

## Zentralität und Sichtbarkeit.<sup>1</sup>

### Mathematik als Hierarchisierungsinstrument am Beispiel der frühen Bibliometrie.

#### 1. Einleitung

Seitdem Computer zielstrebig in jede Pore gegenwärtiger Gesellschaften eindringen, wächst das Interesse am systematisierten Umgang mit Information. Die Stichwortsuche im *World Wide Web* (WWW), die Verwaltung von Bookmarks und Dokumenten oder das Navigieren in Online-Enzyklopädien, Datenbanken und Webshops haben Praktiken der computergestützten Auffindung und Organisation von Information zu alltäglichen Aufgaben und die damit verbundenen Kompetenzen zu elementaren Kulturtechniken werden lassen. Und obwohl Bibliothekskataloge, Nachschlagewerke und Archive aller Art seit Langem ähnliche Problematiken mit sich bringen, hat wohl erst das Internet für die notwendige Erfahrungsdichte gesorgt, um eine breite gesellschaftliche Sichtbarkeit zu bewirken.

Diese neue Aufmerksamkeit lässt sich nicht nur für die breite Öffentlichkeit diagnostizieren, sondern ebenso für die Wissenschaft, wo verschiedene Disziplinen den technischen und konzeptionellen Umgang mit Information zum Thema gemacht haben. Der klassische Untersuchungsgegenstand der Bibliotheks- und später Informationswissenschaft – die sich im deutschen Sprachraum allerdings nie ganz als eigenständige universitäre Disziplinen etablieren konnten – findet sich immer öfter im Fokus der Medienwissenschaft, der Geschichtsforschung oder der Philosophie. Dabei ist bemerkenswert, dass die historisch-konzeptionelle Lesekopf in den Geistes-, Kultur- und Sozialwissenschaften in der Regel entweder bis hin zur »Erfindung« des Computers oder aber, wie etwa im Fall der in den letzten Jahren sehr aktiven *Science and Technology Studies*, in der Gegenwart angesetzt wird. Mit wenigen Ausnahmen beginnt auch für die »interne« Historiographie der Informatik und der Informationswissenschaft die Geschichte der digitalen Informationsverwaltung erst in den späten 60er und 70er Jahren, wobei sich das Interesse zumeist auf konkret technische Objekte, wie etwa Speichermedien, Programmiersprachen, Datenbanken oder Netzwerke, richtet; abstraktere konzeptionelle Fragen konzentrieren sich auf den Bereich mathematischer Grundüberlegungen wie der Berechenbarkeitstheorie oder der Informationstheorie. Ein Loch klafft zwischen der Phase der frühen Systematisierung und Mechanisierung des Umgangs mit Information, die bis hin zu den ersten Computern reicht, und der Gegenwart, deren Beginn meist mit der Erfindung des Mikroprozessors und der damit verbundenen explosionsartigen Verbreitung digitaler Technologie angesetzt wird.

Dabei sind aber gerade die 50er, 60er und 70er Jahre zentral für die Entwicklung und

---

<sup>1</sup> Erschienen als B. Rieder (2012). Zentralität und Sichtbarkeit. Mathematik als Hierarchisierungsinstrument am Beispiel der frühen Bibliometrie. In T. Brandstetter, T. Huebel & A. Tantner (Hg.), Vor Google. Eine Mediengeschichte der Suchmaschine im analogen Zeitalter. Bielefeld: Transcript, pp.223-251.

Stabilisierung vieler grundlegender Methoden und Paradigmen, welche die Informatik und Informationswissenschaft bis heute dominieren. Denn mit der geglückten Konstruktion einer universellen symbolverarbeitenden Maschine<sup>2</sup> war es keineswegs getan. In den Jahrzehnten nach 1945 lässt sich deshalb nicht nur die Transformation von fehleranfälligen technischen Experimenten in relativ verlässliche Industrieprodukte beobachten, sondern vor allem die Entstehung und Emanzipation einer Softwareebene, die immer unabhängiger von den Widrigkeiten der Hardware ihren eigenen Gegenstandsbereich definiert. Während die 30er und 40er Jahre von Fragen der prinzipiellen Berechenbarkeit symbolischer Systeme und ihrer mechanischen Umsetzung dominiert waren, verschiebt sich nun das Interesse auf die schlussendlich alles andere als banale Frage, was sich denn nun alles mit diesen Maschinen machen ließe und wie dabei am besten vorzugehen sei. Die neuen Programmiersprachen der »dritten Generation«, wie FORTRAN oder COBOL, erleichtern das Programmieren außerordentlich und eröffnen einen Möglichkeitsraum, in dem der Computer viel mehr sein wird als eine schnelle Rechenmaschine. Für den Bereich der Informationsverwaltung ist diese Periode ebenfalls von hoher Wichtigkeit, weil es nun nicht mehr genügt, manuelle Verfahren zu mechanisieren – womit sich Firmen wie IBM ein lukratives Geschäft aufgebaut hatten –, sondern die Frage virulent wird, wie spezifisch für das Computerzeitalter konzipierte Konzepte und Methoden aussehen könnten. Dabei geht es zwar zunächst immer noch um Lochkarten, Mikrofilme und Sortiertechniken, analog zur Informatik verschiebt sich aber auch hier das Interesse zu konzeptionellen Fragenstellungen, zu Fragen der *Software* eben. In dieser Scharnierzeit treffen die Denkweisen des analogen Zeitalters auf eine Technologie, die sich zwar aus einer klassisch-industriellen Maschinenlogik entwickelt, sich aber dennoch fundamental von dieser unterscheidet.

Im folgendem Text soll gezeigt werden, wie im Bereich der frühen Bibliometrie, also der Anwendung von Zählverfahren auf Publikationen aller Art, eine Vorstellung vom Umgang mit Information entsteht, die dem Wechsel der technischen Möglichkeiten ein konzeptionelles Pendant zur Seite stellt. Es handelt sich dabei um die *Zitationsanalyse*, ein Verfahren, das heute nicht nur die Messung wissenschaftlicher Produktivität dominiert, sondern auch in modernen Suchmaschinen wie *Google* oder *Bing* mitbestimmt, welche Seite an erster und welche an letzter Stelle einer Ergebnisliste steht. Allgemeiner gesprochen geht es um die Anwendung von Mathematik als Hierarchisierungsinstrument, um die Entwicklung einer *Produktionsform* für Ordnung, die sich anschickt, das junge 21. Jahrhundert maßgeblich mitzugestalten.

## **2. Assoziation und Klassifikation**

Die relevanten Themen der Gegenwart strukturieren unseren Blick auf die Vergangenheit.

---

<sup>2</sup> Computer sind universelle Turing-Maschinen, die jede beliebige korrekt formulierte Befehlskette ausführen können. Als direkte Konsequenz sind sie ohne Software unbrauchbar, können aber eben auch Millionen unterschiedlicher Programme, von der Textverarbeitung bis zum Videospiel, mit der gleichen Hardware ausführen.

Das wachsende Interesse für die systematische und v.a. digitale Handhabung von Information ist, wie schon erwähnt, dem sich seit Mitte der 90er Jahre explosionsartig verbreitenden Internet zuzuschreiben. Da die gesellschaftliche Wahrnehmung des Internets von einer spezifischen Anwendung – dem WWW – beherrscht wird, ist es nicht weiter verwunderlich, dass dem Konzept des *Hypertexts* große Aufmerksamkeit geschenkt wird: bis heute ist die historische Herleitung der digitalen »Medien« geprägt durch die Fahndung nach den Vorläufern des *Hyperlinks*. Ein spezifischer Text, Vannevar Bushs *As We May Think*<sup>3</sup>, spielt dabei in der Regel eine wesentliche Rolle<sup>4</sup>, und diese privilegierte Position zwingt uns gewissermaßen, unser Argument an dieser Stelle beginnen zu lassen.

## 2.1. Der *Memex*

In den Untersuchungen zur Vorgeschichte des Internets wird kein Text so oft genannt wie der 1945 in *The Atlantic Monthly* erschienene Essay, der dem amerikanischen Ingenieur und Technokraten den Titel des »Erfinders«<sup>5</sup> von *Hyperlink* und *Hypertext* einbringen sollte – auch wenn diese Begriffe erst 1967 von Ted Nelson eingeführt werden. Während seiner Tätigkeit als Wissenschaftsadministrator – u.a. als Direktor des *Office of Scientific Research and Development*, welches bis 1943 das *Manhattan Project* betreute – diagnostiziert Bush, dass es für Wissenschaftler angesichts der hohen Publikationsdichte kaum mehr möglich sei, sich Überblick über mehr als ein Spezialfeld zu verschaffen, weswegen fortwährend wichtige Entdeckungen in der Masse untergehen würden. Für den Pionier im Bereich der analogen Computer schien eine technische Lösung für die Probleme der modernen Wissenschaft Erfolg versprechend, und der zitierte Artikel beschreibt eine – nie gebaute – Maschine, den *Memory Extender*, kurz *Memex*. Dabei handelte es sich um eine Art Bibliotheksterminal mit Funktionen zur Annotation und Verknüpfung von auf Mikrofilm gespeicherten Quellen. Wissenschaftler könnten Verbindungsketten – sogenannte »trails of association« – von einem Dokument zum nächsten legen und diese Ketten untereinander austauschen; die entstehenden Textnetzwerke würden die »Künstlichkeit« bestehender Methoden der Indizierung durchbrechen und zu einer »natürlichen«, d.h. der Funktionsweise des Gehirns entsprechenden, Form der Informationsorganisation und -suche führen. Dabei ist es wohl nicht die technisch-konzeptionelle Dimension von Bushs Idee, der zentrale Wichtigkeit zukommt:

»[Bush] is best known in the field of information retrieval, even though his systems hardly worked, his ideas were not new, he did not really understand what he was talking

---

<sup>3</sup> Bush, Vannevar: »As We May Think«, in: *The Atlantic Monthly*, Juli 1945.

<sup>4</sup> Vgl. Nyce, James M./Kahn, Paul (Hg.): *Vannevar Bush and the Mind's Machine*, Boston: Academic Press 1992. oder Buckland, Michael: »Overview of the History of Science Information Systems«, in: Mary Ellen Bowden/Trudi Bellardo Hahn/Robert V. Williams (Hg.), *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems*, Medford, NJ: Information Today 1999, S. 3-7.

<sup>5</sup> Zachary, G. Pascal: »The Godfather«, in: *Wired*, Nov 1997.

about, and he chose not to acknowledge the priority of others.«<sup>6</sup>

Diese Einschätzung von einem der wenigen informationswissenschaftlich gebildeten Historiker und dem ehemaligen Präsidenten der *American Society for Information Science & Technology* (ASIS&T) muss im Lichte der mythischen Stellung des *Memex* und seines Erfinders gesehen werden, die in vielerlei Hinsicht wichtigere Beiträge überschattet. Und tatsächlich, der *Memex* ist keine Maschine zur Informationssuche, sondern ein Werkzeug zur Unterstützung einer Aktivität, die wir heute wohl »persönliches Informationsmanagement« nennen würden. Was in kleinem Rahmen ein nützliches und durch die Mikrofilm-Speicherung ausgesprochen praktisches Werkzeug gewesen wäre, würde durch die fehlenden Such- und Ordnungskonzepte doch schnell an seine Grenzen stoßen. Wie Buckland betont, sei die Bedeutung Bushs für die Entwicklung der Informationstechnik vielleicht weniger in den Details seiner Ausführungen, als vielmehr im Renommee seiner Person als »engineer of the American century«<sup>7</sup> zu suchen. Bushs Text entwirft ein evokatives Bild, in dem Informationstechnik personalisiert und domestiziert als *Medium* erscheint; aber wenn sich spätere Pioniere wie Douglas Engelbart, J. C. R. Licklider und Ted Nelson explizit auf den *Memex* beziehen, dann nicht zuletzt, um vom Prestige seines berühmten Erfinders zu profitieren. Im Kontext dieses Beitrags ist es v.a. wichtig darauf hinzuweisen, dass Bushs Beitrag das Konzept der »Assoziation«, also der flexiblen Verknüpfung von Ideen oder Dokumenten, als mögliches Such- und Ordnungsprinzip legitimiert.

Bevor wir uns der Zitationsanalyse zuwenden, ist es nötig, kurz die Arbeit des Belgiers Paul Otlet ins Auge zu fassen. Dieser nimmt einerseits viele Ideen Bushs vorweg und bringt andererseits die Logik der *Klassifizierung* auf ihren Höhepunkt.

## 2.2. Das Mundaneum

Als Herausgeber des *Sommaire périodique des revues de droit*, einer Übersicht über den Inhalt der in Belgien erscheinenden juristischen Zeitschriften, beginnt der studierte Jurist und Anwalt Paul Otlet Anfang der 1890er Jahre, sich intensiv mit Fragen der Bibliographie, also der »Buchkunde«, auseinanderzusetzen. Gemeinsam mit seinem Freund und Kollegen Henri Lafontaine wird 1895 das *Office international de bibliographie* gegründet und ein utopisches Projekt beschlossen, das darauf abzielt, alle jemals publizierten Bücher zu katalogisieren, was zu dem Zeitpunkt auf gut acht Millionen Titel gekommen wäre.<sup>8</sup> Die Frage der Systematik war angesichts des ambitionierten Unternehmens besonders sensibel. Historisch gesehen war die Entwicklung des Bibliothekskatalogs erst relativ spät in Bewegung gekommen, das Zettelkastensystem als flexibler Speicher etwa wurde erst nach der Französischen Revolution erfunden. Auf der Ebene der Klassifikation war die Arbeit Diderots und d'Alemberts an der *Encyclopédie* wegweisend, aber als Ausgangspunkt für Otlet kam nur

---

<sup>6</sup> M. Buckland: Overview, S. 4

<sup>7</sup> Ebd., S. 4

<sup>8</sup> Vgl. Levie, Françoise: *L'homme qui voulait classer le monde*, Bruxelles: Les Impressions Nouvelles 2006.

die damals am europäischen Kontinent noch wenig bekannte *Dewey Decimal Classification* (DCC) des amerikanischen Bibliothekars und Aufklärers Melvil Dewey in Frage. Die Organisation dieser Wissenssystematik folgt einer rigiden Baumstruktur, in der sich die zehn Kernbereiche menschlichen Wissens in Blöcken von jeweils zehn Untergruppen immer tiefer verzweigen. In ihrer geometrischen Perfektion liegt aber auch die Schwäche der DCC, und Otlet und Lafontaine antizipieren genügend Probleme, um die Erstellung einer modifizierten und erweiterten Form, publiziert als *Classification Décimale Universelle* (CDU), zu rechtfertigen. Diese auch heute noch häufig verwendete Systematik ist internationaler in ihrer kulturellen Perspektive und erweitert die DCC um eine facettenbasierte<sup>9</sup> Komponente, die es ermöglicht, jedes Werk z.B. mit Informationen zu Sprache, Ort und Form zu kennzeichnen. Darüber hinaus bietet die CDU die Möglichkeit, Querverbindungen zwischen Kategorien herzustellen und Werke an mehreren Stellen im Baum zu platzieren. Die CDU wird 1907 erstmals veröffentlicht, aber schon ab 1895 wird an dem »totalen« Katalog, dem *Repertoire Bibliographique Universel* (RBU), gearbeitet. Erst 1919 wird das Vorhaben durch die Unterstützung der belgischen Regierung auf finanziell einigermaßen sichere Beine gestellt, und mit dem *Palais Mondial* (ab 1924 *Mundaneum*) entsteht mit Hilfe des RBU und einigen geschulten Mitarbeitern die »perfekte analoge Suchmaschine«<sup>10</sup>, die in ihrer Hochzeit per Post 1500 bibliographische Anfragen pro Jahr bearbeitet und dabei auf einen Katalog zurückgreifen kann, der bei seiner Schließung 1934 15 Millionen Indexkarten beinhaltet.

Otlet war seiner Zeit in vielerlei Hinsicht voraus und die Beständigkeit der CDU belegt dies. Für den Historiker W. Boyd Rayward, dessen Arbeit<sup>11</sup> Otlet auch außerhalb der Bibliothekswissenschaft bekannt gemacht hat, ist der Belgier der eigentliche Erfinder des Hypertexts und der 1934 erscheinende *Traité de documentation* könne ohne Einschränkungen als einer der Gründungstexte der Informationswissenschaft bezeichnet werden. Und obwohl das *Mundaneum*, das RBU und die CDU klar von der Möglichkeit einer universellen Ordnung des Wissens ausgehen, entwickelt Otlet eine Methodik, die dem Vorwurf der Perspektivität mit einem System zur Weiterentwicklung der CDU entgegentritt, das an den wissenschaftlichen *peer-review* oder heutige Folksonomien erinnert.

Während Otlet versucht, die klassische Logik der Klassifikation für das Informationszeitalter »fit« zu machen, ist die Zeit nach dem Zweiten Weltkrieg immer stärker von einer philosophisch-fundamentalen Kritik an universalistischen Konzeptionen des Wissens

---

<sup>9</sup> Die Facettenklassifikation basiert auf einer polyhierarchischen Logik, in der mehrere Klassenbäume parallel durchlaufen werden können. Im Fall der UDC erlaubt diese Erweiterung eine Suche z.B. nach allen Lehrbüchern (Zusatzbaum »Form«) zum Thema Landwirtschaft (Dezimalklassifikation 630), die auf Deutsch (Zusatzbaum »Sprache«) erschienen sind.

<sup>10</sup> Wright, Alex: »The Web that Time Forgot«, in: The New York Times vom 17.06.2008.

<sup>11</sup> Rayward, W. Boyd: The Universe of Information. The Work of Paul Otlet for Documentation and International Organisation, VINITI: Moscow, 1975.

gekennzeichnet.<sup>12</sup> Diese theoretischen Zweifel werden von praktischen Versuchen begleitet, andere Methoden der Wissenserschließung zu denken und umzusetzen – nicht zuletzt, um dem immer schneller werdenden Publikationsausstoß Rechnung zu tragen. Genau an dieser Stelle setzt die Zitationsanalyse an.

### **3. Der *Science Citation Index***

Die Zitationsanalyse steht einerseits an einem Wendepunkt was den Vorstellungshorizont für Wissensordnungen anbelangt; andererseits ist sie mit einer Maschine konfrontiert, die den praktischen und intellektuellen Möglichkeitsraum gleichzeitig orientiert und erweitert. Der Computer macht es nicht nur möglich, große Informationsmengen zu speichern, sondern diese auch mit komplizierten Verarbeitungsverfahren zu durchsuchen, zu sortieren, zu filtern und zu analysieren.

#### **3.1. Eugene Garfield und die Anfänge der Zitationsindexierung**

Einer der ersten, die dieses Potential für den Bereich der Wissensverwaltung nicht nur theoretisieren, sondern praktisch nutzbar machen wollen, ist Eugene Garfield, die bis heute zentrale Figur im Bereich der sog. »Zitationsanalyse« (*citation analysis*). Dieses Verfahren besteht darin, die zum Kern des wissenschaftlichen Regelwerks gehörende Praxis des Zitierens zum Ausgangspunkt für allerlei Such- und Analysemethoden zu machen. Während es beim *Memex* in erster Linie um die persönliche Verwaltung von und den Zugang zu Dokumenten ging und Otlets totaler Katalog in immer genaueren Verästelungen das Weltwissen kartographieren sollte, ist die Zitationsanalyse ein abstraktes, mechanisches und auf den ersten Blick simples Prinzip für die Nutzbarmachung wissenschaftlicher Publikationen. Und auch wenn sich die Ansätze wesentlich unterscheiden, ist die Basisdiagnose doch weitgehend identisch:

»There is a literal deluge of literature published each year, and it is increasing in geometric proportions.«<sup>13</sup>

Für Garfield stellt die »Sintflut« vor allem für die Wissenschaft ein Problem dar, weil gerade hier manuelle Indizierungsverfahren dem beschleunigten Produktionsausstoß einer stetig wachsenden Anzahl wissenschaftlicher Zeitschriften einfach nicht mehr nachkommen. Nach dem Zweiten Weltkrieg wird deshalb die Suche nach technischen und konzeptuellen Ansätzen zur Bewältigung der Informationsflut intensiviert. An der *Welch Medical Library* der *Johns Hopkins* Universität startet 1948 ein groß angelegtes Projekt, das die Möglichkeiten der automatischen Indizierung und Durchsuchung medizinischer Literatur erforschen soll.<sup>14</sup> Der

---

<sup>12</sup> Vgl. Lyotard, François: *La condition postmoderne: rapport sur le savoir*, Paris: Minuit 1979.

<sup>13</sup> Garfield, Eugene: »The Crisis in Chemical Literature«, Address to the Maryland Section of the American Chemical Society, September 22, 1952, <http://www.garfield.library.upenn.edu/papers/crisisinchemlit1952.html>.

<sup>14</sup> Vgl. Trolley, Jacqueline/O'Neill, Jill: »The Evolution of Citation Indexing«, in: Mary Ellen Bowden/Trudi Bellardo Hahn/Robert V. Williams (Hg.), *Proceedings of the 1998 Conference on the History and Heritage of*

junge Chemiker<sup>15</sup> Eugene Garfield, der sich schon seit Längerem für die organisatorische Dimension wissenschaftlicher Arbeit interessiert, wird 1951 Mitarbeiter im Bereich der chemischen Literatur. Da das *Welch* Projekt zu diesem Zeitpunkt eines der ambitioniertesten seiner Art ist, kommt Garfield nicht nur mit vielen zentralen Figuren der jungen Computer- und Informationswissenschaften in Kontakt<sup>16</sup>, sondern auch mit dem Stand der Technik im Bereich mechanischer Sortierungsverfahren<sup>17</sup>. Denn das deklarierte Ziel des Projekts ist es, den »Faktor Mensch« zu reduzieren und durch maschinelle Verfahren die Geschwindigkeit der Indizierung zu erhöhen.

Schon während der Arbeit an der Johns Hopkins Universität beginnt Garfield eigenständig weiterführende Projekte zu entwickeln. Als besonders wichtig ist dabei die Zeitschriftenserie *Current Contents* (CC) hervorzuheben, die in regelmäßigen Abständen einen Überblick über Neuerscheinungen in verschiedenen wissenschaftlichen Feldern liefert. Seit 1953, und intensiver ab 1957, erscheinen die CC regelmäßig, und die Erfahrung mit der routinemäßigen Erfassung einer großen Anzahl an Zeitschriften wird sich für die spätere Zitationsindizierung als ausgesprochen nützlich erweisen. Um seine Projekte leichter zu organisieren und zu finanzieren, gründet Garfield 1955 eine Firma, *DocuMation Inc.*, aus der 1960 das *Institute for Scientific Information* (ISI) hervorgeht.<sup>18</sup>

Die zündende Idee, wissenschaftliche Literatur mit Hilfe der Erfassung von Zitaten »navigierbar« zu machen, stammt allerdings nicht von Garfield selbst, sondern von William Adair, dem pensionierten Leiter von *Shepard's Citations*<sup>19</sup>, einem seit 1873 erscheinendem juristischen Nachschlagewerk, das die für die *common law* Rechtstradition so wichtigen Bezüge zwischen Gerichtsurteilen dokumentiert. Adair wird über einen Zeitungsartikel auf das *Welch* Projekt aufmerksam und schreibt 1953 einen Brief an dessen Leiter um das Konzept der Suche über Zitatreferenzen vorzustellen. Das Antwortschreiben kam von Eugene Garfield, der zu diesem Zeitpunkt aber nicht auf die Idee eingehen konnte oder wollte. Erst ein Jahr später, nach seiner Kündigung, nimmt Garfield mit Adair Kontakt auf.<sup>20</sup> Kurz darauf

---

Science Information Systems, Medford, NJ: Information Today 1999, S. 124-126.

<sup>15</sup> Bei der Entwicklung und Anwendung technischer Verfahren zur Informationsverwaltung ist die Chemie ausgesprochen häufig Vorreiterdisziplin.

<sup>16</sup> Garfield lernt z.B. Hans Peter Luhn, den deutschstämmigen Informationspionier, und John Mauchly, einen der Erbauer von ENIAC und EDVAC, kennen.

<sup>17</sup> Vgl. Garfield, Eugene: »On the Shoulders of Giants«, in: Mary Ellen Bowden/Trudi Bellardo Hahn/Robert V. Williams (Hg.), *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems*, Medford, NJ: Information Today 1999, S. 237-251.

<sup>18</sup> Ebd.

<sup>19</sup> *Shepard's Citations* wurde 1996 von *LexisNexis* gekauft und in das Angebot der Firma übernommen.

<sup>20</sup> Vgl. Wouters, Paul: »The Creation of the Science Citation Index«, in: Mary Ellen Bowden/Trudi Bellardo Hahn/Robert V. Williams (Hg.), *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems*, Medford, NJ: Information Today 1999, S. 127-136

reicht Garfield, der nebenbei Bibliothekswissenschaft studiert, eine Seminararbeit mit dem Titel »*Association-of-Ideas Techniques in Documentation. Shepardizing the Literature of Science*« ein und 1955 erscheint ein Artikel<sup>21</sup> in der renommierten Zeitschrift *Science*, in dem die Übertragung des zitatabasierten Verknüpfungsprinzips auf wissenschaftliche Literatur in relativ ausgereifter Form präsentiert wird. So legt der Text nicht nur die eigentliche Idee, sondern schon ein durchdachtes Verfahren zu ihrer Verwirklichung dar. Für den praktisch orientierten Garfield steht das Projekt damit aber erst am Anfang, weil es schließlich darum geht, eine technische Umsetzung auch tatsächlich zu bewerkstelligen.<sup>22</sup>

Der Weg von der Idee zu ihrer Umsetzung ist alles andere als einfach und Garfields herausragende Leistung besteht in der Zielstrebigkeit, mit der das Projekt über die nächsten Jahrzehnte verfolgt wird. Zuerst sollen zwei Pilotprojekte die prinzipielle Machbarkeit der Zitationsindexierung überprüfen.<sup>23</sup> Das erste Vorhaben ist mit der Indizierung von 5000 chemischen Patenten, v.a. der Firma *Merck*, noch relativ klein angelegt. 1960 finanzieren dann aber das *National Institute of Health* (NIH) und die *National Science Foundation* (NSF) ein viel größeres Projekt: die Erstellung einer Zitationsdatenbank für das Feld der Genetik, die 1961 als *Genetics Citation Index* erscheint. Die zu dieser Zeit gerade entstehende molekulare Genetik war aufgrund ihrer Neuheit und Interdisziplinarität – und den sich daraus ergebenden Schwierigkeiten für die klassische Indizierung – das ideale Testfeld für die Zitationsindexierung. Ab 1961 gibt das ISI mit dem *Index Chemicus* auch sein erstes kommerzielles Produkt heraus. Obwohl anfangs defizitär operierend, hatte dieser Zitationsindex der Chemie für den Chemiker Garfield wohl besonderen Wert. Das zentrale Projekt ist aber die Erstellung eines Index für die gesamten Naturwissenschaften, und dieses Unternehmen belastet die Firma finanziell in der Folge bis aufs Äußerste<sup>24</sup>. 1963 gelingt es dennoch, die erste Ausgabe des disziplinenübergreifenden *Science Citation Index* (SCI) auf den Markt zu bringen. Das Projekt erweist sich als lebensfähig: ab 1964 erscheinen jährlich eine Vollaussage sowie vierteljährliche Aktualisierungen, 1965 wird der *Social Sciences Citation Index* (SSCI) gestartet und 1975 kommt schließlich der *Arts and Humanities Citation Index* (A&HCI) dazu.

### 3.2. Wie funktioniert der SCI?

Aber wie muss man sich den SCI genau vorstellen? In erster Linie handelt es sich um eine ausgetüftelten Logistik. Schon in der ersten Ausgabe erfasst der SCI immerhin 613

---

<sup>21</sup> Garfield, Eugene: »Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas«, in: *Science* 122 (1955), S. 108-111.

<sup>22</sup> Vgl. Persson, Olle: »Citation Indexes for Science - A 50 year citation history«, in: *Current Science* 89 (2005), S. 1503-1504.

<sup>23</sup> Vgl. J. Trolley/J. O'Neill: *The Evolution of Citation Indexing*

<sup>24</sup> Cawkell, Tony/ Garfield, Eugene: »Institute for Scientific Information«, in: Einar H. Fredrikson (Hg.), *A Century of Science Publishing. A collection of essays*, Amsterdam: IOS Press 2001, S. 149-160.



Zeitschriftentitel und etwa 1,4 Millionen Zitate. Bis 1966 verdoppelt sich diese Zahl nochmals.<sup>25</sup> Diese Verarbeitungsleistung beruht in erster Linie auf Garfields mittlerweile langjährigem Umgang mit der automatischen Datenverarbeitung: die an der *Welch Library* gewonnene Erfahrung mit Lochkartenmaschinen und anderen mechanischen Verfahren, die durch die *Current Contents* etablierte Routine in der Aufarbeitung wissenschaftlicher Publikationen und die in den Pilotprojekten getesteten Methoden der computerbasierten Sortierung erlauben eine von Anfang an professionelle Umsetzung der Idee in ein bis zum heutigen Tage regelmäßig erscheinendes Produkt. Die Aufgabe war gewaltig, schließlich musste für jeden Artikel und für *jedes einzelne Zitat* per Hand eine Indexkarte erstellt werden, bei der ersten Ausgabe also für alle Texte des Erfassungsjahrs 1961 sowie für alle von diesen zitierte Beiträge. Das ISI war nicht zuletzt deshalb auf über 100 Mitarbeiter angewachsen.<sup>26</sup> Eine anwendbare Kodierungslogik hatte Garfield schon in seinem *Science* Artikel von 1955 vorgestellt: wenn jede Zeitschrift und jeder Artikel mit einer fortlaufenden Nummer codiert wird, kann eine Indexkarte für ein Zitat ganz einfach die Nummern (Journalcode und Artikelcode) des zitierten sowie des zitierenden Artikels erfassen. Mit dem Computer, IBM Maschinen der 1400 und 700/7000 Serien, konnten diese Informationen dann so sortiert werden, dass für jeden Beitrag nicht nur die Liste der zitierten Referenzen, sondern auch die Liste der diesen zitierenden Artikel ausgegeben wurde. Der SCI besteht deshalb auch aus zwei Hauptindices<sup>27</sup>, beide alphabetisch nach Autornamen sortiert: der *Source Index* listet alle erfassten Beiträge sowie deren Referenzen auf; der eigentliche *Citation Index* nennt wiederum alle Texte, die auf einen der erfassten Beiträge des jeweiligen Autors verweisen. Bis auf eine einfache Typisierung – Beiträge sind als Artikel, Literaturübersicht, Kommentar, Brief, etc. gekennzeichnet – bleibt das Prinzip der Klassifizierung nach Form oder Thematik komplett außen vor, der einzige »Inhalt« des Index ist das über hunderte Seiten gespannte Verweisungsnetzwerk zwischen wissenschaftlichen Publikationen. Denn im Gegensatz zu Bushs evokativer Medienmaschine erscheint der SCI zuerst in traditioneller Form – als Buch.<sup>28</sup>

Der SCI war in erster Linie als Suchwerkzeug gedacht und wurde als solches präsentiert und beworben. Als »Suchanfrage« fungierte ein beliebiger Artikel, von dem es in die »Vergangenheit« (zitierte Texte) und in die »Zukunft« (zitierende Texte) navigiert werden konnte. Die Kombination dieser beiden Verfahren, das »cycling«, konnte dabei besonders

---

<sup>25</sup> Malin, Morton V.: »The Science Citation Index: A New Concept in Indexing«, in: *Library Trends* 16 (1968), S. 374-387.

<sup>26</sup> Garfield, Eugene/Stevens, Leo J.: »Über den Science Citation Index (SCI) und Verwandte Entwicklungen der jüngsten Zeit«, in: *Nachrichten für Dokumentation* 16 (1965), S. 130-140.

<sup>27</sup> Ein dritter Index, der *Patent Index*, erschließt alle zitierten Patente.

<sup>28</sup> Seit 1980 werden SCI, SSCI und A&HCI als CD-ROM angeboten. Der SCI war auch eine der ersten Datenbanken, die über das DIALOG System online zugänglich gemacht wurden. Heute verarbeitet das aus dem SCI hervorgegangene *Web of Science* der Firma Thomson Reuters über 10.000 Zeitschriften ab 1900 und ist über das WWW zugänglich.

interessante Ergebnisse liefern. Der SCI verstand sich allerdings nicht als Ersatz für klassische Methoden der Indizierung, sondern als »Lösung« für die spezifischen praktischen Schwierigkeiten die diese Methoden mit sich bringen. Garfield ging es dabei in erster Linie um eine Verschiebung des Blickwinkels von den Bedürfnissen der Bibliotheksorganisation hin zu denen der wissenschaftlichen Praxis.<sup>29</sup> Daraus ergeben sich eine Reihe von Konsequenzen:

- Die Schnelligkeit rückt ins Zentrum. Da die Indizierung nach Zitaten wesentlich einfacher ist als die nach Inhalten, kann das ISI Wissenschaftler sehr schnell über relevante Neuerscheinungen informieren. Mit dem *Automatic Subject Citation Alert* (ASCA) wird es möglich, einen oder mehrere Artikel zu »abonnieren«, um dann wöchentlich über eventuell hinzugekommene Zitierungen benachrichtigt zu werden. Die hohe Geschwindigkeit ergibt sich aus dem weitgehend mechanischen Charakter der Arbeit, die sich in weiten Teilen von Maschinen erledigen lässt. Die manuelle Erstellung der Indexkarten wiederum verlangt keinerlei Expertenwissen, sodass Garfield sogar damit spekuliert, die Arbeit in Billiglohnländer wie Indien oder Sri Lanka auszulagern.<sup>30</sup>
- Indem die Vernetzung der Texte in gewisser Weise an ihre Autoren *delegiert* wird, umgeht der SCI die Frage der »richtigen« Organisation des Wissens und damit auch den Vorwurf der Subjektivität oder der kulturellen Perspektivität, mit denen sich quasi jedes Klassifikationsschema auseinandersetzen muss. Auf den ersten Blick beschränkt sich der SCI darauf, eine aus der wissenschaftlichen Arbeit selbst hervorgehende Wissensordnung sichtbar und navigierbar zu machen.
- Ohne Klassifikationsraster gibt es auch scheinbar keine globale Ordnungsebene, keinen Baum des Wissens, der bei jeder Veränderung erweitert oder angepasst werden müsste. Der SCI ist dadurch nicht nur offen für thematische Verschiebungen, sondern auch agnostisch gegenüber Disziplinengrenzen. Damit scheint er für die Erschließung der interdisziplinären Felder, die sich in den Naturwissenschaften nach dem Zweiten Weltkrieg entwickeln, wie prädestiniert.

Garfield wird nicht müde, diese Vorteile zu betonen, und während der *Science* Artikel von 1955 zumeist Indifferenz oder Ablehnung hervorruft, sind die Reaktionen beim Erscheinen des SCI 1963 zwar immer noch gemischt, aber zum Teil auch ausgesprochen positiv.<sup>31</sup> In den acht Jahren zwischen 1955 und 1963 gelingt es Garfield, das konzeptuelle Prinzip der

---

<sup>29</sup> »Thus rapid timing, essential to researchers, was completely sacrificed to improve the retrieval functions important to librarians.«, aus Garfield, Eugene: »Citation Consciousness. The Origins of Citation Indexing in Science«, in: *Password* 6 (2002), S. 22-25, S. 23

<sup>30</sup> Vgl. Garfield, Eugene: »Retrospective on the Sociological and Historical Uses of Citation Data at ISI«, Meeting on Use of Citation Data in the Study of Science. Baltimore, Maryland, April 1, 1975, <http://www.garfield.library.upenn.edu/papers/retrospective1975.html>.

<sup>31</sup> Vgl. Wouters, Paul: *The Citation Culture*. Unveröffentlichte Dissertation, Amsterdam 1999, S. 19ff.

Zitationsindexierung mit einem platzsparenden Kodierungsverfahren, stabilen mechanischen Verarbeitungsmethoden und einem Geschäftsmodell zu verbinden, das sich nach den ersten Schwierigkeiten als lebensfähig erweist. Mit Latour könnte man sagen, dass ein Netz an heterogenen Verbindungen geknüpft und stabilisiert wird, welches die bloße Idee eines Zitationsindex Realität werden lässt. Die unermüdliche Überzeugungsarbeit, die vielen Publikationen und Konferenzbesuche, das Ausprobieren und Basteln neuer Techniken sowie der Aufbau wirtschaftlicher Strukturen sind nicht vernachlässigbares Beiwerk einer genialen Erfindung, sondern notwendige Komponenten der *Erfindungsarbeit*. Erst die Zusammenführung all dieser Elemente macht es möglich, aus einem entschieden revolutionären Prinzip mit wissenschaftlicher Literatur umzugehen, ein tatsächliches *Ding* zu generieren. Die zugrundeliegende Verschiebung von einer qualitativ-klassifikatorischen zu einer quantitativ-explorativen Perspektive ist dabei allerdings nicht zu vernachlässigen und soll in Folge genauer bestimmt werden.

## 4. Vom Katalog zum Zitationsgraph

Es ist unmöglich, über die Mathematik als Technik der Informationsorganisation zu sprechen, ohne anzumerken, dass es erst die logischen Bausteine des Computers sind, die mathematische Verfahren mechanisch ausführbar und damit »performativ« machen. 1948 schreibt der Mathematiker Warren Weaver über die neuen Maschinen:

»[Computers] will make it possible to deal with problems which previously were too complicated, and, more importantly, they will justify and inspire the development of new methods of analysis applicable to these new problems [...].«<sup>32</sup>

Technik ist gleichzeitig Werkzeug und Möglichkeitshorizont: erst der Computer macht die Umsetzung einer eigentlich alten Idee in dieser Form realisierbar<sup>33</sup>, aber einmal unterwegs erweitert und verschiebt die Computerisierung den Vorstellungsräum und bereitet den Weg für eine Reihe signifikanter Folgeinnovationen. Während es zuerst nur um ein neues Verfahren zur Informationssuche geht, wird doch schnell klar, dass der SCI in noch ganz anderer Weise verwendbar sein wird. Der Computer macht es nicht nur möglich, das hinter dem Index lauende Zitationsnetzwerk als Analyseobjekt zu begreifen, sondern erlaubt es eben auch, praktische Verfahren zu entwickeln, dieses Netzwerk zu untersuchen und »verwertbar« zu machen. Diese Verfahren bestehen im Wesentlichen darin, mathematische Methoden und Vorgehensweisen auf Fragen der Informationsverarbeitung umzulegen.

### 4.1. Vom Katalog zum Netzwerk

Der Bibliothekskatalog und die Enzyklopädie sind von einem enzyklopädischen, räumlichen und universalistischen Bild des Wissens geprägt. Ein Blick auf die verästelten Wissensbäume bei Bacon, Diderot und d'Alembert oder bei Dewey genügt um zu verstehen, dass es hier

---

<sup>32</sup> Weaver, Warren: »Science and Complexity«, in: American Scientist 36 (1948), S. 536-544.

<sup>33</sup> Vgl. P. Wouter: Citation Culture, S. 77

jeweils darum geht, menschliche Erkenntnis in eine Form zu bringen, die, vom Allgemeinen zum immer Spezielleren führend, der Ordnung der Welt selbst entspricht. Während Otlet mit der CDU versucht, diese Logik quasi zu vollenden und gleichzeitig die durch Starrheit und Abschottung entstehenden Probleme einer hierarchischen Wissensordnung zu entschärfen, bricht Garfields Zitationsanalyse mit dem Projekt einer verbindlichen Kartographie des Wissens. Die aprioristisch-universelle wird zugunsten einer empirisch-partikularistischen Perspektive aufgegeben: erstens geht es nicht mehr um die Indizierung menschlichen Wissens als zusammenhängendem Ganzem, sondern einzig um wissenschaftliche Publikationen und damit um die spezifischen Kommunikationsformen einer spezifischen Wissenskultur. Zweitens präsentiert sich der SCI nicht als humanistische Kartographie aller Erkenntnis, sondern als praktisches Werkzeug, das Primärtexte nach ihrer eigenen Zitationspraxis indiziert und zugänglich macht. Die gebotene Leistung, der »Mehrwert«, ist *mechanisch* und nicht *philosophisch*. Drittens ist anzumerken, dass die übersichtliche Baumstruktur prinzipbedingt zugunsten einer impliziten Netzwerkform aufgegeben und die zentralperspektivische Sicht auf das indizierte Wissen durch eine Systematik ersetzt wird, in der jeder Artikel auf den ersten Blick einen gleichberechtigten Einstieg in das Verweisungsgewirr darstellt. Wenn das enzyklopädische Prinzip nach d'Alembert darin besteht, »à placer [...] le Philosophe au-dessus de ce vaste labyrinthe dans un point de vue fort élevé«<sup>34</sup>, dann ist der SCI zuerst einmal nichts anderes als eine Methode, um im Labyrinth des Wissens leichter von einer »connaissance« [sic.] zur nächsten zu gelangen. Der enzyklopädische Wissensbaum operiert »top-down«, der SCI »bottom-forward«. In Anlehnung an Bush unterstreicht Garfield, dass es dem SCI wohl weniger um Ordnung geht, als darum, Verbindungen zu schaffen, auch über Disziplingrenzen hinaus:

»By virtue of its different construction, [a citation index] tends to bring together material that would never be collated by the usual subject indexing. It is best described as an association-of-ideas index, and it gives the reader as much leeway as he requires.«<sup>35</sup>

Die Aufgabe einer Zentralperspektive wirft nun aber selbst viele neue Fragen auf: Wie soll man in das Netzwerk-Labyrinth einsteigen? Welchen Verbindungen soll man folgen? Wie lässt sich die Orientierung behalten? Die Antworten auf diese Fragen müssen zuerst *von Außen* kommen, gestützt von einem Expertenwissen, ohne dem der SCI quasi nutzlos ist. Der Bruch mit der übersichtlichen Baumstruktur bedingt in gewisser Weise auch den Bruch mit der pädagogischen Tradition der Aufklärung. Sobald der SCI als Datenbestand und Forschungsgegenstand ins Blickfeld rückt, tauchen aber Denkansätze auf, die den gestellten Fragen auf neuen Wegen zu begegnen suchen.

Eine zentrale Rolle spielt dabei Derek J. de Solla Price, der heute als der Begründer<sup>36</sup> der

---

<sup>34</sup> Diderot, Denis/d'Alembert, Jean le Rond: Encyclopédie, ou Dictionnaire raisonné des sciences, des arts et des métiers, Tome 1, Paris: Briasson et al. 1751, S. XV.

<sup>35</sup> E. Garfield: Citation Indexes for Science, S. 108

<sup>36</sup> Die statistische Beschäftigung mit Wissenschaft begann zwar schon in den 1920er mit den Arbeiten von

Scientometrie gilt, also der »Messung« der Wissenschaft mit Hilfe verschiedener mathematischer Methoden. Als Eugene Garfield seinem Bekannten 1962 vorschlägt, die Daten des SCI zu untersuchen, ist dieser »excited«<sup>37</sup>. Das erste Ergebnis erscheint 1965 in *Science*, und während die Herangehensweise noch weitgehend statistisch orientiert ist, fasst Price den SCI als Netzwerk ins Auge:

»This article is an attempt to describe [...] the nature of the total world network of scientific papers. We shall try to picture the network which is obtained by linking each published paper to the other papers directly associated with it.«<sup>38</sup>

Price zeigt unter anderem, dass die Häufigkeit, mit der Artikel zitiert werden, sehr ungleichmäßig verteilt ist: die meisten Texte werden gar nicht oder nur sehr selten zitiert, während einige wenige sehr oft genannt werden. Durch die »skalenfreie« Häufigkeitsverteilung würden Klassiker und »Superklassiker« automatisch, und damit maschinell, identifizierbar.<sup>39</sup> Darüber hinaus lässt die Analyse des SCI die »Halbwertszeit« wissenschaftlicher Publikation als sehr kurz erscheinen und Price gelingt es, eine »Forschungsfront« zu bestimmen, dichte Artikelnetzwerke, die in kurzem Abstand zueinander erscheinen und deren Zitationshäufigkeit schnell abfällt.

Diese ersten Untersuchungen des SCI als Datenbestand müssen wohl als tatsächlich neue Formen des Umgangs mit Wissenschaft – und in weiterer Folge mit Wissen überhaupt – bezeichnet werden. Es sind Herangehensweisen, in denen die Ordnung des Wissens eben nicht mehr als philosophisch-aprioristisches Problem erscheint, sondern als empirische Problematik, die »explorativ« behandelt werden muss – am besten mit den Mitteln der Mathematik. Die statistischen Ansätzen werden dabei schnell um graphentheoretische Methoden erweitert.

## 4.2. Vom Netzwerk zur Graphentheorie

Ein wichtiger Schritt in der Mathematisierung des Umgangs mit dem SCI ist 1967 getan, nachdem Ralph Garner zeigt, dass sich Zitationsnetzwerke komplett als Graph formalisieren lassen.<sup>40</sup> Dies öffnet die Tür zur Anwendung graphentheoretischer Methoden, mit denen anspruchsvollere Analysen möglich werden als mit klassischen statistischen Verfahren. Die elementare Formalisierung ist dabei wenig überraschend: Artikel werden zu Knoten, Zitate zu

---

Alfred J. Lotka und George Kingsley Zipf, aber erst Price etabliert sich quantitative Wissenschaftsforschung als Disziplin.

<sup>37</sup> P. Wouters: *Citation Culture*, p. 80

<sup>38</sup> de Solla Price, Derek J.: »Networks of Scientific Papers«, in: *Science* 149 (1965), S. 510–515, S. 510

<sup>39</sup> Ebd., S. 511

<sup>40</sup> Garner, Ralph: »A Computer-Oriented Graph Theoretic Analysis of Citation Index Structures«, in: Ralph Garner/Lois Lunin/Lois Baker (Hg.), *Three Drexel Information Science Research Studies*, Philadelphia: Drexel Press 1967, S. 3-46

Kanten. Im Vergleich zu den detaillierten und inhaltsschweren Klassifikationen bei Dewey oder Otlet erscheint der Zitationsgraph karg und wenig aussagekräftig. Sein intellektuelles Potential liegt einerseits in der Struktur seiner Verbindungen und andererseits in seiner enormen Größe. Genau deshalb kann der SCI zum Datenmaterial für eine Form des Betrachtens werden, die mit Hilfe algorithmischer Datenverarbeitung das Zitationsnetzwerk als Wissensquelle erschließt. Die konkreten Verfahren nehmen dabei unterschiedliche Formen an und dienen unterschiedlichen Zwecken.

Das »bibliographic coupling«<sup>41</sup> und die etwas kompliziertere »co-citation analysis«<sup>42</sup> etwa vergleichen die Referenzlisten zwischen Artikeln und etablieren indirekte Verbindungen, wenn zwei Texte, die sich einander nicht zitieren, eine oder mehrere Referenzen gemeinsam haben. Auf ähnliche Weise lassen sich Wissenschaftler mit ähnlichem thematischen Profil gruppieren. Stark verknüpfte Stellen im Graph – unabhängig davon ob diese durch direkte Zitate oder indirekte Gemeinsamkeiten bedingt sind – können auf »unsichtbare Universitäten«<sup>43</sup> hinweisen oder auf der Makroebene dazu verwendet werden, disziplinäre Grenzen sozusagen »von unten« festzustellen. Klassifikationen und andere Metainformationen sind hier nicht mehr das Produkt eines Reflexionsprozesses, sondern werden maschinell aus den Daten selbst abgeleitet. Die Geste des Klassifizierens weicht der Geste des Rechnens und das subjektive Organisieren von Wissenskategorien der scheinbar objektiven Feststellung von einfachen Basisindikatoren (Zitaten). Es ist nicht mehr die Karte des Wissens die Universalität beansprucht, sondern die mathematische Verfahrenslogik, mit deren Hilfe sich je nach Algorithmus und Parameter unendlich viele Karten generieren lassen.

Die Verwendung graphentheoretischer Ansätze erlaubt es auch Price, seine scientometrischen Analysen zu vertiefen. In einem Artikel<sup>44</sup>, der heute als einer der Gründungstexte der Netzwerkwissenschaft gilt, zeigt er, dass die Frage, welche Artikel wie oft zitiert werden, wohl nicht alleine von deren Qualität abhängt, sondern von einer Netzwerkdynamik, der Robert K. Merton später den Namen »Matthäus-Effekt« geben wird.<sup>45</sup> Price formalisiert ein Wachstumsmodell für Netzwerke, in dem schon ein relativ kleiner Unterschied in der Anfangsverteilung von Verbindungen zu einem »kumulativen Vorteil« und der oben genannten skalenfreien Häufigkeitsverteilung führt. Auf die Wissenschaft umgelegt, lässt sich also feststellen, dass ein Artikel, der früh mehrmals zitiert wird, eine gute Chance hat, durch die erhöhte Sichtbarkeit noch viel öfter zitiert zu werden: »wer hat, dem wird gegeben«.

---

<sup>41</sup> Kessler, Michael M.: »Bibliographic coupling between scientific papers«, in: American Documentation 14 (1963), S. 10-25.

<sup>42</sup> Small, Henry: »Co-citation in the scientific literature: A new measurement of the relationship between two documents«, in: Journal of the American Society of Information Science 24 (1973), S. 265-269.

<sup>43</sup> R. Garner: A Computer-Oriented Graph Theoretic Analysis, S. 37

<sup>44</sup> de Solla Price, Derek J.: »General Theory of Bibliometric and other Cumulative Advantage Processes«, in: Journal of the American Society for Information Science 27 (1976), S. 292-306.

<sup>45</sup> Merton, Robert K.: »The Matthew Effect in Science«, In: Science 159 (1968), S. 56-63.

Das spezifisch Besondere an all diesen Verfahren ist die Einfachheit, mit der aus Methoden, die zuerst der Analyse und Beschreibung dienen, Techniken abgeleitet werden können, die sich zur Navigation, Suche, Filterung und Hierarchisierung von Information verwenden lassen. Aus Zentralität im Netzwerk wird Sichtbarkeit in einer Liste.

## 5. Von der Beschreibung zur Hierarchisierung

Obwohl der SCI auch heute noch ein auf die Wissenschaft fokussiertes Spezialwerkzeug darstellt, kann er doch als Beispiel für die fortschreitende Mathematisierung unterschiedlicher gesellschaftlicher Abläufe gedacht werden. Zwei Anwendungsfälle verdienen dabei besondere Aufmerksamkeit: der viel diskutierte »Impact Factor« zeigt besonders deutlich, wie Rechenverfahren administrative Entscheidungsprozesse durchdringen, und der Gebrauch des Prinzips der Zitationsanalyse in modernen Web-Suchmaschinen veranschaulicht, wie einfach sich computerbasierte Verfahren von einem Bereich auf einen anderen umlegen lassen.

### 5.1. Der *Impact Factor*

Es ist bemerkenswert, dass direkt auf Zitationsanalyse basierende Formen der Evaluierung wissenschaftlicher Produktivität und Signifikanz heute wohl bekannter sind als der SCI selbst. Die Rede ist vom *Impact Factor* und seinen Derivaten<sup>46</sup>. Die prinzipielle Idee, aus der Häufigkeit von Zitaten Indikatoren für *Relevanz* abzuleiten, ist allerdings älter als der SCI selbst – sie wird zum ersten mal 1927 genannt und beispielhaft umgesetzt. Die leitende Fragestellung ist dabei ausgesprochen pragmatisch:

»What files of scientific periodicals are needed in a college library successfully to prepare the student for advanced work, taking into consideration also those materials necessary for the stimulation and intellectual development of the faculty?«<sup>47</sup>

Gerade für kleinere Universitätsbibliotheken stellt sich angesichts der schnell wachsenden Zahl wissenschaftlicher Zeitschriften die sehr konkrete Frage der Ressourcenverteilung. Um diese zu beantworten, analysieren Gross und Gross eine Jahresausgabe des *Journal of the American Chemical Society* und kompilieren eine Tabelle mit allen 247 zitierten Titeln sowie der jeweiligen Anzahl an Nennungen. Nach Zitationshäufigkeit gereiht, ordnet sich diese Liste *scheinbar wie von selbst* und – wenn man ein Zitat als zählbaren Indikator für Relevanz versteht – sogar nach Wichtigkeit. Auf dieser Grundlage empfehlen die Autoren dann auch spezifische Abonnements sowie den Nachkauf einzelner Jahresausgaben. Mit Hilfe des SCI kann dieses Prinzip nun quasi ohne zusätzlichen Aufwand angewandt und in verschiedene Richtungen erweitert werden.

Schon 1955 verweist Garfield auf die potentielle Nützlichkeit des SCI für die Erforschung der

---

<sup>46</sup> Ursprünglich viel weiter gefasst, bezieht sich der *Impact Factor* heute einzig auf die Bedeutung wissenschaftlicher Zeitschriften.

<sup>47</sup> Gross, P.L.K./Gross, E.M.: »College Libraries and Chemical Education«, in: *Science* 66 (1927), S. 385-389, S. 368.

Geschichte einer Idee. Weil gerade neue Ansätze oft Disziplin- und Kategoriegrenzen überschreiten, seien klassische Kataloge dafür nur bedingt geeignet. Der hochgradig formalisierte SCI hingegen mache es leicht, die Zitationshäufigkeit einzelner Texte zu bestimmen und einen historischen »Impact Factor« zu kalkulieren.<sup>48</sup> Während es also ursprünglich nur um Geschichtsforschung gehen soll, so wird doch schnell klar, dass ein umfassendes Zitationsregister notwendigerweise die Frage der Untersuchung und Evaluierung zeitgenössischer Phänomene aufwirft. In einem Gespräch mit Garfield antizipiert der Computerpionier Allen Newell die potentielle Sprengkraft dieser Idee:

»It is rather easy to predict, I think, that the publication and wide availability of an extensive citation index will have strong social consequences along the line of becoming a controlling variable for the advancement and employment of scientific personnel... It makes little difference whether one likes this or not.«<sup>49</sup>

Zwei Jahre später weißt Price darauf hin, dass sich das Prinzip der Reihung nach Zitationshäufigkeit auf alle möglichen Analyseeinheiten umlegen lässt:

»With such a topography established, one could perhaps indicate the overlap and relative importance of journals and, indeed, of countries, authors, or individual papers by the place they occupied within the map, and by their degree of strategic centralness within a given strip.«<sup>50</sup>

Erste Anwendungen der Zitationsanalyse als Evaluierungsinstrument in der Forschungsadministration folgen in den 70er Jahren auf eine 1976 von Francis Narin publizierte Studie, die wohl nicht zufällig von der *US National Science Foundation* finanziert worden war.<sup>51</sup> Angesichts der amerikanischen und europäischen Bestrebungen, Wissenschaftsförderung an »sichtbare« Leistung zu knüpfen – besonders die Niederlande und Großbritannien sind diesbezüglich als Vorreiter zu nennen –, werden zitationsbasierte Verfahren in den 80er Jahren zum dominanten Evaluierungsinstrument wissenschaftlicher Produktivität und Relevanz, besonders in den exakten Wissenschaften. Heute schlagen sich diese Verfahren in Universitäts- und Länderranglisten, in Vergabegremien für Förderungen und Preise, bei Reihungen wissenschaftlicher Zeitschriften und Budgetentscheidungen von Bibliotheken und selbst in Karrierekommissionen und Gehaltsverhandlungen nieder; überall wird nach Zitationsstatistik *gerankt*.

Dabei spielt die mechanische Funktionsweise des SCI und der abgeleiteten Verfahren eine zentrale Rolle. So werden Zitate automatisch erfasst, und Indikatoren wie *Impact Factor* oder

---

<sup>48</sup> E. Garfield: *Citation Indexes for Science*, S. 109

<sup>49</sup> Garfield, Eugene: »Citation Indexes in Sociological and Historical Research«, in: *American Documentation* 14 (1963), S. 289-291, S. 290.

<sup>50</sup> D. J. de Solla Price: *Networks of Scientific Papers*, S. 515

<sup>51</sup> Vgl. Leydesdorff, Loet: »Evaluation of research and evolution of science indicators«, in: *Current Science* 89 (2005), S. 1510-1517.



*H-Index* etabliert, die auf feststehenden Formeln basieren. Es sind gewissermaßen die Wissenschaftler selbst, die durch ihre Zitationspraxis die Wichtigkeit von Publikationen, Personen, Journals, Universitäten und ganzen Ländern bestimmen. Wir werden in der Folge auf die Frage der »Objektivität« dieser Prozesse zurückkommen müssen, doch zuvor muss eine Entwicklung explizit gemacht werden, die latent hinter den hier besprochenen Veränderungen steht, die aber, obschon sie das gesamte 20. Jahrhundert begleitet, erst in den letzten Jahren klar beschreibbar geworden ist. Es handelt sich um eine Konsequenz der oft beschworenen Informationsflut, um die Verschiebung vom *Suchen* zum *Filtern*. Nirgendwo wird diese Verschiebung deutlicher als am Beispiel des WWW, und die Zitationsanalyse spielt dabei eine wesentliche Rolle.

## 5.2. Das WWW als Zitationsnetzwerk

Während Techniken und Konzepte, wie sie an Otlets *Institut International de Bibliographie* entwickelt wurden, auf die Auffindung von Information abzielen, rückt im Falle des SCI eine andere Funktion schrittweise ins Zentrum. Schon Ende der 60er Jahre fordert Herbert Simon einen Paradigmenwechsel für die mechanische Informationsverwaltung:

»The task is not to design information-distributing systems but intelligent information-filtering systems.«<sup>52</sup>

Um es schematisch zu auszudrücken: spätestens seit der Erfindung der Rotationspresse steuern zumindest westliche Gesellschaften in Richtung eines Informationsüberflusses, der ab Mitte des letzten Jahrhunderts, wie weiter oben angemerkt, explizit als Problem wahrgenommen wird. Sobald zu einem gewissen Thema hunderte oder tausende Publikationen vorliegen, handelt es sich nicht mehr einfach um eine Frage der *Auffindung*, sondern in erster Linie um eine der *Relevanz* von Information. Welche der vielen gefundenen Publikationen soll man rezipieren, wenn es unmöglich geworden ist alles zu lesen? Vor diesem Hintergrund ist es nicht weiter erstaunlich, dass die auf dem SCI basierenden Rankingmethoden eine immer stärkere Rolle in Suchprozessen spielen. Der SCI kann sicherlich dabei helfen, Publikationen zu finden. Seine eigentliche Stärke entwickelt er aber als Filter- und Hierarchisierungsmaschine, die Gefundenes auch gleich ordnet. Es ist daher nicht verwunderlich, dass moderne Suchverfahren im WWW sich mehr oder weniger direkt aus der Zitationsanalyse entwickeln.

Die auf diesen Seiten skizzierte Entwicklung von einer manuellen Logik der Klassifizierung hin zu automatischen Zählverfahren, lässt sich am Beispiel des WWW wie im Zeitraffer nachvollziehen. Während anfänglich handgemachte Kataloge wie das *Yahoo! Directory* das WWW über klassische Kategorienbäume erschließen, laufen ihnen vollautomatische Suchmaschinen Ende der 90er Jahre schnell den Rang ab. Die hohe Qualität seiner Suchergebnisse verschafft einem von zwei Doktoranden entwickelten System den entscheidenden Vorteil, und *Google* wird in Folge zum Synonym für die Suche im Internet.

---

<sup>52</sup> Simon, Herbert A.: *The Sciences of the Artificial*, Cambridge MA: MIT Press 2001 (orig. 1969), S. 144.

Der zentrale Algorithmus, *PageRank*, steht dabei in direkter Verbindung zur Zitationsanalyse:

»Academic citation literature has been applied to the web, largely by counting citations or backlinks to a given page. This gives some approximation of a page's importance or quality. PageRank extends this idea by not counting links from all pages equally, and by normalizing by the number of links on a page.«<sup>53</sup>

Schon 1976 stellen Gabriel Pinski und Francis Narin im Bereich der Zitationsanalyse einen Algorithmus vor, der *PageRank* insofern ähnelt, als einem Zitat abhängig vom *Impact Factor* der zitierenden Zeitschrift ein unterschiedlicher Wert zukommt.<sup>54</sup> Dieses Prinzip wird von Larry Page und Sergei Brin auf das Web umgelegt und folgendermaßen beschrieben:

»PageRank nimmt eine objektive Bewertung der Wichtigkeit von Webseiten vor. [...] Anstatt die direkten Links zu zählen, interpretiert PageRank im Wesentlichen einen Link von Seite A auf Seite B als 'Votum' von Seite A für Seite B. PageRank bewertet dann die Wichtigkeit einer Seite nach den erzielten Voten.«<sup>55</sup>

Ungeachtet der »demokratischen« Rhetorik erkennen wir hier das schon für die Zitationsanalyse herausgestrichene Prinzip der *Delegation* des Rankings an die betroffenen Akteure selbst – in diesem Fall nicht an zitierende Wissenschaftler, sondern an all jene, die Hyperlinks im WWW platzieren. Die mechanische Anwendung des Verfahrens begründet wiederum den Anspruch auf Objektivität. Genau dieser Anspruch soll folgend genauer untersucht werden.

### 5.3. Mechanische Rationalität

Es ist bezeichnend, dass ausgerechnet die *International Mathematical Union* in Kooperation mit dem *International Council of Industrial and Applied Mathematics* und dem *Institute of Mathematical Statistics* 2008 einen Bericht veröffentlicht, in dem die Zitationsanalyse als dominantes Evaluierungsinstrument stark kritisiert wird.<sup>56</sup> Obwohl Theodore Porters Arbeit<sup>57</sup> nicht genannt wird, kommen die Autoren zu einer erstaunlich ähnlichen Diagnose wie der angesehene Historiker und Spezialist für die Geschichte der Quantifizierung:

»The drive towards more transparency and accountability in the academic world has

---

<sup>53</sup> Brin, Sergei/Page, Lawrence: »The Anatomy of a Large-Scale Hypertextual Web Search Engine«, in: Seventh International World-Wide Web Conference, April 14-18, 1998, Brisbane, Australia.

<sup>54</sup> Pinski, Gabriel/Narin, Francis: »Citation influence for journal aggregates of scientific publications: Theory with application to literature of physics«, in: *Information Processing & Management* 12 (1976): 297–312.

<sup>55</sup> <http://www.google.de/corporate/tech.html> vom 31.12.2010. Der Text wurde mittlerweile entfernt.

<sup>56</sup> Adler, Robert/Ewing, John/Taylor, Peter: *Citation Statistics*, Report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS), published June 12, 2008, <http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf>

<sup>57</sup> Porter, Theodore M.: *Trust in Numbers*, Princeton NJ: Princeton University Press 1995, S. 48.

created a 'culture of numbers' in which institutions and individuals believe that fair decisions can be reached by algorithmic evaluation of some statistical data; unable to measure quality (the ultimate goal), decision-makers replace quality by numbers that they can measure.«<sup>58</sup>

Für Porter ist es genau dieser erhöhte Anspruch an Transparenz und Verantwortlichkeit, der in demokratischen Bürokratien dazu führt, dass sich selbst Experten immer weniger auf Standesautorität und Erfahrungswissen berufen können und stattdessen zunehmend quantitative Verfahren in Entscheidungsprozessen den Ausschlag geben. Der Computer – »widely revered in part because it is incapable of subjectivity«<sup>59</sup> – spielt dabei eine zentrale Rolle: solange die Maschine zählt und nicht der Mensch, scheint es mitunter sekundär, *was* gezählt wird.

Für die Zitationsanalyse – und damit auch für die linkbasierte Suche im WWW – drängt sich natürlich die Frage auf, wie weit der zugrundeliegenden Zitierungs- und Verlinkungspraxis Vertrauen geschenkt werden kann. Schon 1973 weisen Cole und Cole darauf hin, dass Wissenschaftler oft Freunde, Kollegen, Mentoren oder besonders eminente Personen zitieren.<sup>60</sup> Und wie Werber im WWW sogenannte »Linkfarmen« zur Verbesserung ihres Rankings bei *Google* einsetzen können, so sind auch Indikatoren wie *Impact Factor* und *H-Index* anfällig für Eigenzitate und Seilschaften. Denn wie sollte man auch zwischen den dichten Verknüpfungen einer »unsichtbaren Universität« einerseits und einem betrügerischen Zitierungskartell andererseits automatisch unterscheiden?

Aber auch ganz ohne Unaufrichtigkeit zu unterstellen, werfen graphentheoretische Hierarchisierungsmaschinen die Frage auf, inwieweit sie bestehende Sichtbarkeitsdynamiken wie den Matthäus-Effekt nicht noch verstärken. Das weiter oben genannte Prinzip des kumulativen Vorteils operiert zwar auch ohne hierarchisierenden Algorithmus allein durch die höhere Wahrscheinlichkeit, in zufällig gefundenen Artikeln auf einen vielzitierten Autor zu stoßen; die Einführung einer Sichtbarkeitsstufe, die nach Verweisungshäufigkeit sortiert, verschärft aber den Unterschied zwischen jenen wissenschaftlichen Texten oder Websites, auf die häufig verwiesen wird, und solchen, die nur selten genannt werden.

Schließlich muss auch eine über Zitationspraxis und Sichtbarkeitsdynamik hinausweisende Frage gestellt werden, die der »mechanischen Rationalität« des Rechnens selbst gilt. Wie können Rechenverfahren ohne Zögern objektiv genannt werden, wenn alternative Berechnungsalgorithmen zwar zu ähnlichen, aber bei weitem nicht identischen Ergebnissen kommen? Loet Leydesdorff, einer der weltweit führenden Experten der Zitationsanalyse, weist unmissverständlich darauf hin, dass die Idee einer widerspruchsfreien, »objektiven« Messung der Qualität menschlicher Wissensproduktion prinzipiell unerreichbar ist:

---

<sup>58</sup> R. Adler/J. Ewing/P. Taylor: *Citation Statistics*, S. 3

<sup>59</sup> T.M. Porter: *Trust in Numbers*, S. 74

<sup>60</sup> Cole, Jonathan R./Cole, Stephen: *Social Stratification in Science*, Chicago: University of Chicago Press 1973.

»[C]onstructs remain ‘soft’, that is, open for debate and reconstruction. De Solla Price’s dream of making scientometric mapping a relatively hard social science can with hindsight be considered as fundamentally flawed. When both the data and the perspectives are potentially changing, the position of the analyst can no longer be considered as neutral.«<sup>61</sup>

Quantifizierung muss deshalb immer als »soziale Technologie«<sup>62</sup> verstanden werden, als Steuerungs- und Verwaltungstechnik, durch die Fragen der Macht – denn um nichts anderes geht es hier – neu verhandelt, konfiguriert und stabilisiert werden. Die Gegenüberstellung von traditionellen Klassifikationssystematiken und zitationsbasierten Verfahren verweist nicht nur auf unterschiedliche Wissensordnungen, sondern auch auf fundamentale Differenzen in den jeweiligen Produktionsweisen, welche wiederum die Einbeziehung unterschiedlicher Akteure und Kompetenzen bedingen. Notwendigerweise gibt es dabei Gewinner und Verlierer.

## 6. Schluss

Abschließend stellt sich die Frage, wie mit den genannten Verfahren von geistes- und kulturwissenschaftlicher Perspektive aus umzugehen ist. In diesem Beitrag wurde versucht zu zeigen, wie mathematische Methoden mit Hilfe des Computers in den Umgang mit Information eindringen und Zugänge zu Wissen generieren, die einerseits empirisch-explorativ orientiert sind und sich andererseits besonders einfach zum Zwecke der hierarchisierenden Reihung und Filterung heranziehen lassen. Die Ausbreitung dieser und ähnlicher Techniken ist dabei schon viel weiter fortgeschritten, als wir mit unserem Exkurs in das Gebiet der Web-Suchmaschinen angedeutet haben. Menschliche Alltagspraktiken entfalten sich immer öfter in digitalen Medien, und diese Medien sind eben immer auch Maschinen, genauer, Computer. Jeder Tastaturanschlag und jeder Mausklick produziert Daten, die wiederum für das Filtern und Sortieren von Information herangezogen werden können. Von der Singlebörse bis zum Nachrichtenportal spielen solche Algorithmen eine zentrale Rolle für die Modulation der *Sichtbarkeit* von Ideen, Neuigkeiten, Produkten oder anderen Nutzern.

In der Bewertung dieser Mechanismen sollten wir uns nicht von der sterilen Debatte leiten lassen, die in den Sozialwissenschaften seit Jahrzehnten qualitative und quantitative Methoden gegeneinander aufstellt. Nichts wäre schädlicher, als mathematischen Verfahren Aussagekraft und Erkenntnispotential abzuspochen – nicht zuletzt, weil es illusorisch ist anzunehmen, dass eine globalisierte und immer komplexer werdende Welt ohne die automatisierte Kognitionsleistung von Maschinen noch in irgendeiner Weise verwaltbar sei. Eine maschinenstürmerische Geistes- und Kulturwissenschaft würde sich selbst ins Abseits stellen.

Der oft emotionale und undifferenzierte Umgang mit Technik verweist auf ein zentrales

---

<sup>61</sup> L. Leydesdorff: Evaluation of research, S. 1512

<sup>62</sup> T.M. Porter: Trust in Numbers, S. 48

Paradoxon unserer gegenwärtigen Gesellschaften: sie sind einerseits von mathematischen Verfahren durchdrungen, andererseits aber erstaunlich desinteressiert, wenn es darum geht, die gesellschaftliche Bedeutung dieser Verfahren zu untersuchen und zu diskutieren.<sup>63</sup> Folglich wird die *Gemachtheit* dieser Technologien aus den Augen verloren:

»We think of statistics as facts that we discover, not numbers we create.«<sup>64</sup>

Wenn wir auf diesen Seiten die »Objektivität« der vom SCI abgeleiteten Indikatoren in Frage gestellt haben, dann nicht um quantitative Ansätze generell abzulehnen, sondern um zu zeigen, dass erst die Analyse der spezifischen *Perspektivität* eines Algorithmus eine Debatte ermöglicht, in der politische und administrative Ziele nicht mehr hinter einer etwaigen *Neutralität* der Maschine versteckt werden können. Wie die Autoren des oben genannten Berichts der *International Mathematical Union* anmerken, sind nicht die Zitationsstatistiken das Problem, sondern der unreflektierte und einseitige Umgang in Evaluationsgremien:

»Using the impact factor alone to judge a journal is like using weight alone to judge a person's health.«<sup>65</sup>

Mit der nötigen Distanz und als Teil eines weiter gefassten Zugangs, können Zitationsstatistiken als ein wertvolles Instrument für die Bewertung wissenschaftlicher Leistungen dienen. Selbiges gilt für maschinelle Techniken in vielen anderen Bereichen. Es kann nicht darum gehen, algorithmische Verfahren endlich wieder durch »menschliches Urteil« zu ersetzen, sondern mechanische Rationalität *als eine spezifische Form menschlichen Urteils* einer aufgeklärten Kritik zuzuführen, ohne die eine sinnvolle Handhabung nur schwer möglich ist.

Die Rückführung moderner Techniken auf historische Formen kann dabei helfen, nicht nur die geschichtliche Kontingenz, sondern auch technische Mechanismen und Faktoren besser zu verstehen. Es wurde hier versucht am Beispiel des SCI darzulegen, wie aus einer relativ einfachen Idee ein reales Suchwerkzeug gemacht wird, und wie dessen Mathematisierung neue Anwendungsfelder öffnet, neue Akteure auf den Plan ruft und neue Problematiken aufwirft. Und wenn auch vieles erst im Rückblick verständlich wird, so handelt es sich doch um eine Serie von Entscheidungsprozessen, deren Analyse uns dabei helfen kann, gegenwärtige Weichenstellungen besser zu verstehen und gegebenenfalls neu zu orientieren. Denn moderne Verfahren zur Informationssuche, -selektion, -sortierung und -hierarchisierung sind nicht mehr Kinder des Bibliothekszeitalters, sondern Anwendungen

---

<sup>63</sup> Douglas Hofstadter spricht in diesem Kontext von einer »anumerischen« (in Analogie zu »analphabetisch«) Zahlenkultur, in der Zahlen und Rechenverfahren zwar eine enorm wichtige Rolle spielen, aber kaum verstanden werden. Vgl. Hofstadter, Douglas R.: »Number Numbness, or Why Innumeracy May Be Just as Dangerous as Illiteracy«, in: *Scientific American* 246 (1982), S. 20-34.

<sup>64</sup> Best, Joel: *Damned Lies and Statistics. Untangling Numbers from the Media, Politicians, and Activists*, Berkeley: University of California Press 2001, S. 160.

<sup>65</sup> R. Adler/J. Ewing/P. Taylor: *Citation Statistics*, S. 2

immer komplexerer mathematischer und algorithmischer Verfahren auf Informationseinheiten, deren Kontext und Inhalt. Die explosionsartige Verbreitung dieser Techniken ruft nach einer *Politikwissenschaft des Algorithmus*, nach einer Form der Analyse und Kritik, die zwischen der Oberflächlichkeit des Feuilletons und der Spezialisierung des Ingenieurwissens einen Raum schafft, in dem algorithmische Verfahren als das gedacht werden können, was sie sind: automatisierte Formen der Gestaltung von Kultur und Gesellschaft.