Big Data – Challenges for Computer Science Education

This is an author draft.
The final publication is available at Springer via http://dx.doi.org/10.1007/978-3-319-09958-3_4

Andreas Grillenberger and Ralf Romeike

Friedrich–Alexander–Universität Erlangen–Nürnberg (FAU)
Department of Computer Science, Computing Education Research Group
Martensstr. 3, 91058 Erlangen, Germany
{andreas.grillenberger,ralf.romeike}@fau.de

Abstract. Data processing is a central topic of computer science and hence also in secondary computer science education, which includes strategies for storing, managing and retrieving data. In the context of Big Data, this field changes tremendously: established ideas, such as avoiding redundancies and storing data in a persistent and consistent way, are dropped in order to speed up the access to distributed stored data as well as its availability. Furthermore, with the rapidly growing impact of data processing on everyone's daily life, computer science education needs to address these aspects as well as their social and ethical implications, such as privacy issues.
This paper points out the major challenges that arise from the outlined developments by evaluating whether database concepts and examples commonly used in CS education need to be updated.

Keywords: Big Data · NoSQL · Data Management · Databases · Data Analysis · Data Privacy · Challenges

1 Introduction

The concept of data processing is central to computer science and hence to computer science education. Not only manipulating data through programming, but also the efficient storage, management and retrieval of data are seen as central aspects. These topics are strongly affected by Big Data, a phenomenon that arose in recent years and which introduced several new innovations. For example, established concepts of data management, such as avoiding redundancies and inconsistencies by saving data in normalized relational database management systems (RDBMS), are dropped. Instead, newer database management systems (DBMS), like the NoSQL databases, are optimized for performance and distributed storage [11].

Other innovations come from the way Big Data affects everybody's life, as Big Data is fundamental to the functionality of various popular applications: search engines, translation tools, social media (e.g. for friend finders), online shops (e.g. for product recommendations) and so on.

In this paper we will point out major challenges computer science education will have to deal with, when considering the new aspects arising from the outlined developments into teaching. First, Big Data is discussed in terms of its innovations to data management. Then, we will provide the educational context by analyzing the state of research on databases and Big Data in computer science education. On this basis, we will derive and discuss the relevance of changes arising from these developments and we will describe the challenges within computer science education.

2 Managing Big Data

The term Big Data describes the management and analysis of large amounts of data with high complexity and varying structure. When dealing with Big Data, typical database systems are reaching their limit: even if they can handle the large ***volume*** of data, there will remain at least two even more complex problems. The high changing rates of these data (***velocity***) require highly responsible databases, ideally without any blocking operations. Additionally, the ***varying structure*** of data prevents database administrators of defining a data schema. Such a schema is needed for ensuring data consistency when using common relational databases (RDBMS). As Laney [18] summarizes, Big Data can be characterized by the three Vs "volume", "velocity" and "variety" (cf. fig. 1).

These challenges when using the more or less standard RDBMS led to the development of non-relational database management systems, commonly known as NoSQL databases. This term is meanwhile used as a generic name for all non-relational databases and is interpreted as "not only SQL"[1] [11]. The main characteristics of NoSQL databases are [11]:

- a non-relational data model,
- distributed and horizontally scalable,
- schema-free or only with weak restrictions on schema.

Also, in contrary to traditional databases, the NoSQL databases do not guarantee the ACID properties (**A**tomicity, **C**onsistency, **I**solation, **D**urability) [12]. Instead, they are described as **B**asically **A**vailable, **S**oft-state, **E**ventually consistent (BASE) [11]. Such databases offer fast access to and high availability of data, but cannot guarantee failure safety, as data are only written to permanent storage periodically. Also, consistency of data cannot be ensured, because typically NoSQL databases do not enforce a defined data schema in order to reach high performance and partition tolerance. According to the CAP theorem [4], it is not possible to guarantee the three properties availability, consistency and partition tolerance at the same time. Only two of these three characteristics can be ensured concurrently, cf. fig. 2. So, traditional RDBMS mainly focus on consistency and availability, and therefore need to drop partition tolerance—they are

[1] In the original work on the NoSQL database [23], the term NoSQL was used in the way "no usage of SQL", because this RDBMS did not support SQL.

described as consistent-available (CA) according to the CAP theorem. In contrast, the NoSQL databases are typically optimized for distributed databases with high performance and therefore must drop consistency (so they are described as available and partition-tolerant (AP)). The third type—consistent and partition-tolerant (CP)—is less widely spread and not associated to whether RDBMS or NoSQL, as both types can be found in this category. In contrast to these developments, the main topic in today's database education is teaching the concepts of (relational) database systems and the knowledge for dealing with such systems. In order to fill up this gap between CS and database education, the influence of these developments on CS education has to be discussed.

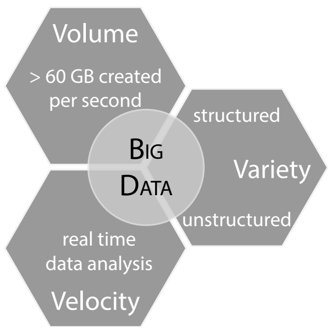

Fig. 1. Three Vs of Big Data

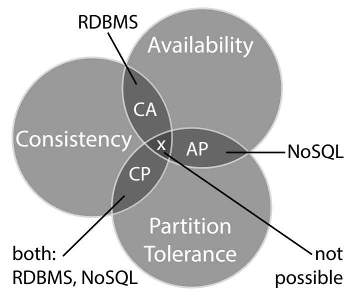

Fig. 2. CAP theorem

While the topic databases plays an important role in secondary school education, with the growing impact of data and its analysis this topic needs to be expanded. *Data management* describes the processes of storing and retrieving data from databases as well as planning, organizing and utilizing these methods [3]. This includes "practices that control, protect, deliver and enhance the value of data and information assets" [8] which can be summarized as data safety, data privacy and data analysis. Aspects of data management are often already considered in computer science curricula, e.g. under the term "information management" in the ACM K-12 curriculum [24]. So, data management offers the chance to bring these aspects together in one topic and to include several additional aspects.

3 Databases and Data Management in Computer Science Education

The topic databases, which comprises the efficient storage, management and retrieval of data, is central to secondary computer science education in several countries (e.g. Germany [16], Austria [6, 7]). The ACM curriculum for K-12 computer science [24] considers data management in the context of information management: database systems, data modeling and the relational model, query languages, data mining, hypertext and hypermedia, digital libraries. The UNESCO/IFIP informatics curriculum for secondary schools [25] considers databases in two ways: Working with a database and designing a database. Numerous publications especially in the late 1980s / early 1990s, discuss the relevance of the topic for computer science at school (e.g. [26]). However, even though the importance and influence of databases and data management is growing rapidly, these topics were hardly discussed in computer science education within the last years: Antonitsch [1] proposes databases as a topic in the context of information retrieval by questioning the common approaches for database instruction. Other publications on these topics mostly propose and/or discuss tools for enhancing database education and in particular for supporting the teaching of SQL (for example [14], [20]). The shortage of research in this field may be explained by the few changes in the concepts of (relational) databases since they were established by Edgar F. Codd. The main changes affected platforms; these are relevant for the DBMS and for tools for using them, but not for the concepts of databases themselves or for educational purposes. In contrary to the discussion in the data management community, new models for database, like the NoSQL databases, have not yet been discussed in the context of database education at school. The same applies to higher education, where recently claims are made to consider the emergence of NoSQL databases in the curricula [19].

In addition to these aspects from computer science, CS education is also affected by social and ethical implications: while current teaching only considers data privacy and safety as marginal topics, mainly integrated with other topics like databases, nowadays the relevance of these aspects is changing tremendously. As new possibilities are opened up for dealing with large amounts of data and to gain new information by analyzing them, also new threats are emerging. Especially, there is a strong impact on data privacy and data security. Furthermore, everyone needs to deal with data in daily life, as the impact of data and data-driven applications is continuously increasing. Hence, there are clear differences between the requirements coming from daily life and current teaching practice.

4 Challenges for Computer Science Education

As pointed out in the previous section, a gap between curricula / teaching practice and the state-of-the-art of computer science emerges from the recent developments. Moreover, there are also recognizable differences between the demands of daily life and current CS education. In acknowledge to this issues, organizations such as the German Informatics Society start claiming that “Big Data is a topic for education” [13]. However, by considering Big Data for teaching, computer science education is faced with several challenges, which include revising the curricula for the topics databases and data management in a critical way. Key questions that need to be considered are:

- Which key concepts and principles are sustainable?
- Which topics may become outdated or need to be discussed in a different light in the future?

– Which aspects does Big Data add?
– What does everyone need to know about and for dealing with Big Data?

Therefore, we will hereafter discuss the major challenges, with which computer science education will have to deal when considering the new aspects coming from the described topics in teaching of data management. This analysis does not focus on a specific curriculum but applies generally to CS education, because while typical curricula strongly differ in many aspects, there is a clear agreement on important aspects in database education.

4.1 Discussing the Relevance of Database Concepts

There is agreement that general school education needs to emphasize persistent concepts over skills that are only usable for a short time span. For computer science, various sets of criteria for selecting topics with general educative value have been proposed. The most influential ones are the "Great Principles of Computing" by Denning [10] and the "Fundamental Ideas of Computer Science" by Schwill [21]. One typical criterion is, that topics for general education need to be relevant in various contexts. This criterion is named "horizontal criterion" by Schwill, "broadly influential" by Denning. It prevents special aspects, which are only relevant in few contexts, from being considered as relevant for general education. Another criterion points out that only those concepts shall be taught that are and will be historically relevant. Due to the recent developments in data management, e.g. Big Data, the concepts and principles currently taught in school are now challenged. For example, RDBMS use an explicit data schema [12], while NoSQL databases use an implicit one [11]. By forcing data into matching the data schema defined by the user/administrator, RDBMS ensure consistency. This approach leads to long outage times if the schema has to be modified, because then data has to be adopted to the new schema immediately. Since today data often vary in structure over time, this would happen too frequently. So, NoSQL databases prevent such outages by not-enforcing a schema (so they have only an implicit one), but therefore they must deal with inconsistent data. Hence, it has to be examined whether always enforcing a data schema holds for a key concept of databases.

Another challenge is the altered significance of redundancy. While in RDBMS data are typically stored in a in a normalized way in order to avoid redundancies [12], the concept of redundant storage is used intentionally in NoSQL databases [11] in order to store data in one place instead of spreading it over multiple tables. Hereby, reading data from NoSQL databases is accelerated, because in comparison with RDBMS no join operations are needed. Therefore, the challenge for computer science is to rethink the idea of redundancy from something that generally should be avoided to a method that is applied in order to achieve certain specific goals. While redundancy in the context of RDBMS has been only discussed in the context of database normalization, it should be discussed, if the concept of redundancy holds for a fundamental idea of data management.

Even the small selection of differing concepts between RDBMS and NoSQL databases described above, leads to important considerations, as differing aspects cast a new light on meeting the criteria for general educational topics: until Big Data became popular, it seemed obvious that the main concepts of relational databases are as well main concepts of databases in general, because there were mainly RDBMS. Nowadays, it is necessary to differentiate between relational and non-relational databases, so it is easily recognizable, that for example the concept of strictly normalizing data schemata is only fundamental to relational databases. Therefore, the concepts concerning databases in general, and not only concerning either RDBMS or NoSQL databases, have to be detected. When looking at the described fundamental ideas and great principles, it is obvious that only the concepts concerning the topic databases in general should be selected for teaching in order to ensure teaching of general educative topics.

In conclusion, current considerations on which concepts of databases should be part of computer science education need to be reconsidered in the light of Big Data and data management by posing the question *"Which concepts are fundamental to databases?"*.

4.2 Involving Big Data Examples into Teaching

Even though the intention to use examples of students' everyday life by discussing offline databases held true until recent years, in future web databases and large data collections will become more common to students and affect their daily life more than offline databases. Furthermore, such databases offer the chance to use public and up-to-date (open) data sources and hence open up the possibility for students to gain insight into data by using analysis methods themselves. At the moment, database education especially discusses offline databases, for example a member database of a sports club, of products the students (or fictive persons) bought, or of music groups [15]. Some main aspects of databases can hardly be demonstrated by such small examples, especially because some of them, like the product database, could be equally addressed in a spreadsheet application. Since small databases are often used as local copy on each computer, characteristics such as the multi-user capabilities of databases cannot be realized. Taking into account the impact of Big Data, this problem will be even intensified, because using small examples will not be possible for illustrating the main characteristics of Big Data and data analysis, because they are particularly based on the high number of available data. Especially, by using small data sets it is not possible to obtain satisfying analysis results of statistical relevance. According to the mathematical law of large numbers, scattering is minimized when analyzing large amounts of measures (data), while results are strongly distorted when only analyzing small amounts of data.

Also, the changes in concepts of databases strongly affect selection of examples for teaching. While typical examples in database education are often chosen in order to discuss databases considering the relational data model, normalization, consistency and so on, these criteria for selection will clearly differ when considering Big Data as topic for education. In the future the examples discussed in class need to be able to clarify the possibilities of data analysis in order to enable students to recognize main concepts and methods of data analysis.

Summarizing, small data sets will not be sufficient in future data management education anymore; instead students need to be enabled to work with large amounts of data themselves, so that they can recognize the advantages of databases and data management. This also implies storing the data on a central server, because of the quantity of data. Therefore, two main functionalities of databases are emphasized implicitly: storing data in a central place and accessing them concurrently with many clients [12]. This results in the challenge to determine and open up the possibilities to use the available data, mainly coming from open data projects, in class. Hereby, teachers are enabled to provide examples that better fit the usage of databases in economy, but at the same time, students are enabled to analyze data themselves and so recognize which kind of data may be obtained by combining different data sets.

4.3 Teaching Data Analysis for Understanding Data Mining

Nowadays, data are a valuable resource. They often contain more information than visible at a first glance. Hence, data mining is one of the most rapidly growing business factors as well as a threat to data privacy. In order to provide students with an understanding of the value of data as well as the possibilities and threats of collecting, processing and evaluating personal data, this aspect of data management should be added to the curriculum. According to the World Economic Forum, "personal data will be the new 'oil' - a valuable resource of the 21st century" [27]. Just like any other resource, raw data must be made usable before being able to benefit from them. For data, this is done by analysis: unnecessary data are stripped, remaining data are aggregated, combined with other data, and so on until a final result is usable. Yet in typical computer science education examples, there is an emphasis on structuring data, storing them in relational databases and retrieving them efficiently. However, only limited new information is extracted from these data. A reason for this may be that data are mainly retrieved from the database in the same way they were stored before. Often these data are supplemented with results of aggregate functions, but they only add few new information. In contrast, data analysis methods offer the chance to discover more information and coherences by using the main methods for data analysis [17], like:

- *Clustering* sorts new data sets into related groups (clusters) according to determine similarities. For example, in social networks clusters of users sharing the same interests may be determined in order to propose interesting groups to a user.
- *Classification* as well as clustering aims for finding patterns in the data. However, by classification methods the categories are defined beforehand. Emphasis is put on determining characteristics and attributes of these categories.

– *Association* is used for determining interdependent occurrences of certain events. This allows for finding inferences for interdependencies between data sets. Such inferences are formulated as "if-then-relations" and can be used in order to predict e.g. future behavior after occurrences of certain events.

Summarizing, a challenge for computer science education is to enable students to understand data mining processes as well as using data analysis methods in order to acquire new information. Therefore, concepts and topics of data analysis need to be analyzed with respect to the criteria for selecting topics for general education described above [10], [21].

4.4 Changes in the Relevance of Data Modeling

Key aspects of data analysis are the methods for structuring data, for interpreting them and for recognizing their important aspects. Especially for structuring, another fundamental idea of computer science is involved: (data) modeling. In current teaching, static data modeling is typically applied in order to define data structures before inserting the data into a database. In the future, the objective of modeling will change: When working with data in order to analyze them, typically prestructured data will be used, for example coming from Open Data projects. Even if they are not structured explicitly, at least an implicit structure is defined by the way they are stored. Therefore, structuring such data by hand is not necessary anymore. But when using data analysis methods, static data modeling will still be relevant for providing an overview over data sets. Also, especially when combining different data sets, it will be necessary to visualize data structures as in most cases these data sets are not aligned for being combined, because at the moment of structuring, future uses are often unknown.

Additionally, when working with NoSQL databases, the relevance of static data modeling will decrease, because these databases do not enforce a data schema. Instead, data is stored less spread, but rather coherent in ideally only one database table/collection often structured exactly the way they are received from the data sources. This is the case, as Big Data does not fit into a defined schema, because of its varying structures. In particular, by saving data in a coherent way retrieving them from the database is accelerated.

Not only static data modeling, but also dynamic modeling techniques are involved when discussing data analysis. In particular, as data analysis consists of complex processes and sequences of operations, the analysis process can clearly be visualized and clarified by using sequence diagrams. By discussing data analysis in this way, the recurrence of the typical methods and operations of data analysis is clearly visible. Therefore, these techniques would be emphasized in comparison to only using a data analysis tool without discussing the methods and operations used. In addition, sequence modeling also has a strong relevance in daily life, as visualizing and discussing sequences of operations is necessary in various fields, not only in computer science. Therefore, emphasizing sequence modeling in data management education also raises its general educational value.

Summarizing, this leads to another challenge: At the moment, data modeling is the main modeling technique when dealing with data in CS education. When considering Big Data and data analysis, on the one hand the usage of data modeling changes from mainly structuring data into also visualizing prestructured data. On the other hand, also process / sequence modeling will increase in importance for planning data analysis processes.

4.5 Sharpening the View on Data Privacy

Since Big Data strongly affects the topic "data privacy", it is not surprising that several publications discuss aspects of Big Data and data analysis in this context. These topics are commonly taught in the context of database education. For example, the "simulation game data privacy"[2] is a learning environment in which students may discover threats for data privacy on their own. Such material typically cover fundamental data privacy problems. But problems caused by Big Data, which strongly differ, generally are not yet taken into consideration. Therefore, data implicitly given to the vendors, like when clicking on a web page link, increases in value. In this context, actual material on data privacy needs to be reviewed and revised in order to cover new challenges for this topic.

Also, new possibilities raise new threats in this field. Today, data is captured continuously in daily life, for example obviously by smartphones and tablets, but also in a more hidden way by cars, smart electricity meters, and such. With the ability to analyze such large amounts of data, large parts of daily life may be reconstructed and users profiles may be generated. By combining information from different sources, gathering private data like friendship relations, hobbies, habits and so on, is possible. This applies not only for obviously private data, like friendship relations, or for personal data like name and birth date, but also for data that seem harmless at the first glance: particularly meta-data, like date and time a text message was sent on or the position a photo was taken at.

When discussing the threats for data privacy, methods for preventing such analysis of own data should be pointed out. As with modern devices, (web) applications and services, data are often collected in a hidden way, a first challenge is to recognize that a specific application/service/device might collect data. Even when suspecting such a collection of data, it is hardly possible to prevent it other than not using the application. Although, in certain cases it is possible to distort data analysis, for example by using pseudonyms or by entering incorrect data. Thereby, no correct conclusions can be drawn from the data analysis, when intentionally wrong data are spread in order to prevent differing between genuine and false data, or when consistently using pseudonyms. But even with pseudonymization, it cannot be ensured that no private data may be recognized by data analysis: for example, when AOL released a strongly pseudonymized set of search results in 2006, it was possible to discover contact data of persons contained in this data set by using simple analysis methods [2].

[2] http://www.informatik-im-kontext.de/index.php/entwuerfe/planspiel-datenschutz-2-0, in German (last retrieved: 20th April 2014).

As discussed in several publications, CS education also should raise students awareness on data protection and privacy, so these topics are part of various curricula and educational standards (like the CSTA K-12 Computer Science Standards [22] or the German Educational Standards for Computer Science [5]). In current curricula, these aspects are often integrated with other topics, especially databases, and thus mentioned only marginally. This cannot satisfy the relevance of these topics, neither at the moment nor in the future when privacy problems will become even more important because of Big Data. This leads to the challenge that the relevance of the topic data privacy as well as material for teaching this topic at school need to be revised in order to fit new requirements.

5 Discussion

It is a grand challenge for computer science education on the one hand to aim for teaching of long lasting concepts and principles of computer science, and on the other hand to be consistent with current scientific developments. In the field of data management, traditional concepts and teaching examples are challenged by recent developments, such as Big Data, NoSQL and Open Data. By considering these in data management education, various new and motivating aspects can be found for teaching—with strong references to daily life of the students.

Using these new possibilities in data management education will clearly change the face of this topic, which is currently limited to concepts of databases and their application. In contrast, new concepts call for a stronger emphasis on broader aspects of data management, e.g. still data storage (as in databases), but in combination with data usage and data analysis. This will provide even more relations to students' daily life, especially by being educated in dealing with their own (personal) data. In order to take the chance of involving such new possibilities, several changes in CS education curricula will be required. Therefore, it is necessary to take a more in-depth look on these aspects for being able to evaluate which changes are promising and which are not.

Also, discussing Big Data in class allows a view on modern or newly emerging professions, like the "data scientist" [9], which involves aspects from computer science but also from mathematics and statistics. By discussing such fields of applied computer science, also the view on the field of CS may be sharpened: a typical bias on computer science is, that it mainly consists of programming. But by having a look on such interdisciplinary fields like data analysis, it is obvious that computer science involves more aspects than writing programs.

Nowadays everybody is affected by this field of computer science, because Big Data is being collected and analyzed in many different systems of daily use. Therefore, considering Big Data as topic in computer science education also enables students to form better founded opinions on actual topics in society, like on (early) data retention or on programs of intelligence services like the NSA's PRISM program. While it is commonly known, that in such projects huge amounts of data are collected, often the consequences remain vague. At this point, computer science education will help understanding possibilities of analyzing such (big) data, as well as the related consequences.

References

1. Antonitsch, P.: Databases as a Tool of General Education. In: Mittermeir, R. (ed.) Informatics Education The Bridge between Using and Understanding Computers, Lecture Notes in Computer Science, vol. 4226, pp. 59–70. Springer Berlin Heidelberg (2006)
2. Barbaro, M., , Zeller, Jr., T.: A Face Is Exposed for AOL Searcher No. 4417749. http://www.nytimes.com/2006/08/09/technology/09aol.html (August 2006)
3. Bodendorf, F.: Daten- und Wissensmanagement [Data and Knowledge Management]. Springer-Lehrbuch, Springer (2005)
4. Brewer, E.: CAP twelve years later: How the "rules" have changed. Computer 45(2), 23–29 (2012)
5. Brinda, T., Puhlmann, H., Schulte, C.: Bridging ICT and CS: Educational Standards for Computer Science in Lower Secondary Education. In: Proceedings of the 14th Annual ACM SIGCSE Conference on Innovation and Technology in Computer Science Education. pp. 288–292. ITiCSE '09, ACM, New York, NY, USA (2009)
6. Bundesministerium für Unterricht, Kultus und Kultur: Lehrplan der AHS-Oberstufe, Fach Informatik, Pflichtgegenstände [Curriculum for the Austrian Secondary School, Subject "Informatics", compulsory topics], http://www.bmukk.gv.at/medienpool/11866/lp_neu_ahs_21.pdf
7. Bundesministerium für Unterricht, Kultus und Kultur: Lehrplan der AHS-Oberstufe, Fach Informatik, Wahlpflichtgegenstände [Curriculum for the Austrian Secondary School, Subject "Informatics", compulsory topics], http://www.bmukk.gv.at/medienpool/11876/lp_neu_ahs_21.pdf
8. DAMA International: The DAMA Guide to the Data Management Body of Knowledge - DAMA-DMBOK. Technics Publications, LLC, USA (2009)
9. Davenport, T.H., Patil, D.J.: Data scientist: the sexiest job of the 21st century. Harvard business review 90(10), 70–77 (2012)
10. Denning, P.J.: Great Principles of Computing. Commun. ACM 46(11), 15–20 (2003)
11. Edlich, S., Friedland, A., Hampe, J., Brauer, B., Brckner, M.: NoSQL [in German]. Hanser, Carl Gmbh + Co. (2011)
12. Elmasri, R., Navathe, S.: Fundamentals of Database Systems. ADDISON WESLEY Publishing Company Incorporated (2011)
13. GI (Gesellschaft für Informatik e.V.): Handlungsempfehlungen an die politischen Akteure [Recommendations for action at the Political Actors] (Big Data Days). http://www.gi.de/fileadmin/redaktion/Hauptstadtbuero/Handlungsempfehlungen.pdf (2013)
14. Grillenberger, A., Brinda, T.: eledSQL: A New Web-based Learning Environment for Teaching Databases and SQL at Secondary School Level. In: Proceedings of the 7th Workshop in Primary and Secondary Computing Education. pp. 101–104. WiPSCE '12, ACM, New York, NY, USA (2012)
15. ISB (Staatsinstitut für Schulqualität und Bildungsforschung): Informatik am Naturwissenschaftlich-technologischen Gymnasium, Jahrgangsstufe 9 (Handreichung) [Informatics at the Scientifiy-Technological Secondary School in Bavaria (Recommendations)] (2007)

16. ISB (Staatsinstitut für Schulqualität und Bildungsforschung): Lehrplan für das Gymnasium in Bayern, Fach Natur und Technik [Curriculum for the Bavarian Secondary School, Subject "Informatics", Scientific & Technical Branch] (2009)
17. Kemper, A., Eickler, A.: Datenbanksysteme: eine Einführung [Database Systems: An Introduction]. Oldenbourg (2006)
18. Laney, D.: 3D Data Management: Controlling Data Volume, Velocity, and Variety. Tech. rep., META Group (February 2001)
19. Luukkainen, M., Vihavainen, A., Vikberg, T.: Three Years of Design-based Research to Reform a Software Engineering Curriculum. In: Proceedings of the 13th Annual Conference on Information Technology Education. pp. 209–214. SIGITE '12, ACM, New York, NY, USA (2012)
20. Sadiq, S., Orlowska, M., Sadiq, W., Lin, J.: SQLator: An Online SQL Learning Workbench. In: Proceedings of the 9th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education. pp. 223–227. ITiCSE '04, ACM, New York, NY, USA (2004)
21. Schwill, A.: Fundamental ideas of computer science. Bull. European Assoc. for Theoretical Computer Science 53 (1994)
22. Seehorn, D., Carey, S., Fuschetto, B., Lee, I., Moix, D., O'Grady-Cunniff, D., Boucher Owens, B., Stephenson, C., Verno, A.: K12 Computer Science Standards. Computer Science Teachers Association, Association for Computing Machinery (2011)
23. Strozzi, C.: NoSQL: a non-SQL RDBMS. http://www.strozzi.it/cgi-bin/CSA/tw7/I/en_US/nosql/Home%20Page (1998)
24. Tucker, A., Deek, F., Jones, J., McCowan, D., Stephenson, C., Verno, A.: A Model Curriculum for K–12 Computer Science. Final Report of the ACM K-12 Task Force Curriculum Committee, CSTA (2003), https://www.acm.org/education/education/curric_vols/k12final1022.pdf
25. UNESCO / IFIP: Information and Communication Technology in Secondary Education (2000), http://wwwedu.ge.ch/cptic/prospective/projets/unesco/en/
26. Witten, H.: Datenbanken - (k)ein Thema im Informatikunterricht? [Databases - No Topic in Computer Science Education?]. LOG IN 2 (1994)
27. World Economic Forum: Personal Data: The Emergence of a New Asset Class. http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf (2011)

All electronic sources were at last retrieved on 17th April 2014.