

Datenmanagement: Grundlagen und Ideen für den Informatikunterricht

Andreas Grillenberger und Ralf Romeike¹

Abstract: Obwohl aktuelle Innovationen im Umgang mit Daten, insbesondere im Kontext von *Big Data*, vielfältige Einflüsse auf Informatik, Wirtschaft, Politik und Gesellschaft haben, werden diese Themen und ihre Auswirkungen im Informatikunterricht bisher üblicherweise allenfalls als Randthemen diskutiert. In diesem Workshop werden daher die aktuellen Entwicklungen im Datenmanagement aus fachlicher und fachdidaktischer Sicht beleuchtet und verschiedene Beispiele vorgestellt und gemeinsam diskutiert, die die Auswirkungen auf den Alltag und den Informatikunterricht verdeutlichen. Dieser Beitrag beschreibt diese Themen zusammenfassend und geht dabei auf die drei wesentlichen den Informatikunterricht betreffenden Säulen ein: die Innovationen in der Fachwissenschaft, deren Bedeutung für den Alltag sowie Herausforderungen bei der Umsetzung im Informatikunterricht.

Keywords: Datenmanagement, Datenbanken, Datenstromsysteme, Big Data, NoSQL

1 Einleitung und Motivation

In der informatischen Bildung stellen *Datenbanken* ein grundlegendes Thema dar. Die Vielzahl innovativer Entwicklungen auf diesem Gebiet, insbesondere bei der Speicherung und Verarbeitung großer und komplexer Datenmengen in immer kürzerer Zeit, *Big Data*, führt dazu, dass sich das Thema Datenbanken zum Fachgebiet *Datenmanagement* ausgeweitet hat: dieses beinhaltet viele zusätzliche Aspekte, beispielsweise Metadaten, die Verwaltung weniger strukturierter Daten oder auch Maßnahmen zur Überprüfung der Datenqualität [DA09] – diesen Aspekten kommt heute eine immer größere Bedeutung zu. Hingegen konzentriert sich Informatikunterricht weiterhin insbesondere auf das tradierte Fachgebiet *Datenbanken*: die Speicherung und Verarbeitung von Daten in Datenbanken unter Nutzung des *relationalen Datenmodells* [GR14a]. Um die Grundlagen und die Funktionsweise moderner Datenspeicherung, wie sie beispielsweise bei Google, Facebook, Twitter oder Amazon stattfindet, nachvollziehen und verstehen zu können, reicht dieses Wissen jedoch nicht mehr aus: zur Bewältigung der riesigen Datenmengen müssen diese beispielsweise auf verschiedene Standorte verteilt und möglichst in Echtzeit verarbeitet werden, während relationale Datenbanken insbesondere auf die Sicherstellung einer hohen Datenqualität optimiert sind.

Auch im täglichen Leben werden Schülerinnen und Schüler, insbesondere bei der Verwaltung ihrer privaten Daten, aber auch bei der Teilnahme am gesellschaftlichen Diskurs, mit vielfältigen Aspekten von Datenmanagement konfrontiert: seien es *Redundanz* und

¹ Friedrich-Alexander-Universität Erlangen-Nürnberg, Didaktik der Informatik, Martensstr. 3, 91058 Erlangen
andreas.grillenberger@fau.de, ralf.romeike@fau.de

Konsistenz, die im Kontext von *Backup* und *Synchronisation* eigener Daten relevant werden, oder *Metadaten*, die einerseits Gefahren für die Privatsphäre bergen, andererseits für Komfort sorgen, indem sie beispielsweise Suchfunktionen in der heute gewohnten Art und Weise (z. B. eine Bildersuche) erst ermöglichen. Aktuelle Innovationen aus der Informatik, beispielsweise die zunehmende Vernetzung von Geräten und Gegenständen des täglichen Lebens durch das *Internet der Dinge* („Internet of Things“, kurz IoT) sorgen für eine weitere Zunahme der Relevanz dieser Themen: Die Vision des IoT sieht vor, dass alle Gegenstände mit denen der Mensch interagiert eine eigene digitale Identität erhalten [Br09] und somit auch Daten produzieren. Im Gegenzug zur Bedeutung dieser Themen zeigt sich jedoch ein mangelndes Bewusstsein über die Gefahren, die mit einem unangemessenen oder unkritischem Umgang mit Daten einhergehen: Nicht nur Schülerinnen und Schüler gehen mit Daten und Passwörtern oft allzu sorglos um, sondern auch Passanten auf der Straße verraten ihre Passwörter teils ohne sich über mögliche Konsequenzen bewusst zu sein [Sp15].

2 Innovationen im Datenmanagement

Den aktuellen Entwicklungen im Kontext von *Big Data* liegen umfangreiche Innovationen in der Informatik zugrunde. Diese Innovationen sind insbesondere im bisherigen Fachgebiet *Datenbanken* zu erkennen, das sich immer stärker ausweitet und eine Vielzahl neuer Themen beinhaltet: beispielsweise Metadaten, Datenqualität und neue Ansätze zur Datenspeicherung. Das damit entstehende Gebiet wird häufig mit dem Begriff *Datenmanagement* bezeichnet und schließt das bisherige Fachgebiet Datenbanken ein. In Folgenden werden wir kurz einige der wesentlichen Entwicklungen in diesem Fachgebiet skizzieren.

Das Schlagwort ***Big Data***, das im Zusammenhang mit Innovationen im Datenmanagement andauernd fällt, wird typischerweise durch die sog. „3 Vs“ *volume*, *velocity* und *variety* charakterisiert [La01]: Neben der offensichtlich immer stärker zunehmenden Datenmenge (*volume*) wird heute bei der Datenspeicherung und -verarbeitung eine immer höhere Geschwindigkeit, idealerweise Echtzeit, erwartet, während gleichzeitig die Geschwindigkeit mit der Daten auf ein Datenverarbeitungssystem einströmen und von diesem verarbeitet werden sollen, immer stärker zunimmt (*velocity*). Neben diesen beiden Aspekten stellt auch die *variety* das Datenmanagement vor neue Herausforderungen: Während Daten für relationale Datenbanken meist in einer definierten Struktur vorlagen oder mit vertretbarem Aufwand in eine solche gebracht werden konnten, liegen Daten heute überwiegend in unstrukturierter Form vor.

Die Lösung dieser Herausforderungen ist mit herkömmlichen Mitteln nicht möglich: Während heute Daten immer schneller verarbeitet und ausgewertet werden sollen (*velocity*), sorgt die große Datenmenge (*volume*) dafür, dass häufig Daten nicht mehr auf einem einzelnen Gerät/Datenspeicher abgelegt werden können, sondern verteilt gespeichert werden müssen. Gleichzeitig kommt der Anforderung Daten konsistent zu Speichern weiterhin große Relevanz zu. Das ***CAP-Theorem*** [Br12] verdeutlicht jedoch den inhärenten Widerspruch einer schnellen Verarbeitung verteilt gespeicherter Daten unter gleichzeitiger Sicherstellung von Konsistenz. Es können daher nur zwei der drei Eigenschaften Konsis-

tenz (*Consistency*), hohe und schnelle Verfügbarkeit (*Availability*) und verteilte Speicherung (*Partition tolerance*) gleichzeitig erreicht werden können. Im Rahmen vieler aktueller Anwendungen muss daher zwischen diesen drei Eigenschaften abgewogen werden: Wenn auf die verteilte Speicherung aufgrund großer Datenmengen nicht verzichtet werden kann, muss entweder auf die Sicherstellung der Konsistenz des Datenbestandes oder auf eine schnelle und hohe Verfügbarkeit verzichtet werden.

Eine der meist diskutierten Veränderungen auf diesem Gebiet, die Entwicklung neuer Datenbankmodelle die oft unter dem Begriff *NoSQL* zusammengefasst werden, setzt genau hier an: Typische NoSQL-Datenbanken ermöglichen die verteilte Speicherung von und den schnellen Zugriff auf Daten, vernachlässigen jedoch die Konsistenz, die damit in den Verantwortungsbereich der Anwendung verlagert wird – im Gegensatz zu relationalen Datenbanken, die insbesondere auf konsistente Datenspeicherung und schnellen Zugriff optimiert sind. Der Begriff *NoSQL* wird daher heute als „Not only SQL“ verstanden [Ed11], da es nicht um die Vermeidung von SQL als Abfragesprache sondern um neue Datenbankmodelle geht, die je nach konkreter Implementierung auch durch SQL-ähnliche Sprachen angesteuert werden können, jedoch unterschiedliche Anwendungsbereiche haben. Um Missverständnisse zu vermeiden, bietet es sich daher an, von nicht-relationalen Datenbanken zu sprechen. Obwohl diese Datenbanken teils als Ersatz für bisherige relationale Datenbanken gehandelt werden, zeichnet sich jedoch eher ab, dass sie für eine umfassende Vielfalt auf diesem Gebiet sorgen und in manchen Bereichen relationale Datenbanken ersetzen können – eine komplette Ablösung jedoch kaum stattfinden wird.

Weitere Entwicklungen im Datenmanagement stellen komplett neue Ansätze zur Datenspeicherung und -verarbeitung dar, wie beispielsweise *Datenstromsysteme*: Im Gegensatz zu Datenbanken werden hier Daten nicht dauerhaft gespeichert und zu beliebigen Zeitpunkten analysiert, sondern direkt bei Übermittlung an das Datenstromsystem verarbeitet und Ergebnisse direkt ausgegeben. Diese Systeme unterscheiden sich in ihrer Konzeption wesentlich von Datenbanken, einerseits durch die (außer für Cache-Zwecke) nicht stattfindende Speicherung, andererseits durch die direkte Reaktionsmöglichkeit auf sich verändernde Daten. Andere Entwicklungen stellen beispielsweise die immer stärker dezentralisierte und an spezialisierte Anbieter ausgelagerte Speicherung und Verarbeitung von Daten in die *Cloud*, neue Verfahren und Anwendungen zur Datenanalyse und ähnliches dar.

3 Bedeutung von Daten und Datenmanagement im Alltag

Neben diesen offensichtlichen Veränderungen im Fachgebiet erlangen Daten auch im Alltag und im gesellschaftlichen Diskurs eine immer größere Bedeutung: Neben der zunehmenden Prägung aktueller Themen durch datenbezogene Aspekte, beispielsweise beim NSA-Skandal, der Vorratsdatenspeicherung oder der PKW-Maut, nutzen immer mehr Personen eine Vielzahl von Geräten die Daten nicht nur verarbeiten sondern auch erfassen. Beispielsweise hat schon 2013 jeder im Schnitt 2,9 digitale Geräte bei sich getragen [Tr13], die meist eine Vielzahl datengenerierender Anwendungen und Sensoren mitbringen. Gleichzeitig zeichnet sich aber auch ein oft allzu sorgloser Umgang mit Daten ab: In

einem Experiment auf einem Künstlermarkt in New York konnte anhand eines Fragebogens festgestellt werden, dass viele Personen bereit waren, persönliche Daten anzugeben, um dafür Punkte zu erhalten, die sie wiederum in Kekse umtauschen konnten [Be14]. Ein besonders interessanter Aspekt bei diesem Experiment war, dass viele Besucher Fragen wie „*What was the name of your first pet?*“ bereitwillig beantworteten ohne zu erkennen, dass es sich dabei um eine der typischen Sicherheitsfragen im Internet handelt. Das Bewusstsein für die Relevanz der Sicherheit eigener Daten scheint daher oft nur gering ausgeprägt zu sein. Auch bei der Verwendung von modernen Anwendungen, wie beispielsweise *WhatsApp*, existieren jedoch nicht nur die offensichtlichen und bereits ausführlich diskutierten Probleme in Bezug auf den Datenschutz, derer sich viele Anwender bewusst sind (z.B. die Veröffentlichung von „zuletzt online“-Zeitstempeln): Beispielsweise erhalten derartige Anwendungen – und damit deren Betreiber – unter anderem Zugriff auf das komplette Adressbuch, in dem auch Daten fremder Personen enthalten sind. Dabei stellt sich die Frage, ob diese Datenweitergabe durch den Eigentümer des Gerätes an den Betreiber des Dienstes überhaupt durch die Personen deren persönliche Daten weitergegeben werden gewünscht bzw. geduldet wird, sodass sich an dieser Stelle mindestens ethische Probleme eröffnen können. Verstärkt wird dieses Problem durch die immer größere Unkontrollierbarkeit von Informationen und Daten: Sobald diese einmal übermittelt wurden, liegen sie außerhalb der eigenen Kontrolle. Obwohl Anwendungen wie SnapChat und ähnliches eine scheinbare Sicherheit der eigenen Daten vermitteln (in diesem Fall durch Löschung der übermittelten Bilder nach einer bestimmten Zeitspanne), ist eine Sicherstellung der Löschung von Daten heute kaum mehr möglich – trotz der Diskussion eines „Rechts auf Vergessenwerden“ [MS10] und eines diesbezüglichen Urteils des Gerichtshofs der Europäischen Union [Ge14]. Es ergeben sich daher völlig neue Herausforderungen bei der Speicherung, Nutzung und Verwaltung von Daten, mit denen heute jeder konfrontiert wird.

Für den verantwortungsbewussten Umgang mit eigenen aber auch mit fremden Daten sind in der heutigen datengetriebenen Welt auch vielfältige neue Schlüsselkompetenzen nötig. Diese Kompetenzen betreffen dabei beispielsweise informatische Konzepte, wie Konsistenz und Redundanz (auch im Zusammenhang mit Synchronisation), die Auswahl geeigneter Datenspeicher, das Bewusstsein um Gefahren für die eigene und fremde Privatsphäre, Maßnahmen zur Sicherstellung der Datensicherheit (Backup, Verschlüsselung) und vieles mehr [GR14c, GR15b]. Solche Kompetenzen können jedoch nur durch einen Informatikunterricht gefördert werden, der neben den grundlegenden Konzepten auch aktuelle Beispiele miteinbezieht und somit einen lebensweltlichen Bezug zum Alltag der Schülerinnen und Schüler schafft. Gerade bei der Thematisierung aktueller Entwicklungen, wie dem Internet der Dinge, ergibt sich neben der Schaffung des lebensweltlichen Bezugs dabei gleichzeitig die Möglichkeit, nützliche Fertigkeiten zu entwickeln, die die Schülerinnen und Schüler im eigenen Alltag einsetzen können.

4 Umsetzung im Informatikunterricht

Modernes Datenmanagement stellt damit nicht nur eine Herausforderung für die Informatik dar, sondern fordert gleichzeitig auch Informatikdidaktik und Informatikunterricht [GR14b]: Neben dem Finden neuer und für den Informatikunterricht geeigneter Beispiele

müssen tradierte Konzepte, die seit Jahren teil des Informatikunterrichts sind, unter einem neuen Licht betrachtet werden, wodurch sich ihre Relevanz wesentlich verändern kann. Gleichzeitig müssen Werkzeuge und Datenquellen gefunden werden, die es ermöglichen auch im Informatikunterricht große Datenmengen einzusetzen, um die Möglichkeiten aber auch Gefahren der aktuellen Entwicklungen im Datenmanagement aufzeigen zu können und insbesondere die Gefahren auch für die Lernenden erkennbar zu machen statt nur vor diesen zu warnen. Gerade bei der Übernahme solcher moderner Themen aus der Fachwissenschaft stellt sich daher die grundlegende Frage, welche dieser Inhalte für die Schule relevant sind, wie diese geeignet didaktisch reduziert werden können und warum jede Schülerin und jeder Schüler heute diese Inhalte kennenlernen und Kompetenzen im Umgang mit diesen Themen erwerben soll.

Der aktuelle Informatikunterricht konzentriert sich im Bereich *Daten* hingegen typischerweise auf das Thema *Datenbanken* [GR14a]. Dabei sind Bemühungen zu erkennen, diese Themen mit aktuellen Beispielen zu verknüpfen, was jedoch häufig an Grenzen stößt. Wird beispielsweise Facebook als alltagsnahes und für die Lernenden häufig motivierendes Beispiel für die Nutzung von Datenbanken herangezogen, besteht die Gefahr zu versuchen, die Komplexität dieses Systems zu stark zu reduzieren: Werden relationale Datenbanken zur Erklärung der Datenspeicherung von Facebook herangezogen ist dies zwar grundsätzlich zutreffend, viele der Möglichkeiten von Facebook, insbesondere die 2013 eingeführte *Graph Search* lassen sich damit jedoch kaum erklären. Diese Graphsuche könnte auf relationalen Datenbanken nur wesentlich langsamer Ergebnisse liefern, da die Daten nicht in geeigneter Form vorliegen und nötige Anfragen somit wesentlich komplexer wären. Stattdessen wird der *Social Graph* in einer höheren Schicht abgebildet, die für diese Nutzung optimiert ist. Die Erklärung von modernen Anwendungen mittels traditioneller Datenbankkonzepte gerät damit an ihre Grenzen, die auch durch didaktische Reduktion oft nicht überwunden werden können, sondern bewusst gemacht werden müssen.

Im Rahmen bisheriger Arbeiten wurde die Möglichkeit der didaktischen Reduktion von komplexen Inhalten exemplarisch anhand von Datenstromsystemen und Datenanalysen im Kontext von Big Data (Data Mining) [GR15a] verdeutlicht. Als wesentliche Leitlinie dienten dabei die fundamentalen Ideen nach Schwill [Sc93]: Durch die Konzentration auf Aspekte die die Kriterien für fundamentale Ideen erfüllen, sollten aus der Menge möglicher Unterrichtsinhalte diejenigen herausgearbeitet werden, die für den Informatikunterricht an allgemeinbildenden Schulen relevant sind – eine Art Übertragung des Konzepts *Data Mining* auf die fachdidaktische Forschung. Im Workshop werden verschiedene Ansätze in denen die zuvor beschriebenen Herausforderungen im Informatikunterricht aufgegriffen werden vorgestellt und gemeinsam diskutiert werden.

5 Fazit

Wie in diesem Beitrag und den referenzierten Quellen gezeigt wurde, eröffnet sich durch aktuelle Themen der Informatik wie *Big Data* bzw. durch die Neuerungen im *Datenmanagement* eine Vielzahl interessanter und für den Unterricht wertvoller Themenbereiche. Die Aufgabe der Fachdidaktik Informatik ist es daher, geeignete Unterrichtsinhalte und

von den Schülerinnen und Schülern zu erwerbende Kompetenzen aus dieser Vielzahl von Möglichkeiten zu extrahieren. Im Workshop werden daher verschiedene Möglichkeiten und Ansätze sich mit diesen Themen zu beschäftigen diskutiert und in kurzen Praxisphasen ausprobiert.

Literaturverzeichnis

- [Be14] Beckett, Lois: People Are Willing to Give Away Their Personal Data for a Cinnamon Cookie. Mashable, 2014. <http://mashable.com/2014/10/01/data-for-cookies/>, zuletzt geprüft: 09.04.2015.
- [Br09] Brand, Leif; Hülser, Tim; Grimm, Vera; Zweck, Axel: Internet der Dinge - Perspektiven für die Logistik. Bericht, VDI Technologiezentrum GmbH, 2009. https://www.vdi.de/fileadmin/vdi_de/redakteur/dps_bilder/TZ/2009/Band%2080_IdD_komplett.pdf, zuletzt geprüft: 09.04.2015.
- [Br12] Brewer, E.: CAP twelve years later: How the "rules" have changed. *Computer*, 45(2):23–29, 2012.
- [DA09] DAMA International: The DAMA Guide to the Data Management Body of Knowledge - DAMA-DMBOK. Technics Publications, LLC, USA, 2009.
- [Ed11] Edlich, Stefan; Friedland, Achim; Hampe, Jens; Brauer, Benjamin; Brückner, Markus: NoSQL. Hanser, Carl GmbH + Co., 2011.
- [Ge14] Gerichtshof, Europäischer: Urteil in der Rechtssache C-131/12. 2014.
- [GR14a] Grillenberger, Andreas; Romeike, Ralf: A Comparison of the Field Data Management and its Representation in Secondary CS Curricula. In: *Proceedings of WiPSCE 2014*. ACM, Berlin, 2014.
- [GR14b] Grillenberger, Andreas; Romeike, Ralf: Big Data – Challenges for Computer Science Education. In (Gülbahar, Yasemin; Karataş, Erinc, Hrsg.): *Informatics in Schools. Teaching and Learning Perspectives*, Lecture Notes in Computer Science 8730, S. 29–40. Springer International Publishing, 2014.
- [GR14c] Grillenberger, Andreas; Romeike, Ralf: Teaching Data Management: Key Competencies and Opportunities. In (Brinda, Torsten; Reynolds, Nicholas; Romeike, Ralf, Hrsg.): *KEY-CIT 2014. Commentarii informaticae didacticae*. Universitätsverlag Potsdam, 2014.
- [GR15a] Grillenberger, Andreas; Romeike, Ralf: Big Data im Informatikunterricht: Motivation und Umsetzung. In: *INFOS 2015. Lecture Notes in Informatics (LNI)*. Köllen Druck+Verlag, Bonn, 2015.
- [GR15b] Grillenberger, Andreas; Romeike, Ralf: Schlüsselkompetenzen im Umgang mit Daten: Ein Thema für den Informatikunterricht. In: *INFOS 2015. Lecture Notes in Informatics (LNI)*. Köllen Druck+Verlag, Bonn, 2015.
- [La01] Laney, Douglas: 3D Data Management: Controlling Data Volume, Velocity, and Variety. Bericht, META Group, February 2001.
- [MS10] Mayer-Schönberger, V.: *Delete: die Tugend des Vergessens in digitalen Zeiten*. Berlin University Press, 2010.
- [Sc93] Schwill, Andreas: *Fundamentale Ideen der Informatik*. Zentralblatt für Didaktik der Mathematik, 1993.

- [Sp15] Spiegel Online: So schnell verraten Passanten ihre Passwörter. 2015. <http://www.spiegel.de/netzwelt/netzpolitik/a-1013721.html>, zuletzt geprüft: 09.04.2015.
- [Tr13] Truong, Kimberly: , INFOGRAPHIC: Users weighed down by multiple gadgets - survey reveals the most carried devices, 2013. <http://nakedsecurity.sophos.com/2013/03/14/devices-wozniak-infographic>, zuletzt geprüft: 09.04.2015.