

# Die Elektronische Edition der Schriften Immanuel Kants

von Winfried Lenders, Bonn  
und Hans-Christian Schmitz, Frankfurt am Main

Die Universität Bonn verfügt über ein elektronisches Korpus von Immanuel Kants gesammelten Schriften gemäß den Abteilungen 1–3 der Akademie-Ausgabe. Dieses Korpus bildet die Grundlage einer elektronischen Edition der Schriften Kants, auf die über die Webseite des ehemaligen<sup>1</sup> Instituts für Kommunikationsforschung und Phonetik zugegriffen werden kann: <http://www.ikp.uni-bonn.de/kant/>. Im vorliegenden Artikel wird über den Umfang und den Zustand des Bonner Korpus und der elektronischen Edition berichtet.

## 1. Das Bonner Kant-Korpus

Seit den 1960er Jahren wurden an der Universität Bonn im Rahmen der Arbeit an einem Allgemeinen Kant-Index die Werke, der Briefwechsel und der handschriftliche Nachlass Kants systematisch digitalisiert. Die Datenerfassung wurde mehrfach unterbrochen und zog sich über Jahrzehnte hin.<sup>2</sup> Es kamen verschiedene Techniken zum Einsatz – manuelles Stanzen von Lochkarten, Eintippen mit der Tastatur und Erfassung mittels *Optical Character Recognition* (OCR) –, außerdem wurden die Daten auf verschiedene Weisen mit Metainformation annotiert. Das Ergebnis war ein Korpus, das alle Kant-Schriften aus den Abteilungen 1–3, damit aus den Bänden 1–23 der Akademie-Ausgabe, umfasste, das allerdings nicht einheitlich kodiert und teilweise mit Fehlern belastet war. Im Zuge eines von der DFG geförderten Projekts zur Bereitstellung und Pflege der Schriften Immanuel Kants in elektronischer Form wurde das elektronische Korpus in den letzten Jahren ergänzt, weitgehend korrigiert und in ein standardisiertes Format überführt. Im Zuge dieser Bearbeitung zeigte sich, dass ein solches elektronisches Korpus mit der heutigen Informationstechnologie in einer Weise auswertbar ist, wie sie in den 1960er und 1970er Jahren noch nicht absehbar war.

<sup>1</sup> Das Institut für Kommunikationsforschung und Phonetik ist 2005 im neu gegründeten Institut für Kommunikationswissenschaften der Universität Bonn aufgegangen.

<sup>2</sup> Zur Geschichte des Bonner Kant-Korpus vgl. Lenders (1982), Lenders (2000) und Lenders: *Zur elektronischen Edition der Werke Immanuel Kants* in Schmitz (Hg.) (2003).

## 1.1 Der Umfang des Korpus

### 1.1.1 Werke und Briefwechsel

Das Korpus enthält alle von Kant verfassten Werke aus den Bänden 1–9 (erste Abteilung der Akademie-Ausgabe). Es enthält außerdem den kompletten Briefwechsel gemäß der zweiten Auflage der Bände 10–13 (zweite Abteilung), darunter sowohl die von Kant verfassten als auch die an Kant gerichteten Briefe. Der Band 13 besteht bis auf einen Brief ausschließlich aus Herausgeberanmerkungen zu den Briefen der Bände 10–12. Diese Herausgeberanmerkungen sind im elektronischen Korpus nur auszugsweise vorhanden. Herausgeberanmerkungen aus den Werk-Bänden 1–9 sind überhaupt nicht in das Korpus aufgenommen worden.

### 1.1.2 Handschriftlicher Nachlass

Der in der dritten Abteilung der Akademie-Ausgabe enthaltene handschriftliche Nachlass (Bände 14–23) gliedert sich in die von Erich Adickes und seinem Nachfolger Friedrich Berger herausgegebenen Bände 14–19 und die von Gerhard Lehmann herausgegebenen Bände 20–23. Die von Adickes und Berger herausgegebenen Schriften sind in 8141 Einzelnotizen aufgeteilt.<sup>3</sup> Jede Notiz beginnt mit einem Notizkopf, in dem die jeweilige Notiznummer, eine oder mehrere mögliche Datierungen, die Fundstelle und gegebenenfalls weitere Anmerkungen stehen. Zum elektronischen Korpus gehören alle Notizen mit den Notizköpfen und den Herausgeberanmerkungen, welche innerhalb der Notizen stehen. Es handelt sich dabei sehr häufig um Angaben darüber, wie die einzelnen Textstellen der Notizen in der Original-Handschrift angeordnet sind. Beispielsweise werden in den Reflexionen zur Mathematik aus Band 14 (Seiten 1–61) die relativen Positionen von Formeln und Skizzen durch den Herausgeber spezifiziert. Das Korpus enthält – mit wenigen Ausnahmen aus den Bänden 17 und 18 – nicht die Herausgeberanmerkungen, die als Fußnoten in den gedruckten Bänden erscheinen. Zum Korpus gehören darüber hinaus alle Schriften Kants aus den von Lehmann herausgegebenen Nachlassbänden 20–23, einschließlich der Herausgeberanmerkungen, die in diese Texte integriert sind. Ferner sind die Fußnoten des Herausgebers aus den Bänden 20–22 vollständig und die Fußnoten aus Band 23 zu einem kleinen Teil vorhanden. Auch die in der Akademie-Ausgabe abgedruckten Skizzen und Bilder wurden in das elektronische Korpus aufgenommen.

<sup>3</sup> Die Notizen sind von 1 bis 8112 durchnummeriert. Der Unterschied zwischen der Anzahl der Notizen – 8141 – und der letzten Notiznummer – 8112 – kommt dadurch zustande, dass einzelne Notiznummern mehrfach vergeben und durch einen Buchstaben ergänzt wurden. Beispiele sind die Notizen 158 und 158a in Band 15.

## 1.1.3 Bezugstexte

Kant verfügte in seiner Bibliothek über Ausgaben eigener Werke und von Werken fremder Autoren, in denen er handschriftliche Notizen hinterlassen hat. Die Notizen beziehen sich zum Teil unmittelbar auf den Text, neben den sie geschrieben wurden. Die entsprechenden Bezugstexte sind zu einem großen Teil in der Akademie-Ausgabe abgedruckt. Kants eigene Werke, in denen er Notizen hinterlassen hat, sind in der ersten Abteilung der Akademie-Ausgabe vorhanden und sind damit Teil des elektronischen Korpus. Es handelt sich dabei um die *Beobachtungen über das Gefühl des Schönen und Erhabenen* (1764), die *Kritik der reinen Vernunft* (1781) und die *Kritik der praktischen Vernunft* (1788). Die Bezugstexte fremder Autoren sind zu einem großen Teil in den Bänden der dritten Abteilung abgedruckt. Sie wurden digitalisiert und in das elektronische Korpus aufgenommen. Es handelt sich dabei um die folgenden Texte:

- Gottfried Achenwall, *Juris naturalis pars posterior complectens jus familiae, jus publicum et jus gentium*. Editio quinta emendatior, Gotingae 1763.
- Alexander Gottlieb Baumgarten, *Metaphysica*, Ed. IIII, Halae Magdeburgicae 1757.
- Alexander Gottlieb Baumgarten, *Initia Philosophiae Practicae*, Halae Magdeburgicae 1760.
- Friedrich Bouterwek, Rezension von Kants *Metaphysische Anfangsgründe der Rechtslehre*, Göttingische Anzeigen von gelehrten Sachen unter der Aufsicht der Königlichen Gesellschaft der Wissenschaften 18. 02. 1797.
- Johann August Eberhard, *Vorbereitung zur natürlichen Theologie zum Gebrauch akademischer Vorlesungen*, Halle 1781.
- Johann Gottlieb Kreutzfeld, *Dissertatio Philologico-Poetica de Principiis Fictionum Generalioribus*, zweite Dissertation vor Antritt zur Professur der Dichtkunst an der Universität Königsberg, verteidigt am 28. 02. 1777.
- George Friedrich Meier, *Auszug aus der Vernunftlehre*, Halle 1752.
- Johannes Schulz, Rezension von Johann August Eberhard, *Philosophisches Magazin II. Bd.*, Jenaer Litteraturzeitung 1790, 281–284.
- Schüleraufsatz, unbetitelt.

Alle Bezugstexte außer dem von Gottfried Achenwall sind vollständig in der Akademie-Ausgabe abgedruckt. Sie wurden digitalisiert, indem sie aus der Akademie-Ausgabe eingescannt und mittels des OCR-Programms Abbyy Fine Reader 7.0 zeichenweise erkannt wurden. Die Akademie-Ausgabe diente als Vorlage, weil in ihr die Bezugstexte mit automatisch gut erkennbaren Antiqua-Lettern gedruckt sind. In ihren Originalausgaben sind die Bezugstexte in Frakturschrift gedruckt, die sich mit dem verwendeten OCR-Programm nicht verlässlich erkennen ließ. Der Achenwall-Text ist in der Akademie-Ausgabe nur auszugsweise abgedruckt; er konnte deshalb auch nur auszugsweise in das Korpus aufgenommen werden.<sup>4</sup>

<sup>4</sup> Mittlerweile gibt es mit dem Abbyy Fine Reader 9 (vgl. <http://www.frakturschrift.de>), ein

Das Bonner Kant-Korpus enthält also alle Texte und Bilder aus den Bänden 1–23 der Akademie-Ausgabe außer den als Fußnoten abgedruckten Herausgeberanmerkungen. Die Fußnoten sind nur zu einem kleinen Teil im Korpus vorhanden.

## 1.2 Die Kodierung der Daten

Die Texte des elektronischen Korpus sind zeichengetreu kodiert worden. Das bedeutet beispielsweise, dass ein ‚ß‘ als „ß“ und nicht etwas als „ss“ kodiert ist. Umlaute wie ‚ä‘ erscheinen als einzelne Zeichen („ä“) und nicht etwa als Zeichenketten wie „ae“. Griechische Textstellen sind in griechischen Lettern kodiert und nicht etwa ins Lateinische transliteriert worden.<sup>5</sup>

Die zeichengenau kodierten Rohtexte des Korpus sind mit Metainformation im XML-Format – ‚XML‘ steht für die *eXtensible Markup Language* – versehen worden. Durch die XML-Kodierung werden erstens Kant-Texte von Herausgeberanmerkungen und Texten fremder Autoren unterschieden. Bei den Notizen aus dem handschriftlichen Nachlass wird zwischen den eigentlichen von Kant verfassten Notizen und den von den Herausgebern hinzugefügten Notizköpfen unterschieden; die Notizköpfe wiederum enthalten jeweils eine Notiznummer, mögliche Datierungen, die Fundorte und sonstige Anmerkungen, die jeweils als solche ausgezeichnet sind. Zweitens werden Texteigenschaften, die sich vor allem im Druckbild manifestieren, gekennzeichnet: Im Korpus sind Überschriften, Paragraphen und Randtexte, weiterhin Seiten- und Zeilenwechsel und Besonderheiten im Drucksatz wie Kursiv- und Fettdruck, Wechsel der Schrifttype etc. markiert, ebenso Worttrennungen, auch wenn diese zumeist durch das Druckbild der Akademie-Ausgabe bedingt sind. Drittens wurden inhaltlich motivierte Metainformationen aufgenommen: Eigennamen sind als solche ausgezeichnet, und nicht-deutschsprachige Textstellen sind mit Angabe der jeweiligen Sprache versehen. Viertens und letztens wurden von Kant ausgelassene Wiederholungen aufgenommen: An einigen Stellen – zum Beispiel in Aufzählungen – kommt es vor, dass Kant Auslassungszeichen setzt, um dadurch bestimmte Wörter oder Phrasen nicht wiederholen zu müssen. Solche Auslassungen sind in den Korpus-Daten markiert und mit dem ausgelassenen Text versehen worden. (vgl. das Beispiel im Abschnitt 2.2 des vorliegenden Artikels.)

Eine genauere und ausführlichere Darstellung der Korpus-Kodierung geben Schmitz et al. (2006).

OCR-Programm, das Frakturschrift verlässlich erkennt. Mithilfe dieses Programms könnten die fehlenden Teile des Achenwall-Textes nachträglich digitalisiert werden.

<sup>5</sup> Die Korpus-Texte sind gemäß UTF-8 (*8-bit Unicode Transformation Format*) kodiert.

## 2. Die Verwendung des Korpus

Das Bonner Kant-Korpus kann in vielfacher Hinsicht verwendet und ausgewertet werden. Besonders interessant ist erstens die Möglichkeit, verschiedene Darstellungen oder ‚Sichten‘ des Materials zu erzeugen und über das *World Wide Web* (WWW) zugänglich zu machen. Texte und Textstellen, die sich aufeinander beziehen, wurden durch Hyperlinks verknüpft; dadurch wurden die Texte zu einem ersten umfassenden Hypertext verbunden, in den sich jederzeit auch neue Verknüpfungen, Fundstücke etc. integrieren lassen. Zweitens kann das Material nach vielfachen Gesichtspunkten durchsucht werden. Dazu wurden eine Suchmaschine installiert und ein Personenindex bereitgestellt. Schließlich können die Texte als Datenbasis für verschiedene Zwecke vorgehalten werden, so dass philologische Arbeiten am Werk Kants unterstützt werden können. Die Adresse der Webseite, über die die elektronische Kant-Edition gelesen und durchsucht werden kann, lautet <http://www.ikp.uni-bonn.de/kant>.

### 2.1 Verschiedene Darstellungen der Schriften Kants

Ein Zweck des Kant-Korpus besteht darin, die Korpustexte über das WWW zugänglich und lesbar zu machen. Dazu wurden mehrere Darstellungen der Texte erzeugt.

#### 2.1.1 Druckähnliche Darstellung der Bände 1–23 der Akademie-Ausgabe

Für die Bände 1–23 der Akademie-Ausgabe wurde eine druckähnliche elektronische Version erstellt, die über das WWW abgerufen werden kann. Dazu wurden die XML-kodierten Daten seiten- und zeilengetreu gemäß der Druckausgabe in die *HyperText Markup Language* (HTML) – die Kodierungssprache für Dokumente des WWW – übertragen. Für jede Druckseite, die Text enthält, der auch im elektronischen Korpus enthalten ist, wurde eine eigene HTML-Seite erzeugt; für leere Seiten und für Seiten, die ausschließlich Herausgeberanmerkungen enthalten und deren Inhalt nicht Teil des elektronischen Korpus ist, wurde keine HTML-Seite erzeugt.

Auch die Inhaltsverzeichnisse der Bände 1–23 wurden nach HTML transferiert; darüber hinaus wurde ein Gesamtverzeichnis für alle Bände erstellt. Das Gesamtverzeichnis ist durch Hyperlinks mit den Inhaltsverzeichnissen der Einzelbände und deren Inhaltsverzeichnisse mit den Textseiten der jeweiligen Bände verknüpft worden, so dass über das Gesamtverzeichnis und die einzelnen Inhaltsverzeichnisse auf die Textseiten zugegriffen werden kann. Jede Textseite ist außerdem mit der ihr vorhergehenden und der nachfolgenden Seite verknüpft, so dass gleichsam ‚geblättert‘ werden kann.

Die druckähnlichen HTML-Seiten sind als Tabellen aufgebaut. In der von links aus gesehen ersten Tabellenspalte stehen die Zeilennummern des nachfolgenden Textes, dahinter, in der zweiten Tabellenspalte, befindet sich der Text der Akademie-Ausgabe. Satztechnische Besonderheiten der Druckausgabe – wie Kursivdruck und so weiter – sind reproduziert; allerdings wird – anders als in der Druckausgabe – keine Frakturschrift verwendet. Außerdem werden Wörter, die in der gedruckten Ausgabe durch einen Zeilenumbruch getrennt sind, in der elektronischen Edition nicht getrennt; sie stehen jeweils vollständig in der Zeile, in der sie in der gedruckten Ausgabe beginnen.

Die dritte Tabellenspalte ist für Randtexte und für Verknüpfungen zu verwandten Textstellen oder externen Dokumenten vorgesehen. Zwei Beispiele: (a) In Band 7 befindet sich Kants Werk *Anthropologie in pragmatischer Hinsicht*, und in Band 15 befinden sich seine handschriftlichen Notizen zu diesem Werk. Die Notizen aus Band 15 sind in Anlehnung an das Inhaltsverzeichnis des Werkes aus Band 7 sortiert; sie können deshalb den Abschnitten des Werkes zugeordnet werden. Die Zuordnung wurde erreicht, indem die Notizen und die ihnen entsprechenden Abschnitte des Werkes wechselseitig durch Hyperlinks verknüpft wurden. (b) Handschriften und Briefe können mit ihren Faksimiles verknüpft werden. Für das *Opus postumum* (Bände 21 und 22) wurden solche Verknüpfungen hergestellt. Die Originalhandschrift des *Opus postumum* wurde von der Berlin-Brandenburgischen Akademie der Wissenschaften vollständig digitalisiert; die daraus resultierenden Bilder der einzelnen Manuskript-Seiten wurden im WWW veröffentlicht. Die druckähnliche HTML-Version des *Opus postumum* ist mit diesen Bildern verknüpft, so dass Nutzer des elektronischen Kant-Korpus ergänzend zu Gerhard Lehmanns Edition des *Opus postumum* die entsprechenden Seiten des Original-Manuskripts einsehen können.<sup>6</sup>

Fußnoten stehen an den Seitenenden, an denen sie auch in der gedruckten Ausgabe stehen. Es wurden nur die Fußnoten der Bände 20–22 berücksichtigt. Die Fußnoten der anderen Bände wurden nicht in die HTML-Edition aufgenommen, weil sie entweder nicht oder nur teilweise im elektronischen Korpus vorhanden sind. Die Bezugstexte aus den Nachlassbänden sind in gesonderten Dateien gespeichert; sie befinden sich – anders als in der gedruckten Ausgabe – nicht unterhalb der Kant-Texte. Druckseiten, die ausschließlich einen Bezugstext enthalten, wurden wie leere Seiten behandelt, d.h. für sie wurden keine druckähnlichen HTML-Seiten erstellt.

<sup>6</sup> An dieser Stelle sei der Berlin-Brandenburgischen Akademie der Wissenschaften für die Genehmigung, die digitalisierten Abbildungen der Manuskriptseiten mit den Bonner Daten zu verknüpfen, herzlich gedankt.

### 2.1.2 Alternative Anordnungen

Ein Vorteil einer elektronische Edition gegenüber einer gedruckten Edition besteht darin, dass sie bezüglich der Anordnung der Texte nicht festgelegt ist: Die Texte können parallel in verschiedenen Anordnungen präsentiert werden. Außerdem können neue Texte leicht integriert werden, ohne dass dadurch die Editionsprinzipien verletzt und die Übersichtlichkeit beeinträchtigt würden.

#### *Briefe*

Für Kants Briefwechsel wurden zwei verschiedene Darstellungen erzeugt. Erstens ist der Briefwechsel bereits ein Teil der oben beschriebenen druckähnlichen Darstellung. Darin entspricht eine HTML-Seite einer Druckseite, weshalb sich einerseits ein Brief über mehrere HTML-Seiten erstrecken kann und andererseits verschiedene Briefe auf derselben Seite anfangen oder enden können. Ergänzend zur druckähnlichen Darstellung wurde eine briefweise Darstellung erzeugt, indem – statt für jede Druckseite – für jeden Brief eine HTML-Seite erzeugt wurde. Der Zugang zu den Einzelbriefen erfolgt über ein eigens erstelltes chronologisch sortiertes Gesamtverzeichnis des Briefwechsels. Die briefweise Darstellung hat zwei Vorteile gegenüber der druckähnlichen Darstellung: Erstens ist sie übersichtlicher, denn sie fügt die Briefe der Bände 10–13 und 23 in chronologischer Ordnung zusammen. Zweitens ermöglicht sie es, neue bisher nicht in der Akademie-Ausgabe abgedruckte Briefe zu integrieren, ohne dazu nennenswerten Aufwand zu betreiben oder die Übersichtlichkeit der Edition zu beeinträchtigen.

#### *Rekonstruktion der handschriftlichen Notizen*

Die von Erich Adickes und Friedrich Berger herausgegebenen handschriftlichen Notizen Kants (Bände 14–19) sind in der gedruckten Akademie-Ausgabe thematisch angeordnet. Zusammen mit jeder Notiz ist ihre Fundstelle angegeben. Die Notizen können deshalb – ergänzend zur thematischen Anordnung – gemäß ihren Fundorten sortiert werden. Ein Beispiel: Kant verfügte über eine mit unbedruckten Seiten ‚durchschossene‘ Ausgabe von Alexander Gottlieb Baumgartens *Metaphysica*, in der er einen Großteil der in der Akademie-Ausgabe abgedruckten Notizen hinterlassen hat. Fast die Hälfte, nämlich 4022 der 8141 von Adickes und Berger herausgegebenen Notizen stammen aus der *Metaphysica*-Ausgabe. Diese Notizen sind über die Bände 14–18 verteilt; nur in Band 19 ist keine *Metaphysica*-Notiz enthalten. Die *Metaphysica*-Notizen wurden gemäß ihrem Auftreten sortiert: Notizen von Seite 1 der *Metaphysica*-Ausgabe wurden Notizen von Seite 2 vorgeordnet, Notizen von Seite 2 wurden wiederum Notizen von der Seite 2 gegenüberliegenden Durchschusseite vorgeordnet, und so weiter. Die Notizen wurden dadurch in eine an

ihren Fundorten orientierte Halbordnung gebracht. Es handelt sich nur um eine Halbordnung, weil Notizen, die Adickes zufolge auf derselben *Metaphysica*-Seite standen, untereinander nicht in eine durch die Fundstelle begründete Ordnung gebracht werden konnten; die Herausgeberanmerkungen darüber, wo auf der betreffenden Seite die jeweiligen Notizen standen, reichten für eine vollständige Ordnung nicht aus.

Im Anhang von Band 16 der Akademie-Ausgabe ist das Faksimile einer Durchschusseite aus Kants Handexemplar von Georg Friedrich Meyers *Auszug aus der Vernunftlehre* abgedruckt. Die Seite ist übersät mit teilweise einander überlappenden Notizen Kants. Laut Auskunft von Werner Stark handelt es sich dabei um eine in extremer Weise beschriebene Seite; nicht alle Notizseiten von Kant seien so unübersichtlich. Es scheint weder möglich noch sinnvoll zu sein, den Aufbau solch unübersichtlicher Seiten vollständig zu rekonstruieren. Es kann allenfalls sinnvoll sein, Bilder der Originalhandschriften bereitzustellen. Das bedeutet allerdings nicht, dass eine vollständige Rekonstruktion von Notizseiten generell unmöglich oder unsinnig ist. Wie eine mögliche Rekonstruktion von Notizen aussehen kann, konnte anhand eines Foliobogens mit einer Reflexion zur Medizin (Notiz 1526, Band 15, Seiten 939–953) gezeigt werden. Allein aufgrund der Herausgeberanmerkungen wurden die Textbausteine der Reflexion – Haupttext, durchgestrichene Passagen, Einschübe und Randtexte – auf den Seiten des Bogens positioniert.<sup>7</sup> Der Aufbau des Gesamttextes wird dadurch klar. Die Frage, ob eine solche Rekonstruktion der Rezeption des Textes überhaupt dienlich ist, ist damit natürlich noch nicht beantwortet.

Den von Adickes und Berger herausgegebenen Notizen sind neben ihren Fundorten auch mögliche Datierungen – das heißt Angaben darüber, in welchen Zeiträumen die Notizen gemacht worden sein können – beigegeben. Man könnte nun versuchen, die Notizen aufgrund dieser Angaben in eine chronologische Ordnung zu bringen und in dieser Ordnung zu präsentieren.<sup>8</sup> Es hat sich allerdings herausgestellt, dass eine solche Anordnung nicht praktikabel ist: Für die meisten der 8141 Notizen sind mehrere mögliche Datierungen angegeben, so dass die Anzahl der auf den alternativen Datierungen basierenden Anordnungen unüberschaubar groß wäre. Es ist zu vermuten, dass sie selbst dann übermäßig groß wäre, wenn man sich auf die Anordnung eines signifikanten Anteils der Notizen – zum Beispiel auf die Notizen aus der */Metaphysica/-Ausgabe* – beschränkte und nur die nach Adickes wahrscheinlichsten Datierungen berücksichtigte. Zur Zeit liegen jedoch keine Kriterien vor, die ausreichen, die Menge der möglichen Anordnungen auf ein überschaubares Maß zu reduzieren. Daher wurde auf die Darstellung der handschriftlichen Notizen in chronologischer Anordnung verzichtet.

<sup>7</sup> Die Rekonstruktion des Foliobogens wurde von Bernhard Fisseni durchgeführt.

<sup>8</sup> Vgl. Lenders (1982).



## 2.2 Durchsuchbarkeit

Eine weitere Möglichkeit, das elektronische Kant-Korpus zu benutzen, besteht darin, die Daten nach unterschiedlichen Kriterien zu durchsuchen.

### 2.2.1 Suchmaschine

Um die Kant-Texte durchsuchen zu können, wurde die frei verfügbare Suchmaschine *webglimpse*<sup>9</sup> installiert. Die Suchmaschine operiert auf einem Wortformenindex, der automatisch aus den druckähnlichen HTML-Dateien der Bände 1–23 erzeugt wird. Sie erlaubt die Verwendung komplexer Suchausdrücke und ist also funktionsmächtig. Die Suchmaschine ist selbst erklärend, einfach zu bedienen und funktioniert außerordentlich robust. Die Nutzer der Kant-Edition, von denen bisher Rückmeldungen vorliegen, äußerten sich zufrieden mit den vorhandenen Suchmöglichkeiten. Meist wurden einfache Suchanfragen nach zentralen Inhaltswörtern gestellt und – soweit wir das feststellen können – immer richtig und erschöpfend beantwortet.

Nichtsdestoweniger sollte die gegenwärtig installierte (kostenfreie) Suchmaschine mittelfristig durch eine andere Suchmaschine ersetzt werden, die auf dem XML-kodierten Korpus anstatt auf den HTML-kodierten Dateien operiert. Dies hätte folgende Vorteile: Erstens funktionieren derzeit nur Suchanfragen nach einzelnen Inhaltswörtern robust. Ganze Phrasen werden nur dann gefunden, wenn sie im Text nicht durch Zeilenwechsel unterbrochen werden, und Anfragen mit mehreren Suchausdrücken führen nicht zum gewünschten Ergebnis, wenn die Suchausdrücke auf verschiedenen Seiten stehen. Zweitens würden die Ergebnisse von Suchanfragen nach Phrasen oder komplexen Suchausdrücken deutlich verbessert, wenn das XML-kodierte Gesamtkorpus anstelle der einzelnen HTML-Dateien durchsucht werden könnte. Denn die XML-Daten sind umfangreicher als die HTML-Daten. Sehen wir uns als Beispiel die Seite 499 von Band 15 der Akademie-Ausgabe an. Dort stehen die folgenden Zeilen, die so auch in der HTML-Version der Seite erscheinen:

```
Der Wille nach instincten ist das temperament,  
-- -- -- Grundsätzen -- character.
```

Wenn man den Band 15 nach der Zeichenkette „Der Wille nach Grundsätzen ist character“ durchsucht, wird man kein Ergebnis erhalten, obwohl es wünschenswert wäre, die Seite 499 als Fundstelle zu bekommen, da der gesuchte Text dort wenigstens implizit vorhanden ist. Man erhielte die Seite als Ergebnis, wenn bei der Bearbeitung der Suchanfrage dem System bekannt wäre, durch welche Zeichenketten – nämlich „Der Wille nach“ und „ist“ – die Auslassungsstriche vor und nach „Grund-

<sup>9</sup> <http://webglimpse.net>

satzen“ zu ersetzen sind. Diese Information ist in den XML-Daten vorhanden, so dass eine Suche auf den XML-Daten erfolgreich wäre. Drittens wäre es, wenn die Suchmaschine auf den XML-Daten operierte, möglich, die Suche auf bestimmte XML-markierte Textpassagen einzugrenzen. So könnte beispielsweise die Suche nach einem Schlüsselwort auf das Vorkommen dieses Wortes in Überschriften beschränkt werden.

### 2.2.2 Lemmatisierung

Weder mit der installierten Suchmaschine, noch mit jeder anderen heute verfügbaren Suchmaschine kann allerdings ein Problem der Suche in Volltexten behoben werden, das schon in den 1960er Jahren als Hindernis automatischer Textrecherche erkannt wurde, nämlich das Problem der Lemmatisierung.<sup>10</sup> Es wäre wünschenswert, das Kant-Korpus nach Wörtern in verschiedenen Flexionsausprägungen und nicht nur nach spezifischen Wortformen durchsuchen zu können. Dazu müsste in eine Suchmaschine eine morphologische Analyse integriert werden. Zwar können für Suchanfragen mit *webglimpse* in beschränkter Form reguläre Ausdrücke gebildet werden, so dass man sich bei der Suche nicht eindeutig auf bestimmte Wortformen festlegen muss, die Vorzüge einer Lemmatisierung werden dadurch aber keineswegs ausgeglichen. Versuchsweise wurde von Gerd Willée mithilfe des von ihm geschriebenen Programms LEMMA4 eine lemmatisierte Form von Kants Schrift *Der Streit der Fakultäten* (Band 7 der Akademie-Ausgabe) erzeugt. Das Ergebnis ist ein Text, in dem in jeder Zeile je eine Wortform, ihre Grundform und ihre Wortklasse stehen; für Verben sind außerdem deren Flexionsausprägungen angegeben.<sup>11</sup> Derartige Angaben können leicht in die XML-Kodierung des Korpus integriert und dann für die Durchsuchung der Daten verwendet werden. Insgesamt ist die Lemmatisierung eine von der Umstellung der Suchmaschine unabhängige Aufgabe. Der beschriebene erste Versuch zeigt, dass sie zumindest für die deutschsprachigen Werke Kants mit einem aktuellen Lemmatisierer weitgehend automatisch erfolgen kann – die Morphologie des Deutschen hat sich in den letzten gut 200 Jahren nicht wesentlich geändert. Aber nicht alle Texte des Kant-Korpus sind in deutscher Sprache, und die Lemmatisierung der besonders interessanten Nachlassdaten ist schon deshalb problematisch, weil sie in beträchtlichem Ausmaß verschiedene Schreibweisen von Wörtern, Auslassungen von Wortendungen und Schreibfehler enthalten.

<sup>10</sup> Aus diesem Grund wurden ab Ende der 60er Jahre für zahlreiche literarische Werke lemmatisierte Indices unter großem Aufwand an manueller Bearbeitung erzeugt und publiziert. Auch für die Schriften Kants lagen Anfang der 1970er Jahre von Hand bearbeitete lemmatisierte Indices zu den Bänden 1 und 3 der Akademie-Ausgabe vor, die jedoch wegen überaus zahlreicher Probleme nicht publiziert worden sind. Vgl. dazu Lenders 1982, 440/441,

<sup>11</sup> Der lemmatisierte Text ist auf der Web-Seite der elektronischen Kant-Edition verknüpft (<http://www.ikp.uni-bonn.de/kant/lemmatisiert>); das Format des Textes wird von Willée und Elsen (2006) erläutert.

### 2.2.3 Personenindex

Die Benutzung einer Suchmaschine verlangt, dass der Nutzer weiß, wonach er sucht und dass er eine entsprechende Suchanfrage formulieren kann. Dabei können Listen (Indices) der in einem Datenmaterial enthaltenen Suchwörter außerordentlich hilfreich sein. Im Falle des Kant-Korpus operiert die installierte Suchmaschine intern auf einem Wortformenindex (s. o.). Um dem Benutzer eine Übersicht über die im Wortmaterial vorhandenen Personennamen zu geben, wurde die elektronische Edition um einen Personenindex erweitert.

Dazu wurde der von Katharina Holger, Eduard Gerresheim, Antje Lange und Jürgen Goetze bearbeitete und im Jahr 1969 bei de Gruyter erschienene Personenindex 1. Stufe zu Kants gesammelten Schriften digitalisiert, in HTML kodiert und über das WWW verfügbar gemacht. Bei dem Personenindex 1. Stufe handelt es sich um einen Index aller Personen, die Kant – nach Ausweis der Akademie-Ausgabe – erwähnt hat. Die Personennamen sind alphabetisch geordnet; mit ihnen sind alle Stellen, an denen die jeweiligen Personen genannt werden, angegeben. Die Stellenangaben sind wiederum durch Hyperlinks mit den entsprechenden Stellen in der druckähnlichen HTML-Version des Korpus verknüpft. Mithilfe des Personenindexes können zum einen alle Stellen, an denen Kant eine Person erwähnt hat, gefunden werden; zum anderen gibt der Personenindex einen Überblick darüber, wen Kant überhaupt erwähnt hat.

## 2.3 Bereitstellung und Sammlung von Daten

Das elektronische Kant-Korpus kann als zentrale Datenbasis für die Kantforschung dienen. Bisher schon wurden auf individuelle Anfragen hin die Texte in verschiedenen Formaten zur Verfügung gestellt, unter anderem für die Neuedition der Werke und der Vorlesungen. Im Gegenzug sollte das Bonner Korpus sukzessive ergänzt und aktualisiert werden. Insbesondere sollten Korrekturen und Ergänzungen, die im Zuge von Neueditionen anfallen, eingefügt werden. Ferner sollten Briefe und sonstige Fundstücke, die nicht in die gedruckte Akademie-Ausgabe aufgenommen wurden, hinzugefügt werden. Schließlich können weitere Faksimiles von Briefen und sonstigen Handschriften eingefügt werden.

Von besonderem Interesse ist auch die genauere Untersuchung und Bewertung von Adickes' Datierungen der handschriftlichen Reflexionen Kants aus den Bänden 14–19. Da das Bonner Korpus die entsprechenden Markierungen enthält, kann man sich aus Experimenten mit verschiedenen Anordnungen und aus statistischen Auswertungen interessante Ergebnisse erhoffen.<sup>12</sup>

<sup>12</sup> Möglichkeiten und Verfahren statistisch basierter Untersuchungen werden von Manning und Schütze (1999) erläutert. Wir verweisen speziell auf das Kapitel über Textklassifikation (*Text Categorization*).

### 3. Fazit und Perspektiven für die Weiterarbeit

Die hier beschriebene Fassung des Bonner Kant-Korpus wurde in den letzten Jahren im Rahmen des von der Deutschen Forschungsgemeinschaft geförderten Projekts *Bereitstellung und Pflege der Schriften Immanuel Kants in elektronischer Form* erarbeitet und in dem standardisierten XML-Format bereit gestellt. Ausgehend vom Korpus wurde eine elektronische Edition der Schriften Kants hergestellt und über eine WWW-Schnittstelle allgemein und kostenlos zugänglich gemacht. Die elektronischen Kant-Schriften können nun *online* gelesen und durchsucht werden, ein Angebot, von dem intensiv Gebrauch gemacht wird. Darüber hinaus kann das elektronische Korpus die interpretatorische Arbeit am Werk Immanuel Kants unterstützen, indem es verschiedene Anordnungen des Textmaterials, insbesondere des Nachlasses, erlaubt.

Allerdings ist das Korpus nur so gut, wie seine Textgrundlage, die Akademie-Ausgabe. Da die editorischen Arbeiten an dieser Ausgabe keineswegs abgeschlossen sind (vgl. das Vorwort von Reinhard Brandt zum Sonderheft der Kantstudien „Zustand und Zukunft der Akademie-Ausgabe von Immanuel Kants Gesammelten Schriften“ sowie die Beiträge in diesem Sonderheft), sollte die Bonner elektronische Edition idealerweise zusammen mit der Akademie-Ausgabe weiterentwickelt werden. Dementsprechend wurden bereits Daten für Neueditionen im Rahmen der Akademie-Ausgabe bereitgestellt, und es wurden Absprachen über den Rückfluss von Daten – insbesondere von Korrekturen und Änderungen, die im Zuge der Neueditionen anfallen – getroffen.<sup>13</sup> Das Kant-Korpus und die elektronische Edition sind mindestens solange nicht fertig, als auch die Akademie-Ausgabe nicht fertig ist.

Weiterhin kann die elektronische Edition optimiert werden. Zum Beispiel kann – wie oben bereits angemerkt – die Suchfunktion verbessert werden. Ferner muss davon ausgegangen werden, dass das Korpus und die elektronische Edition nicht vollständig fehlerfrei sind. Bezüglich kritischer Stellen muss sich der Nutzer also anhand der gedruckten Ausgabe absichern. Sobald Fehler erkannt werden, müssen diese korrigiert werden. Eine vollständig korrigierte Version des Korpus könnte schließlich als Datengrundlage für eine neu zu erzeugende gedruckte Edition der Schriften Kants dienen.

Ein weiteres Desideratum betrifft den Personenindex. Denn außer dem Personenindex 1. Stufe existiert der von Katharina Holger und Eduard Gerresheim bearbeitete Personenindex 2. Stufe, allerdings bisher nur in einer als Manuskript vervielfältigten Fassung von 26 Einzelheften. In diesem Register sind alle Stellen, an denen Kant eine Person erwähnt hat, mit den jeweiligen zitierten Stellen aus den Originalwerken und bio- sowie bibliographische Angaben zu den zitierten Autoren aufge-

<sup>13</sup> Ein Beispiel: Kants *Kritiken* werden derzeit an der Berlin-Brandenburgischen Akademie der Wissenschaften neu ediert. Im Rahmen der Bonner Arbeiten wurden dafür die entsprechenden elektronischen Texte bereitgestellt. Die Texte müssen nun nicht mehr neu eingegeben werden, statt dessen können sich die Editoren auf die Korrektur der vorliegenden Texte konzentrieren. Die Arbeitersparnis ist zweifellos beträchtlich.

führt. Mit Eduard Gerresheim wurden die Digitalisierung des Personenindex 2. Stufe und seine Aufnahme in die elektronische Edition bereits vereinbart. Wenn auch die entsprechenden Arbeiten noch nicht durchgeführt werden konnten, so zeigten doch immerhin einige Experimente, dass die zusätzlichen Angaben des Personenindex 2. Stufe problemlos in die elektronische Edition integriert werden können.

Schließlich wäre es attraktiv, die nicht-deutschen Texte der Kant-Edition – das sind vor allem Kants lateinische Schriften sowie die lateinischen Bezugstexte, beispielsweise Baumgartens */Metaphysica/*, durch deutsche Übersetzungen zu ergänzen. Die Rezeption dieser Schriften würde dadurch zweifellos gefördert.

#### *Literatur*

- Bereitstellung und Pflege der Werke Immanuel Kants in elektronischer Form.* Hrsg. von Hans-Christian Schmitz. In: *IKP-Arbeitsbericht* NF 03. Universität Bonn 2003. <http://www.ikp.uni-bonn.de/ikpab/ikpab-nf03.pdf>
- Lenders, Winfried: Der allgemeine Kantindex. Vom Stellenindex zum Informationssystem. In: *Kant-Studien* 73, 1982, 440–451.
- Lenders, Winfried: Kants Gesammelte Werke in elektronischer Form. In: Brandt/Stark, 148–159.
- Manning, Christopher D. und Schütze, Hinrich: *Foundations of Statistical Natural Language Processing*. Cambridge 1999.
- Schmitz, Hans-Christian/Diel, Marcel/Keul, Gert/Lüert, Simon und Lenders, Winfried: Elektronische Kanttexte. In: *IKP-Arbeitsbericht* NF 18. Universität Bonn 2006. <http://www.ikp.uni-bonn.de/ikpab/ikpab-nf18.pdf>
- Willée, Gerd und Elsen, Harald: Funktionsbeschreibung von LEMMA4, einem Wortklassentagger und Lemmatisierer für unrestringierte deutsche Texte. *IKP-Arbeitsbericht* NF 16. Universität Bonn 2006. <http://www.ikp.uni-bonn.de/ikpab/ikpab-nf16.pdf>
- Zustand und Zukunft der Akademie-Ausgabe von Immanuel Kants Gesammelten Schriften.* Hrsg. von Reinhard Brandt und Werner Stark. Sonderheft der *Kant-Studien* 91. Berlin 2000.