erfolgt nach Geburtsnamen, wobei aber jeder Datensatz noch zwei weitere Namensfelder enthält: zum einen den phonetischen Namen = die Originalschreibweise im Dokument (z.B. „Thainerpaur“ statt „Donaubauer“), zum anderen den in Niederbayern früher sehr gebräuchlichen Hausnamen.

Des Weiteren sollte die Herkunft der Datensätze rückverfolgbar sein; es wurde deshalb ein kleines Programm entwickelt, das an jeden Personen-Datensatz in einer eingeschickten gedcom-Datei den Namen des Einsenders anhängt. Dadurch können z.B. bei doppelten Datensätzen zu einer Person, die unterschiedliche Angaben enthalten, die beiden Einsender zur Klärung rückgefragt werden.

Eine Mitgliederbefragung anlässlich der Bezirksversammlung im Mai 2006 ergab eine Reihe weiterer wertvoller Anregungen, die allerdings zum Großteil durch die beschränkten Ressourcen kurzfristig nicht in das Projekt einfließen konnten. Sie sollen zu einem späteren Zeitpunkt berücksichtigt werden.

Erfahrungen im ersten Jahr 2006

Eine Inventur der vorliegenden Datenquellen ergab, dass die überwiegende Anzahl nur auf Papier existiert, da deutlich weniger als die Hälfte der Mitglieder mit einem Ahnenprogramm arbeitet. Die geschätzten bereits erforschten 150.000 Personen müssen also durch Freiwillige nach und nach eingegeben werden. Von den bis Ende November 2006 erfassten 50.000 Personen stammen ca. 18.000 aus eingesandten gedcom-Dateien, die restlichen 32.000 wurden bisher von Hand eingegeben. Wir schätzen den Bestand an noch zu verarbeitenden Papierquellen auf weit über 100.000 Personen.

Eine weitere interessante Quelle sind Abschriften von kompletten Kirchenbüchern im Text- oder Tabellenformat. Um diese nutzen zu können, erarbeiten wir deshalb auch Techniken, um solche Dateien direkt in eine gedcom-Datei umzuwandeln. Dies wurde bereits mit zwei Kirchenbüchern erfolgreich praktiziert: aus zwei Tabellen mit jeweils 3.000 Personeneinträgen konnten gedcom-Dateien erstellt werden, die von unserem Ahnenprogramm problemlos eingelesen und verarbeitet wurden. An der Umwandlung von Textdateien wird noch gearbeitet. Bei diesem Thema arbeiten wir mit dem Lehrstuhl für Informatik einer bayerischen Universität zusammen.

Die automatische Umwandlung von altdeutscher Handschrift in eine gedcom-Datei ist aber leider nicht möglich; hier muss nach wie vor der Mensch seine Erfahrung mit alten Dokumenten einbringen.

Der Stammbaum aus dem Computer

Die Verschmelzung der bereits verarbeiteten Ahnenlisten hat bestätigt, dass durch dieses Gemeinschaftsprojekt viele Lücken geschlossen werden können. Dazu waren weder Archivbesuche noch sonstige Recherchen notwendig, da eine Ahnenliste oft die Antwort auf die Fragen einer anderen Ahnenliste enthält.

Die Möglichkeiten der Auswertung dieser Datenbank sind vielfältig: sie reichen von regionalen Wanderungsstudien über statistische Fragestellungen (Entwicklung von Kinderanzahl und -sterblichkeit, Lebensalter, Todesursachen) bis zu den Grundlagen für Ortsfamilienbücher.

Zugang zur Datenbank

Ab 2007 soll auf mehreren niederbayerischen Stammtischen [z.Z. Freyung, Passau, Regen, Rottal, Pfeffenhausen, Straubing] ein Laptop mit der jeweils aktuellen Datenbankversion zur Verfügung stehen, um die Teilnehmer bei ihren Recherchen zu unterstützen.

Des Weiteren soll in den nächsten Monaten ein Programm erarbeitet werden, das es erlaubt, über Internet (<http://www.ahnenforschung-bayern.de/db/>) den aktuellen Füllstand der Datenbank einzusehen, z.B. über FOKO-Listen der einzelnen Gemeinden.

Dadurch soll Vereinsmitgliedern aber auch externen Forschern die Möglichkeit gegeben werden, über Nachname-Ort-Zeitraum den Inhalt der Datenbank zu prüfen. Im Erfolgsfall können dann Interessierte ihre Forschungsergebnisse als gedcom-Datei einschicken, die dann innerhalb unserer Bezirksgruppe auf Anschlussdaten überprüft werden. Der Einsender erhält im Gegenzug diese Anschlussdatensätze und eventuelle Korrekturen zurück.

Weitere bekannte Datenbank-Projekte

Diese Datenbank füllt eine Lücke zwischen den fünf anderen uns bekannten Ahnen-Datenbankprojekten: die Datenbanken des Bistumsarchivs Passau, des oberösterreichischen Landesarchivs in Linz (beide seit ca. 2004 im Aufbau), der Westdeutschen Gesellschaft für Familienforschung (seit 1998) sowie der Gesellschaft für Familienforschung in Franken (GFF) sind auf diese Regionen beschränkt. Jede dieser DB enthält bereits mehr als 1 Million Datensätze.

Die deutschlandweite Gedbas-DB des Vereins für Computergenealogie (erreicht in Kürze 5 Millionen erfasste Personen) hat nur wenige Verknüpfungen zwischen den einzelnen Datensätzen. Die Suche in dieser DB gestaltet sich erfahrungsgemäß sehr mühsam, da mehrfach vorhandene gleiche Personen nicht bereinigt werden, und die Ortsangaben oft nicht eindeutig sind oder sogar fehlen.

Im europäischen Raum sind noch folgende zwei DB erwähnenswert: der Staat Island hat bereits vor längerer Zeit alle Vorfahren der ca. 300.000 heute lebenden Isländer in einer staatlichen Datenbank erfasst, um die Erforschung genetischer Krankheiten zu unterstützen.

Die Niederlande haben ca. 1995 begonnen, alle Personenstandsunterlagen in den Rijkssarchiven zu erfassen. Z. Z. sind alle Register von 1800 -1900 erfasst (ca. 9 Millionen Personen) und im Internet suchbar (www.genlias.nl).

Zum Thema Datenschutz

Die Datenbank enthält Angaben zu Personen, die zwischen 1550 und 2000 geboren wurden. Dies erklärt sich dadurch, dass es Einsendern freigestellt ist, in den gedcom-Dateien die Angaben über lebende Personen und solche, die dem Datenschutz unterliegen, zu löschen (die meisten Programme bieten diese Option in der Export-Funktion) oder die Datei komplett mit allen Personen einzusenden. Da die Wahrscheinlichkeit, mit vollständigen Datensätzen Anschlussdaten zu finden, deutlich größer ist, erhalten wir überwiegend vollständige Daten.

Dies erfordert natürlich einen verantwortungsvollen Umgang mit der Datenbank, um den gesetzlichen Anforderungen gerecht zu werden. Neben technischen Voraussetzungen (z.B. Passwortschutz) hat nur ein namentlich festgelegter Personenkreis direkten Zugang zur Datenbank.

Weiteres Vorgehen

Unser Ziel ist es, bis Ende 2007 auf einen Datenbestand von 150.000 Personen zu kommen. Dazu ist es notwendig, mehr Freiwillige zu finden, die Ahnenlisten und andere Quellen in ein Programm eingeben wollen. Bis Sommer 2007 soll die Datenbank über Internet suchbar sein, z.B. in Form eines geographischen Zugangs mit FOKO-Listen.

Des Weiteren soll die Umstellung des Datenbank-Programms untersucht werden, da die Mormonen in diesem Jahr angekündigt haben, im Sommer 2007 ihr Programm PAF 5.2 durch eine ausschließlich Internet-gestützte Version PAF 6.0 zu ersetzen. In der näheren Auswahl sind open-source-Genealogieprogramme. Wir werden in regelmäßigen Abständen über den Fortschritt und die Erfahrungen berichten.

Siegfried NYSSSEN, Straubing, 12.11.2006

siegfried.nyssen@arcor.de