



# Exploring Information Systems Curricula

## A Text Mining Approach

Patrick Föll · Frédéric Thiesse

Received: 19 November 2019 / Accepted: 31 March 2021 / Published online: 15 June 2021  
© The Author(s) 2021

**Abstract** The study considers the application of text mining techniques to the analysis of curricula for study programs offered by institutions of higher education. It presents a novel procedure for efficient and scalable quantitative content analysis of module handbooks using topic modeling. The proposed approach allows for collecting, analyzing, evaluating, and comparing curricula from arbitrary academic disciplines as a partially automated, scalable alternative to qualitative content analysis, which is traditionally conducted manually. The procedure is illustrated by the example of IS study programs in Germany, based on a data set of more than 90 programs and 3700 distinct modules. The contributions made by the study address the needs of several different stakeholders and provide insights into the differences and similarities among the study programs examined. For example, the results may aid academic management in updating the IS curricula and can be incorporated into the curricular design process. With regard to employers, the results provide insights into the fulfillment of their employee skill expectations by various universities and degrees. Prospective students can incorporate the results into their decision concerning where and what to study, while university sponsors can utilize the results in their grant processes.

**Keywords** Text mining · Topic modeling · LDA · IS education · Curriculum research

## 1 Introduction

Hardly a week goes by without the digitalization of companies and the workplace being highlighted in the mass media – a reflection of the fact that the digital transformation of the economy has brought about a profound change across industries and the labor market. Skills that are in demand today may become obsolete tomorrow. As future employees, the current generation of students will have to cope with the highly dynamic nature of digital work environments, which makes lifelong learning and up-to-date skills ever more important. As providers of the corresponding study programs, universities and other institutions of higher education (IHE) are under constant pressure to adapt their curricula to rapidly evolving demand on the part of students and employers alike. To address the challenges posed by digital transformation, Jung and Lehrer (2017), for example, present a model curriculum for Information Systems (IS) programs in German-speaking countries proposed by a working group of domain experts from academia and business on behalf of the Academic Commission for Business Informatics (*Wissenschaftliche Kommission Wirtschaftsinformatik*, WKWI) of the German Academic Association for Business Research (*Verband der Hochschullehrer für Betriebswirtschaft*, VHB) and the Department of Business Informatics of the German Computer Science Society (*Gesellschaft für Informatik*, GI). Similar to other model curricula, in particular those developed by the Association for Computing Machinery (ACM) (Topi et al. 2010) and the Association for Information Systems (AIS) (Topi et al.

---

Accepted after two revisions by Natalia Kliewer.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s12599-021-00702-2>.

---

P. Föll (✉) · F. Thiesse  
University of Würzburg, Sanderring 2, 97070 Würzburg,  
Germany  
e-mail: patrick.foell@uni-wuerzburg.de

2017), the guidelines set forth by Jung and Lehrer (2017) “aim to support schools, departments, and individual faculty members in charge of curriculum development and to assist students in program and career choice”.

While model curricula aim to define a standardized set of basic educational contents, the contents of actual study programs in today’s higher education market exhibit a wide variety of different thematic profiles. For example, according to the German Rectors’ Conference (2016a), 247 IS study programs leading to a bachelor’s or master’s degree were offered by German IHEs in 2016. Beyond their sheer number, the programs differ substantially in their focus, which may be attributed, among other things, to specific local boundary conditions, but also to the individual programs’ efforts to differentiate themselves from competing offerings in order to gain visibility and attract more applications from the best students. Due to their abundance, it is hardly possible to compare the large number of curricula efficiently and in a reasonable time using conventional methods like qualitative content analysis. However, such a comparison would be of interest to various stakeholders concerned: (1) IHE administrations may benefit from a comparison of the modules, contents, and goals of similar study programs, which would provide an objective assessment of their curriculum and, thus, an opportunity for improvement and differentiation; (2) employers can focus their search for graduates on those universities which best fulfill their desired qualifications; (3) the data could serve as an aid to prospective students in selecting a university on the basis of their personal preferences and interests with regard to the subject of study; (4) policymakers and university sponsors can more easily obtain knowledge for academic landscape planning and the evaluation prior or future funding measures.

Against this backdrop, the present study considers quantitative content analysis in the form of text mining techniques as an approach to address the needs of the aforementioned stakeholders. The underlying hope and expectation is that text mining might enable a largely automated evaluation of the study contents documented in module manuals with reasonable effort, simple repeatability, and good scalability. While text mining has already been put to practical use in other related areas (e.g., the analysis of business skill expectations), the investigation of curricula has not been a field of application in prior research. In order to fill this gap in the literature, we employ topic modeling techniques to identify, analyze, and compare the skill areas offered by universities to students in their study programs. For illustrative purposes, we consider the example of the IS curricular landscape in Germany. The contribution we thereby aim to make is twofold. On the one hand, we propose a procedure built on the foundation of well-established text mining techniques,

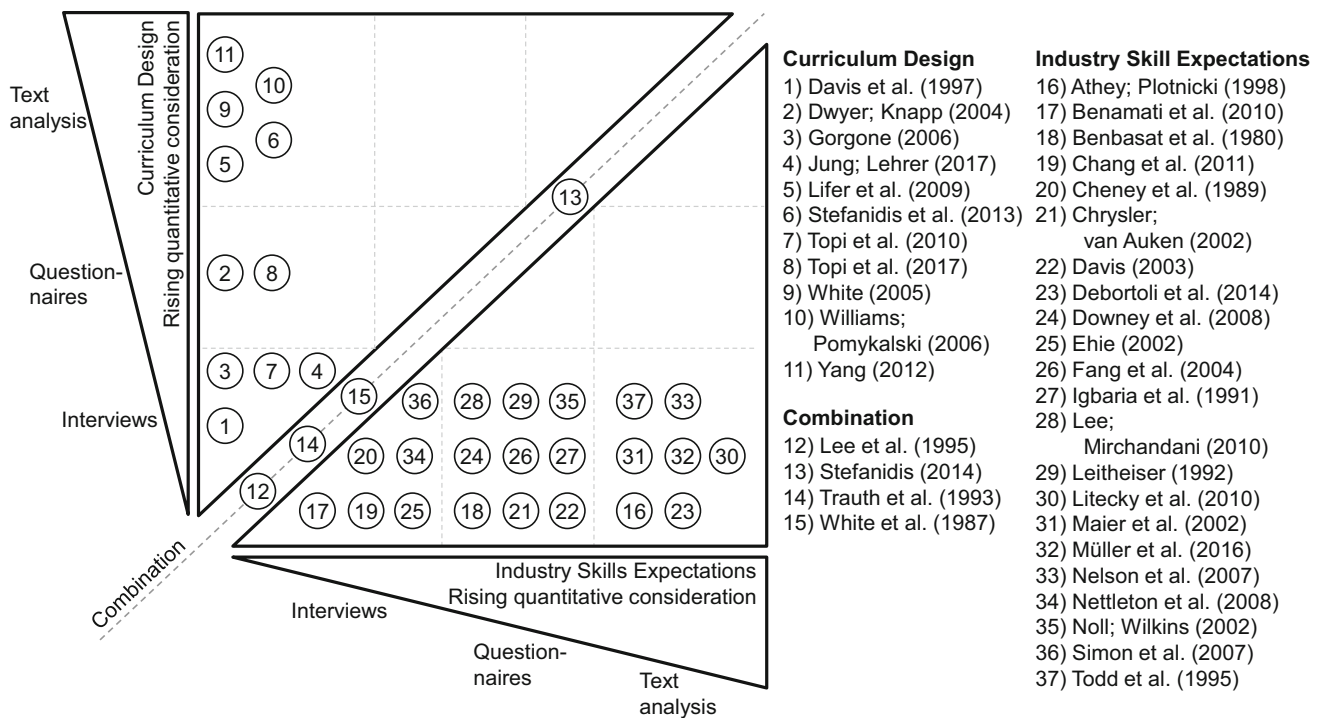
which was specifically designed for the analysis of study curricula from arbitrary academic disciplines. As we argue in the following, this procedure stands out over traditional manual techniques due to its superior scalability and reproducibility of results, which would enable curriculum designers to react more quickly and more often to the dynamics of a constantly changing environment. On the other hand, we provide detailed empirical insights into the diversity of the IS curricular landscape in Germany, which indicate how different IHEs and degree programs distinguish themselves from each other in terms of the skill offerings found in their curricula.

## 2 Related Work

IS and adjacent disciplines are subject to continuous change because technologies, economic conditions, and job-related skills evolve rapidly (Lee et al. 1995; Gallivan et al. 2004). Technological changes can take the form of, for example, new hardware and emerging platforms (e.g., smartphones) or new programming languages and/or paradigms. Digitalization in all its forms – such as the increase in large data sets, as well as their processing and evaluation, or novel ways of doing business that may entail radical innovation in processes and entire business models – also confronts the discipline with fresh challenges. The consequence of this transformation is new needs and requirements for IS students, which should be reflected in the curricula to prepare graduates for this highly dynamic environment. In order to guarantee the successful integration of new topics into existing curricula, employee-skill research is necessary to identify and evaluate changing needs and requirements. The objective of curriculum research is to take these findings into account and incorporate them into the adjustment or redesign of curricula. We use both research strands for a first high-level categorization of prior research related to our work.

To identify related studies in the existing literature, we searched the databases Business Source Premier, ACM Digital Library, and Google Scholar for the following keywords and variations thereof: (1) information systems/IS curriculum/education design university/higher education/(under)graduate skill/competence provision; (2) information systems/technology job/work/employee/industry skill/competence requirements/expectations. For the articles found with these keywords, we conducted a forward/backward search to obtain a current overview of the state of knowledge. The outcome of this process is depicted in Fig. 1.

One group of identified studies pertains to specific guidelines for curriculum design of IS courses (Davis et al. 1997; Gorgone 2006; Topi et al. 2010, 2017; Jung and



**Fig. 1** Prior related research in the literature

Lehrer 2017) and evaluations of how these are used in reality (Yang 2012; Dwyer and Knapp 2004; Williams and Pomykalski 2006; Lifer et al. 2009). Regarding the development of guidelines, the corresponding studies mostly apply traditional data collection methods, such as interviews or questionnaires, to examine the curricular needs of IS. To evaluate the guidelines and their uses, manual content analysis is the preferred data analysis method (Stefanidis et al. 2013; White 2005). In contrast, none of the reviewed studies makes use of text mining methods based on, for example, LSA or LDA.

A second group of studies sets the focus on business skill expectations in IS-related job categories. These cover the skill requirements in fields of work characterized by innovation and technological change, which are often sources of competitive advantage (Gallivan et al. 2004; Hickson 2000; Nelson et al. 2007). Most of these studies are empirical, that is, they gather data via traditional questionnaires and interviews or through data collected from printed or online job postings. Several studies seek to determine the required skills through standardized questionnaires. Questionnaire respondents range from employers (Lee and Mirchandani 2010; Noll and Wilkins 2002; Downey et al. 2008) to employees (Chang et al. 2011; Davis 2003; Igbaria et al. 1991) to students (Fang et al. 2004; Chrysler and van Auken 2002) or faculty members (Benamati et al. (2010). Only a few studies use interviews to identify work skills. The majority of these studies draw

on managers or employers as interview subjects (Benbasat et al. 1980; Cheney et al. 1989; Ehie 2002; Leitheiser 1992; Nettleton et al. 2008; Simon et al. 2007). Early data-driven approaches present analyses of job postings in newspapers, such as that carried out by Maier et al. (2002), Todd et al. (1995) or Athey and Plotnicki (1998). It is only in recent years that researchers have started to apply text mining techniques to the analysis of online job platform data. Early examples of the latter approach are provided by Litecky et al. (2010), who analyze 250,000 IT job postings in the US using data mining, and Debortoli et al. (2014), who examine job postings related to business intelligence and big data, or Müller et al. (2016). We also contributed to this group of studies (Föll et al. 2018) by analyzing the labor market expectations regarding competencies of IS job starters.

Some studies also try to combine the IS curriculum with requirements of the job market in order to determine the correlation between the skills provided by study programs and those demanded by industry (Litecky et al. 2004). Through this combination, the studies aim to facilitate more appropriate offerings at universities in order to better prepare their students for the expectations of the job market. To achieve this goal, various methods, such as interviews (White and Lederer 1987; Trauth et al. 1993), questionnaires (Lee et al. 1995), or manual content analysis approaches (Stefanidis 2014), are used.

### 3 Methodology

As our literature review has shown, machine learning has become part of the methodological toolbox for the analysis of business skill expectations. In contrast, we are not aware of a comparable example of prior research in the area of curriculum research. This outcome of our literature review seems rather surprising, as the corresponding methods excel in a number of respects over traditional approaches, including better reproducibility and scalability as well as more objectifiable results in contrast to qualitative methods. To address this gap in the literature, we propose a methodological approach for the analysis of study programs using topic modeling techniques as depicted in Fig. 2.

Module manuals of accredited study programs in Germany from the respective IHE web pages comprise the foundation for all subsequent steps (1). Since accredited German-language study programs are the focus of the analysis, the module handbooks first pass through various filtering steps (2) until the determination of the final number of module manuals that comprise the data set for topic modeling. From these manuals, we extract the parts relevant for topic modeling (e.g., module description and title). In the next step (3), the extracted documents are preprocessed for topic modeling. Preprocessing includes,

for example, removing stop words, word stemming, and the formation of N-grams. Once preprocessing is completed, we proceed with the iterative generation of the topic model for the document corpus. With the topic model and its document-to-topic matrix (which contains the distribution of all topics to all documents, i.e., modules), analysis and comparison of the IS curricula can take place (4). In the following subsections, we describe in detail the text mining-based procedure conducted for German IS curricula. The corresponding results are then presented in Sect. 5.

#### 3.1 Data Source

As the illustrative example for demonstrating the practical application of our methodological approach, we consider IS study programs in Germany. IS as an academic discipline has decades of history in German-speaking countries, with the first distinct programs being established in the 1970s. In Germany, academic IS programs are typically offered by two types of IHE: traditional universities and universities of applied sciences (UAS). The latter usually feature a practice-oriented curriculum, while the first are more theory- and research-oriented (Standing Conference of the Ministers of Education and Cultural Affairs 2015). IHEs in Germany are mostly state-funded, but there are also privately funded institutes. Both are subject to German

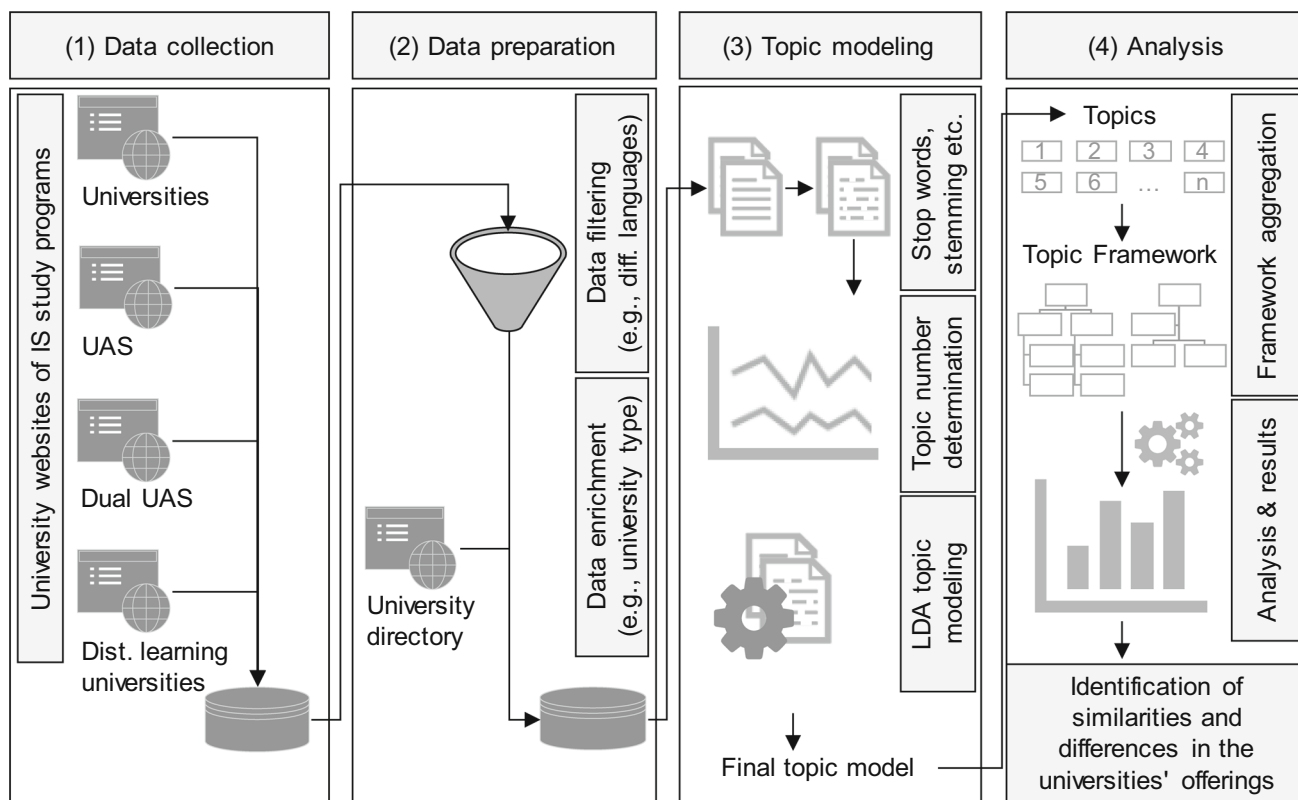


Fig. 2 Methodological approach

higher education legislation (German Rectors' Conference 2016b). As a result of the Bologna Process of 1998, the formerly single-stage German diploma degree programs were largely converted to the modularized bachelor's/master's system (German Rectors' Conference 2017).

The structure and contents of a study program are described in a module manual, study regulations, and examination regulations. The module manual contains information on the modules that make up the study program. For example, these descriptions might include the actual teaching content, module cycle, ECTS (European Credit Transfer System) points, or course type. The study regulations put the modules of the handbook in context, that is, they describe the structure of the degree program, such as compulsory and elective modules. The examination regulations provide information on the procedure and conduct of examinations, as well as their evaluation (Standing Conference of the Ministers of Education and Cultural Affairs 2015).

To ensure a certain uniformity across the study programs and universities, another important aspect of the Bologna reform was the introduction of the accreditation system. It defines standards that must be taken into account when designing a study program (German Council of Science and Humanities 2000). The overarching purpose is a uniform European qualification framework which, by designing qualification profiles and clarifying educational pathways, is intended to increase the transparency and comparability of study programs, thus making it easier for students to switch between programs and institutions (Accreditation Council 2017). This does not necessarily mean that non-accredited study programs are of lower quality, but accreditation offers a binding framework for the design of the abovementioned program-related documents, which form the body of data utilized in the approach presented. The framework was drawn up by the Accreditation Council (2013) and detailed by the individual accreditation institutions (ACQUIN 2014; ZEvA 2016; AQAS 2014). It consists of eleven criteria that must be met for a program to be accredited. Among these, *qualification goals*, *study program concept*, *study program-related cooperation*, and *transparency and documentation* are particularly decisive with regard to the quality and comparability of the module manuals of the study programs considered below. Following the directives of the Accreditation Council (2013) and the accreditors (ACQUIN 2014; ZEvA 2016; AQAS 2014), the criteria ensure the following:

- All modules that can usually be selected within a university's study program are listed (or exceptions are defined, for example, for the recognition of foreign achievements). If the university commissions other

organizations to carry out parts of its program in the form of program-related cooperation, exact details of the existing cooperation are required. In addition, an explanation must be given of how the quality of the study program is guaranteed by the cooperation.

- The qualification goals and the qualifications imparted are listed and described. They include scientific/artistic qualifications, professional skills, opportunities for social commitment, and personal development and must be described in detail.
- Study program concepts are available. The modules offered must meet the requirements of the set qualification goals; appropriate forms of teaching and learning must be available. In particular, information on the awarding of credit points according to the ECTS scheme is required. Furthermore, information should be provided concerning the scope of the compulsory and elective courses.
- The study program, progress of studies, examination and admission requirements must be documented and published. Furthermore, there should also be an indication of how often the module manual is updated.

Overall, these criteria were established with the aim of ensuring study program comparability and quality across universities. We therefore conclude that the module manuals of study programs represent a suitable data source for our research.

## 3.2 Data Collection and Preparation

Module manuals comprise the document corpus for the topic modeling procedure. The process of collecting the module manuals and extracting the modules contained therein is described in the following, as is the enrichment of the extracted module data with university information. Finally, we give an overview of the collected data.

### 3.2.1 Selection and Extraction of the Module Manuals

Of the 247 WI study programs in Germany, the accreditation criterion applies to 188 (German Rectors' Conference 2016a). In this subsection, we discuss further restrictions, which require a step-by-step filtering of the manuals. Table 1 gives an overview of each step and the resulting exclusion of some programs from further analysis.

In the first step, we had to manually locate and download each module manual from the universities' websites. Of the 188 manuals identified, 21 could not be found or were not available to non-university members. We then checked the remaining 167 module manuals for their language. Since we can only evaluate one language in the



**Table 1** Overview of module manual filtering

Step	Filtering for...	# of manuals before step	# of filtered-out manuals	% of filtered-out manuals
1	Manual availability	188	21	11
2	Different languages	167	72	43
3	Processing possibility	95	2	2

Final number of module manuals: 93

topic modeling process, we excluded all non-German module manuals. That left us with 95 module manuals, from which we excluded two from further processing, as they were copy-protected. With the final corpus of 93 module manuals (cf. Online Appendix I, available online via <http://link.springer.com>), we conducted the module extraction procedure in order to obtain the following attributes of each module: module name, module type, recommended semester, number of ECTS points and semester hours per week, assignment to compulsory or elective course area, course rotation, (entry) requirements/prior knowledge, learning objectives, learning content, examination type, form of teaching, and teaching language.

During the extraction process, we were confronted with several challenges: Although all the module manuals contain essentially the same type of information, they do not exist in a structured form that would enable us to use them directly for further processing. The manual documents are all formatted differently, which makes fully automatic data extraction into a structured database impossible. For example, not all module manuals have tables of contents or the modules are structured either by tables or by headings. For several module manuals that were somewhat similarly formatted, we were able to use a PDF parser with self-developed rules for extraction, while others had to be extracted by hand. In the following, we briefly describe the PDF parsing extraction process.

We start by importing the module manual page by page with the PDF parser and saving it in a vector. Each page of the PDF is now stored as a string in the vector. Second, we search for the strings where a new module starts – a prerequisite for this is that a module always starts on a new page. We therefore search for a pattern that must occur within the first characters of the string. We often use the term “module” as a pattern, because it frequently introduces the beginning of a new module description. After having found the beginnings of the module descriptions, we join the modules. Each vector in which we saved a PDF page now serves as the basis for a new vector with a module description as a string in it. In the next step we save the attributes of the modules to the respective vectors. The basic idea behind this step is to find the attribute names (e.g., ECTS, learning objectives) within the full module

description and then shorten the module description so that only the attribute description remains. After extracting the attribute descriptions, we order them according to the sequence in our output file and save them. Furthermore, we add an ID and a link to the module manual for each module. The overall extraction process results in a data table like that shown in the example (translated) extract in Table 2.

### 3.2.2 Data Enrichment and Overview

In order to verify whether the final sample of the IS curricula adequately covers all types and sizes of IHE and all types of degree, we enriched the data with information pertaining to these attributes. For this purpose, we used a list provided by the Foundation for the Promotion of the German Rectors’ Conference (2016). The list contains the attributes mentioned as well as further information (for instance, the years in which the universities were founded). For a better overview, we aggregated the attribute of university size into three groups according to the definition of the Stifterverband (2013):

- *Group 1* small universities (maximum 5000 students)
- *Group 2* medium-sized universities (5001 to 14,999 students)
- *Group 3* large universities (15,000 students or more)

Table 3 shows the distribution of the study programs considered in the final sample among various attributes such as size and type of university. The numbers in brackets indicate the numbers out of the full set of 188 study programs. As can be seen, we cover a wide range of different programs and attributes. Thus, we expect our results to represent Germany’s IS higher education landscape in accordance with the underlying module manual corpus.

Taking a more detailed look into the module manuals, we come away with the following picture: Each of the 93 module manuals contains an average of around 40 modules. Bachelor’s study programs have an average of around 44 modules each, which is higher than in the master’s programs, each of which has 30 modules on average. Figure 3 shows the overall distribution of the modules per

**Table 2** Example of an extracted module description

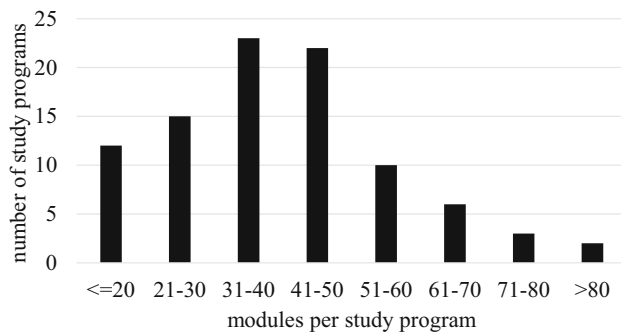
ID	47
Module manual	008
Module name	Introduction to theoretical computer science
Module type	Lecture
Recommended semester	3
Number of ECTS points	2
Semester hours per week	2
Compulsory or elective	Compulsory
Course rotation	Every semester
Prior knowledge	Basics of computer science; Object-oriented modeling; Program development techniques; Mathematics 1
Learning objectives	The students are able to: convert a conceptual solution into a formal representation that can be processed by the computer; describe the syntactical basics of programming languages and their processing on the computer; think analytically and in a structured way with the aim of systematically developing solutions to problems; name and classify fundamental concepts of computability theory, complexity theory, and undecidability
Learning content	The course is intended to yield an insight into the most important sub-areas of theoretical computer science. The goal is to develop theoretical models and descriptions of computers, programming languages, or real problems. In addition to the training of analytical thinking, basic knowledge in the areas of logic, formal languages, computability, decidability, and complexity is to be taught  Among other topics, we deal with: resolution algorithms in logic; the abstract model of a computer, the Turing machine; fundamentals of computability and complexity theory; selected problems such as the stop problem; the NP-completion and some NP-completion problems (SAT, 3SAT, node-cover); some NP-hard problems; formal languages (regular, context-free, context-sensitive); automata theory (finite automata, cellular automata); basics of petri nets
Examination type	Written exam

**Table 3** Overview of curricula data basis

Type of university	Study programs at universities		Study programs at universities of applied sciences
<b>Number</b>	21 (of 57)		72 (of 131)
<b>Type of degree</b>	Bachelor's		Master's
<b>Number</b>	67 (of 129)		26 (of 59)
<b>University size</b>	Study programs at small universities	Study programs at medium universities	Study programs at large universities
	( $\leq 5000$ students)	(5001–14,999)	( $\geq 15,000$ students)
<b>Number</b>	31 (of 55)	46 (of 88)	16 (of 45)

study program. A total of 3752 modules were collected. If we compare the module average of the examined module manuals (around 40 modules) with the average of the excluded module manuals (around 53 modules), we see a higher average in the excluded module manuals. This is due to the fact that we had to exclude more module manuals from larger universities than from smaller. Larger universities usually offer their IS students a wider spectrum of electives from their general study offerings. The

implication is that not only did we look at fewer study programs, but also the total number of modules considered decreased more than the number of module manuals. Nevertheless, we are confident that our corpus consists of a robust number of module manuals and should – by covering half of the accredited IS curricula in Germany – deliver meaningful findings regarding the IS education landscape in Germany. It is by now, to the best of our



**Fig. 3** Modules per study program

knowledge, the largest study on German-language IS curricula.

Table 4 gives an overview of the different module types and the distribution of compulsory and elective modules. The types of each module are aggregated to the five module-type categories shown. The reason that there are more modules for final theses than study programs considered is that some module manuals include thesis and colloquium as separate modules. We were not able to assign 29 of the modules to any of the categories (e.g., orientation tutorial) and for 447 modules, no information could be determined regarding their status as either compulsory or elective.

### 3.3 Data Analysis via Topic Modeling

As we consider unstructured data, for the next step we rely on the use of text mining methods to process the data. Integrating our task definition into the framework of Miner et al. (2012) led us to the field of concept extraction. The text mining methods there are summarized under the term “topic modeling”. The basic idea behind topic modeling is that every document consists of latent topics, which are characterized by specific word allocations. All topic modeling methods try to identify topics through unsupervised learning. One of the most common topic modeling methods is the latent Dirichlet allocation (LDA) method of Blei et al. (2003). We chose LDA because it is an (1) established and (2) generic model. It should therefore be suited to covering the differences in content and appearance among the modules in our document corpus (Lee et al. 2010; Alghamdi and Alfalqi 2015). LDA is a Bayesian model in which each word in a document is assigned to one

or more topics. This contrasts with traditional classification or clustering methods, whereby a data point is assigned to exactly one category: probabilistic topic models allow documents membership in various categories with differing shares (Schmiedel et al. 2018). Schmiedel et al. (2018) give a concrete example of the topic modeling idea by explaining that “the co-occurrence of words like sunshine, temperature, wind, and rain in a set of newspaper articles can be interpreted as a marker for a common topic of these articles, namely weather”. Following DiMaggio et al. (2013), they state that “topic modeling algorithms like LDA take a relational approach to meaning in the sense that co-occurrences of words are important in defining their meaning and the meaning of topics” (Schmiedel et al. 2018).

The approach rests on three basic assumptions as prerequisites for its implementation (Blei 2012):

- The order of the words in a document is not important; they are a “bag of words”.
- The order of the documents within the corpus is not important.
- The number of topics is known.

Blei et al. (2003) describe LDA as shown in Fig. 4. The outer box represents an iteration over all documents ( $M$ ); the inner box represents an iteration over the choice of topics ( $z$ ) and words ( $w$ ) within a document ( $N$ ). Latent random variables are shown as white nodes; the node of the words which are observable as the only variable is shown in grey. The topic-document distribution  $\theta$  is (like the not-shown topic-word distribution  $\varphi$ ) a Dirichlet distribution (Steyvers and Griffiths 2007), which is used to assign the words of the documents to the different topics (Blei 2012). The Dirichlet distribution of the topic-document distribution and the topic-word distribution are controlled by the hyperparameters (Dirichlet priors)  $\alpha$  and  $\beta$ . Their parameterization is, like the definition of  $K$  – the number of topics – of great importance. We describe their parameterization for the present paper in Sect. 3.3.2.

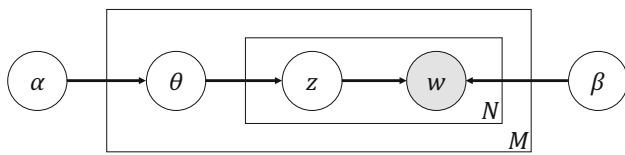
With Fig. 4 and its description in mind, Blei et al. (2003) describe the generative LDA process as follows:

- (1) Choose the topic-document distribution  $\theta \sim Dir(\alpha)$
- (2) Choose the word-topic distribution  $\varphi \sim Dir(\beta)$
- (3) For each word in a document
  - a. select a topic  $z \sim Multinomial(\theta)$
  - b. choose a word  $w \sim Multinomial(\varphi)$

**Table 4** Overview of the manuals’ module types

Module category	Lecture and tutorial	Seminar/project	Internship	Thesis	Other
<b>Number</b>	3027	530	62	104	29
<b>Area</b>	Compulsory	Elective	Not specified		
<b>Number</b>	2006	1299	447		





**Fig. 4** Modules per study program

In order to determine the underlying latent topics for a given document collection with its visible words, the process of inference is necessary. This is due to the fact that it is not the topic-document nor the topic-word distribution but rather the document corpus that is given. Figure 5 shows schematically the relationship between the document corpus and the topics and words within it. The topic-document distribution can be described as a matrix with one column per topic and one row per document. A cell of the matrix contains probability values. They indicate the prevalence of a topic in a document and add up to 100% for each document. Likewise, the topic-word distribution is a matrix with one column per topic and one row per word. A cell of the matrix contains probability values which indicate the relative occurrence of a word in a topic. These cells add up to 100%, as well. Both matrices together represent a summary of the contents of the document corpus (Schmiedel et al. 2018).

Based on the given document corpus, LDA tries to learn the representation of the  $K$  topics in each document and the distribution of the words for each topic by applying the generative process backwards. Unsupervised machine learning is used to estimate the latent variables, since the distribution is generally not computable (Dickey 1983; Blei et al. 2003). In most cases, this approximation of the a

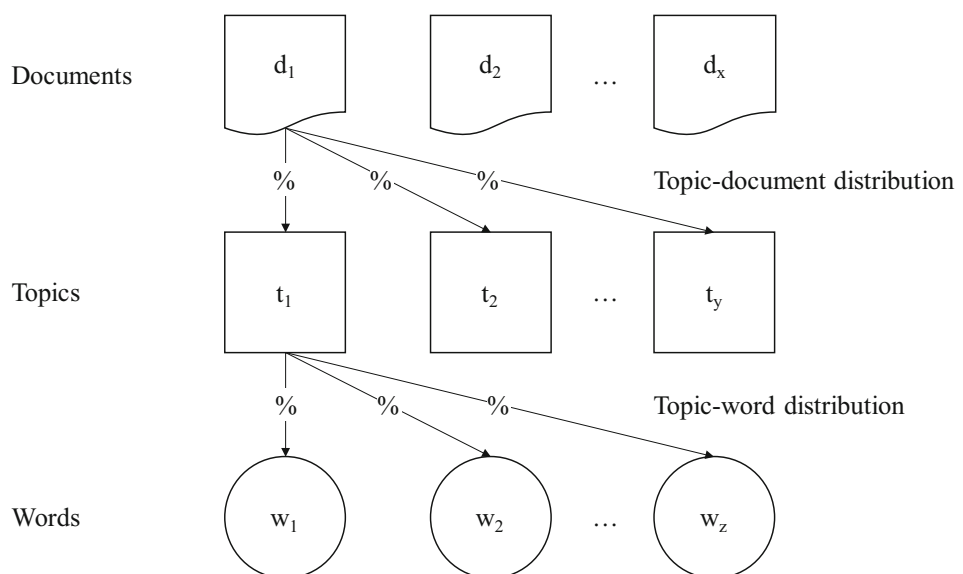
posteriori distribution of the topic structure is done by using the Gibbs sampling algorithm (Blei 2012).

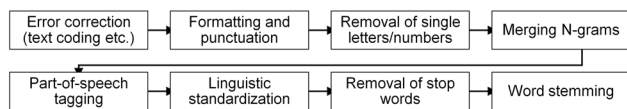
### 3.3.1 Data Preprocessing

In order to be able to train a meaningful topic model going forward, the module manual body remaining after the filtering steps had to go through further preparation steps. First of all, we had to remove irrelevant components from the modules for topic modeling, as these would otherwise impair the result: A document (i.e., a module) contains not only information about the module itself (teaching content, module description), but also information about required prior knowledge or skills. If these were to be included in the topic modeling, the result would be biased, as the algorithm cannot distinguish between knowledge imparted and prior knowledge required. Thus, we only used the module title, module content, and module learning objectives from each module as components of one document for the topic modeling process. After creating the reduced module documents, we further processed them by implementing the steps shown in Fig. 6.

Preprocessing included removing hyphenation and punctuation marks, as well as the transformation of umlauts or other language-specific characters and capital letters. Furthermore, we performed the steps of word stemming (e.g., “working” and “worker” are each traced back to the word stem “work”), the formation of N-grams (related words like “white house”), the removal of stop words (deletion of frequently used but meaningless words like “and”), and part-of-speech tagging (identification of the different word classes). We kept nouns, proper nouns, adjectives, adverbs, verbs and foreign words in the corpus

**Fig. 5** Relationship of documents, topics, and words following Debortoli et al. (2016)





**Fig. 6** Data preprocessing steps

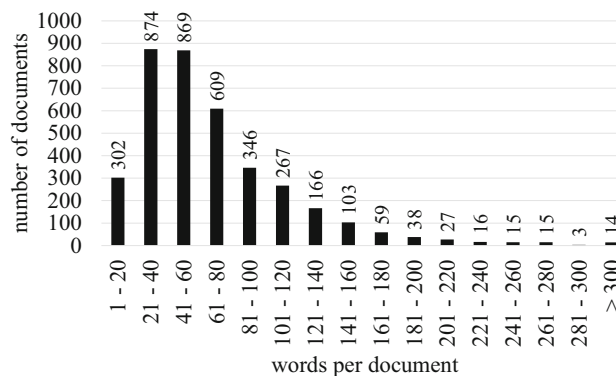
and removed all other part-of-speech tags (e.g., articles or prepositions).

After the preprocessing steps, the document corpus included a total of about 247,000 words, comprising almost 29,000 unique words. 3723 documents remain for topic modeling, which consist of 68 words on average. We show the document-word distribution in Fig. 7, which indicates that most of the documents contain 20–80 words. Only a few documents contain more than 200 words; the longest document has 508 words. This distribution of document lengths should be advantageous for topic modeling as the method usually works best with a large number of rather short documents (Tang et al. 2014).

### 3.3.2 Topic Number Determination

In topic modeling using the LDA approach, three parameters (in addition to the document corpus itself) have a significant influence on the quality of the result: the hyperparameters  $\alpha$  and  $\beta$  as well as the number of topics  $K$ . The LDA approach used for our analysis enables the automatic estimation of the hyperparameters  $\alpha$  and  $\beta$ , which are adjusted after every 20 runs of the algorithm. For the determination of  $K$ , different quantitative – for example, measuring semantic coherence and exclusivity (Mimno et al. 2011) or the density-based approach of Cao et al. (2009) – and qualitative approaches exist. In the present study, we employ a qualitative approach to determine  $K$ . We trained different topic models and let multiple coders compare them. To determine the range of  $K$  in which to train topic models, we followed research like that of Blei et al. (2003), Debortoli et al. (2016), and Schmiedel et al. (2018) to choose  $K$  in such a way that the result remains interpretable by humans, but is also large enough to clearly distinguish the topics from each other. Prior empirical research has shown that human-interpretable topic models typically include 10–50 topics (Schmiedel et al. 2018).

As we were not aware of any earlier studies that applied text mining or topic modeling to curricula, no values for  $K$  were available in the literature to align with. We therefore used the IS model curriculum of Kurbel et al. (2009) as a reference point, which divides the contents of an IS program into 47 subjects from nine aggregated areas (e.g., management and development of IS). At the beginning of our research project, the model curriculum had already



**Fig. 7** Data preprocessing steps

been endorsed by two scientific organizations (*VHB* and *GI*) for more than 10 years. When compared with the more recent but less established curriculum of Jung and Lehrer (2017), we therefore decided in favor of Kurbel et al. (2009) as an older but more suitable framework for assessing the contents of already existing IS programs. In the latter model curriculum, the majority of educational contents are distributed more or less equally among the fields of information systems, computer science, and business and economics. Further subjects include fundamentals from related fields, such as mathematics or law, as well as project seminars, programming courses, theses, and industry internships (Kurbel et al. 2009).

Our assumption was that the 47 subjects proposed by Kurbel et al. (2009) would be found in the module descriptions and therefore should form a good basis for  $K$ . Starting with  $K = 47$ , we first trained various topic models in 10-topic increments with lower or higher topic numbers and qualitatively determined their interpretability. Smaller topic models were discarded as they did not clearly differentiate between themes. In contrast, the larger topic models included duplicate topics that differed only minimally. The coders saw no added value in those duplicates, so we discarded the larger topic models as well. Training and labeling the larger topic models was nevertheless important as our corpus covers some documents that are duplicates, which results from universities offering the same modules in different study programs. Following Schofield et al. (2017), these few duplicates should not have a substantial impact on the topic modeling. If they did in our case, we would have noticed it during the labeling of the topic models with  $K < 47$  by receiving more meaningful topics, as LDA would designate exclusive topics to the duplicate documents at the loss of other topics in the corpus. After that, we trained models with  $K$  close to 47. They provided rather similar (i.e., good) results compared to  $K = 47$  but did not add more input for the following evaluation of the data. We therefore discarded them. In the

next subsection, we describe the process of topic modeling with LDA for the final model with  $K = 47$  topics.

### 3.3.3 Topic Modeling with LDA

During the topic modeling process, LDA randomly divides all words from the corpus into a given number of topics. Thereupon, each word is checked to determine whether it corresponds with the rest of the words in the topic. The procedure assumes that all other (not yet considered) words are correctly distributed among the topics. If similar words are contained in the topic, the word remains where it is. If it does not match the topic in question, it is assigned to the topic to which it has the greatest similarity. This process is repeated for each word until all words have been assigned to the topic that matches best. To implement the algorithm, we used the machine learning tool MALLET (MACHINE Learning for Language Toolkit), a Java-based package that allows for the statistical processing of natural language using machine learning applications for text. The software includes functions for document classification, clustering, and topic modeling (McCallum 2002). We connected to the tool with an R package, so that the actual coding took place in R. The topics that resulted from the algorithm's implementation were evaluated independently by two experts, each of whom reviewed the most likely words in the topics. The topic interpretation and labeling process was, in most cases, straightforward. Minor disagreements – mostly during the labeling of the application knowledge and soft skills area – could be resolved through the consideration of the corresponding top modules of the topics and short discussions amongst the experts. Table 5 illustrates the resulting top-word lists with two topics, one of which (topic 41) is about databases while the other (topic 27) is about project work. We provide all topics with their top 10 words and their top 10 modules (i.e., the modules that have the highest share of that topic) in Online Appendix II (available online via <http://link.springer.com>).

We defined the titles for the individual topics in accordance with the subjects set forth by Kurbel et al. (2009) by matching them on the basis of the words most likely to occur for each topic. Figure 8 illustrates the procedure for assigning topic titles. On the left side, we see the contents with which Kurbel et al. (2009) describe one of the topics, while on the right we see the most likely words for one of the final topics. Due to the fact that the contents are identical, we assigned the subject “programming and testing” to topic 44. The same procedure was applied for all other topics and subjects.

As stated before, the interpretation and labeling procedure was conducted for all trained topic models. The final step was a discussion of which topic model to use for

further processing. Both experts strongly agreed on the model with 47 topics.

### 3.3.4 Topic Framework Design

After the allocation of subjects to topics, the topics were arranged according to superordinate skill areas of the IS discipline in order to enable analyses at a higher aggregation level. In this step, we once again followed the framework proposed by Kurbel et al. (2009), who, for example, assign the subject “database” to the area “computer science”. We aggregated the topics to five areas and then assigned each topic to one of the superordinate areas, as shown in Table 6. The table is structured such that the first four areas are of a more theoretical nature, whereas the area of “application knowledge and soft skills” comprises the skills that are helpful in putting the knowledge of the other areas into practice.

In order to be able to compare the study programs, we used the document-to-topic matrix of the topic model to summarize the individual modules of each study program in such a way that analyses of each study program are possible at the (1) topic, (2) skill area, and (3) study program level (e.g., bachelor's/master's, university/university of applied sciences). We therefore added up the results for one document (i.e., module) with the corresponding counterparts and divided them by the total number of modules to obtain the average. In order to increase the meaningfulness of the following analysis, the modules were weighted beforehand according to their ECTS points. The ECTS points reflect the effort (as measured in time) that students have to invest in the module: completion of normal courses is often rewarded with five ECTS points, whereas final theses are worth ten to thirty credits. A few modules (usually lectures) did not contain any information on the number of ECTS credits; these were weighted with the modal value for lectures (i.e., five ECTS credits).

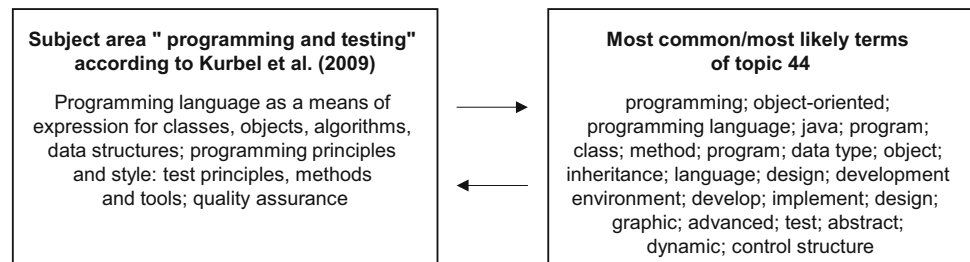
## 4 Results

With the topic model, which was assigned to the skill areas, we were able to perform a descriptive analysis to compare the IS study programs at both the skill area level and the topic level, as well as with regard to their orientation towards either business and economics or computer science. In Fig. 9, we show the results of the topic modeling process as averaged across all documents of all types of universities and degrees (i.e., the average across the examined study offerings of IS in Germany).

Within the study programs, we see the largest share for research methods, followed by topics such as professional practice, project work, and transfer competence. All of

**Table 5** Top words of topics 41 and 27

Topic	Top words (translated)
Topic 41	Database; relational; sql; database system; data model; database design; application; data modeling; model; object-relational; transaction; query; modeling; language; normalize; jdbc; create; design; database application; data
Topic 27	Project; team; presentation; in practice; project work; group; work; implement; result; solution; complex; method; develop; self-reliant; teamwork; documentation; project management; request; group work; in cooperation

**Fig. 8** Assignment of a topic to a subject area according to Kurbel et al. (2009)**Table 6** Topic aggregation to skill areas

Skill area	Topics (study content)
Information systems (IS) 11 topics	Fundamentals of information systems; project management; IT management; IT service management; e-business; process management; ERP systems; operations research; logistical optimization models; business intelligence; data analytics
Business and economics (BE) 11 topics	Fundamentals of business administration; fundamentals of economics; production and logistics; marketing and sales; human resources and organization; corporate management and strategy; entrepreneurship and innovation; cost accounting; financial accounting; investment and finance; management accounting
Computer science (CS) 12 topics	Algorithms and data structures; software development; programming and testing; web programming; theoretical computer science; operating systems and computer architecture; databases; computer networks and network technologies; information and communication technology; distributed systems; mobile applications design; IT security
Adjacent fields of knowledge (AK) – 4 topics	Fundamentals of mathematics; statistics; law; physics, engineering and medicine
Application knowledge and soft skills (AS) – 9 topics	Research methods; technology and implementation competence; self-reliance and transfer competence; presentation techniques; case studies; project work; analysis and evaluation competence; language and culture; professional practice

these are topics that, on the one hand, find themselves in specific modules like research methods in the final thesis, but are also incorporated, to some degree, into numerous other modules from every skill area. This explains their large share in comparison to other topics. The significant share for professional practice, project work, and transfer competence reflects the image of IS as a rather practice-oriented discipline. We see this view further strengthened by the relatively sizeable shares for software development and programming in the computer science (CS) skill area and for corporate management and strategy in the business and economics (BE) skill area. In contrast, only a small

proportion of the course of study is devoted to rather technical topics like logistical optimization and information and communication technology, which is also the case for physics and engineering. The other topics are incorporated into the study programs in somewhat equal proportions, an indication of the generalist approach of IS. The results become more differentiated if we do not consider all study programs at the same time, but rather each skill area, split up according to degree types. Therefore, in the following subsections, we present the results of the analysis on a more detailed scale in order to identify focal points and differences between the study programs.

### 4.1 Skill Area Level

At the level of skill areas for the study programs considered, the picture depicted in Fig. 10 emerges. The box plots show the differences between university types as well as between degree types. The X-axis shows the degree types, while the Y-axis represents the relative study shares for the degree type, university type, and skill area considered. The box shows the range between the upper and lower quartiles, while the line within the box represents the median. The lines extending vertically from the box indicate 1.5 times the interquartile distance above or below the upper or lower quartile. Outlier values are displayed as dots. This type of display remains identical for the box plots that follow later in this section. In the following, we describe the results and compare them to each other. The following figures thereby give insight into the general distribution over the university and degree types. For example, one could determine from

the figures what a “typical” IS study program looks like. The tables following the figures give insight into the “outliers”, that is, the study programs that distinguish themselves from the others with regard to a skill area or one or more topics. This manner of presentation may help future students or employers to identify study programs and universities with emphasis on certain topics that will help them to build their career or fill vacancies. The interpretation of the results (for example, what an ideal share of a skill area would look like) depends on the stakeholder requirements in the academic, industrial, or political sectors and varies on the basis of which stakeholder is considering the data.

The information systems (IS) skill area is rather stable across the groups. Only differences between the master’s programs are apparent. The proportion in the bachelor’s programs is very similar, especially at the universities. Overall, we see a slightly higher focus on the field

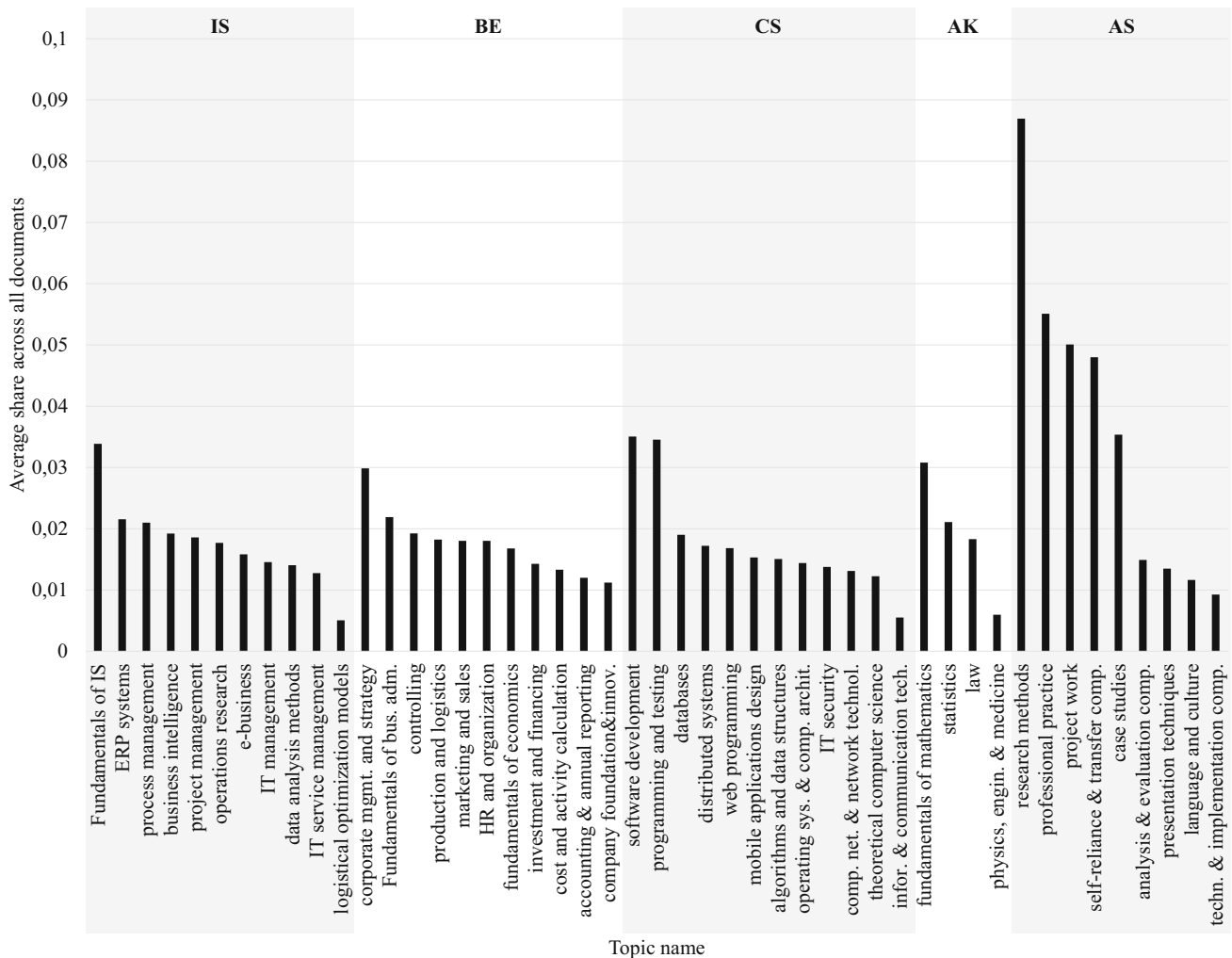


Fig. 9 Topic distribution across all documents



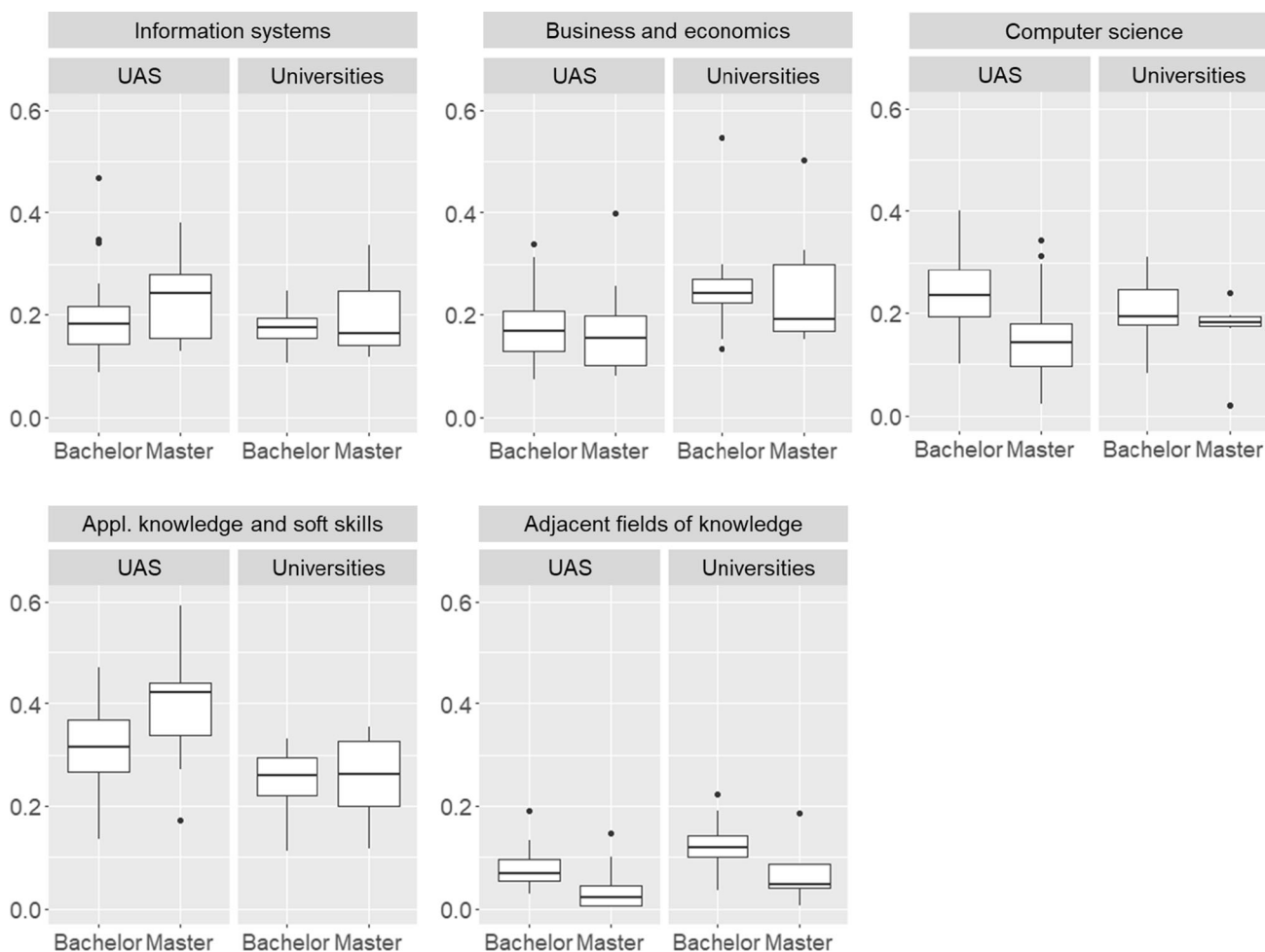
of IS in the master's programs. For business and economics (BE), the picture differs considerably between universities and universities of applied sciences, with the master's degree at universities, in particular, accounting for a higher share. The share of computer science (CS), on the other hand, varies more among universities of applied sciences (UAS). Whereas at universities, the CS share seems more stable for both degrees, especially in the master's programs. The adjacent fields of knowledge (AK) represent low proportions. They comprise, however, a slightly larger proportion at universities and in the bachelor's degree programs in which these foundations are mostly laid. For the last skill area, application knowledge and soft skills (AS), large differences between universities and universities of applied sciences are evident. More than half the master's study programs at universities of applied sciences assign shares above 40% to this area – an indication of the practical orientation of these study programs.

The previously mentioned outliers, which are displayed as points in the box plot, are listed in Table 7. They differ greatly from the distribution of their comparison group.

Examples are the Albstadt-Sigmaringen UAS, with a very high share in the IS area in the bachelor's program, and the Lower Rhine UAS, with a very high share in the EC area in the master's program. The Stralsund UAS has a very high share in both the bachelor's and master's programs in the AS area, while the Bundeswehr University in Munich shows a very high share in the master's program in the adjacent fields of knowledge. The Technical University of Darmstadt, on the other hand, has only a very marginal share of computer science in the master's program.

#### 4.2 Topic Level

Leaving the aggregation level of skill areas and zooming in on the individual skill areas (i.e., at topic level), we obtain a differentiated picture of the study programs. In the following, we take a closer look at each skill area, broken down by the types of degree. We again visualize the differences between the bachelor's and master's degree types with box plots. The topics of the skill areas under consideration are located on the Y-axis, while the X-axis



**Fig. 10** Shares of skill areas by university and degree type

indicates the relative study proportions. The further design of the box plots corresponds to Fig. 10.

#### 4.2.1 Information Systems

Figure 11 shows the results for the IS skill area. For the bachelor’s degree type, we see a focus on *IS fundamentals*. The topics of *process and project management*, *ERP systems*, and *operations research* have a larger share, as well. *Logistical optimization models*, on the other hand, is hardly an integral part of the bachelor’s programs. The same applies to the master’s programs. In contrast to the bachelor’s degree, we see an increased treatment of the topics concerning *business intelligence* and *data analytics* in the master’s programs. As before, at topic level the results contain outliers displayed as points in the box plot, which differ greatly from the distribution of their comparison group. Selected examples are given in Table 8, such as the universities of Duisburg-Essen and Siegen, which show a higher share of the *IS fundamentals* topic in their bachelor’s programs. The same applies to the University of Bamberg, but in this case to the master’s program. The Albstadt-Sigmaringen UAS stands out due to its high values for *IT management* and *IT service management*. The Bundeswehr University in Munich is the only university to offer a significant share of the topic *logistical optimization models* in its master’s program, presumably due to its sponsor, the German Ministry of Defense. *Data analytics* is becoming increasingly prominent in the bachelor’s program at the Hanover UAS, while *ERP systems* are a subject of particular focus in the master’s programs at the Ludwigshafen and Brandenburg UAS.

**Table 7** Overview of study programs focusing on specific skill areas

Degree	University type	University	Focus (+/–)	Share
Bachelor	UAS	Stralsund	(+) Application knowledge and soft skills	0.644
		Albstadt-Sigmaringen	(+) Information systems	0.466
		Lower Rhine	(+) Information systems	0.346
		Hanover	(+) Information systems	0.341
		Bielefeld	(+) Business and economics	0.338
		Wismar	(+) Adjacent fields of knowledge	0.191
	University	Hagen	(+) Business and economics	0.545
		Clausthal	(+) Adjacent fields of knowledge	0.223
		Clausthal	(–) Business and economics	0.133
Master	UAS	Stralsund	(+) Application knowledge and soft skills	0.683
		Lower Rhine	(+) Business and economics	0.399
		Nuremberg	(+) Computer science	0.342
		Brandenburg	(+) Computer science	0.312
		Aalen	(+) Adjacent fields of knowledge	0.147
	University	Siegen	(+) Computer science	0.238
		Munich (Bundeswehr)	(+) Adjacent fields of knowledge	0.187
		Darmstadt	(–) Computer science	0.021

#### 4.2.2 Business and Economics

Figure 12 depicts the results for the topics of the *business and economics* skill area. A certain underweight of the topic *entrepreneurship and innovation* in the bachelor’s programs is noticeable; apart from that, the topics are weighted in a rather balanced manner. The plot suggests that in the majority of bachelor’s study programs, a common pool of fundamentals from the areas of *business administration* and *economics* is taught. In the master’s programs, however, the distribution shifts to a stronger focus on *corporate management and strategy*. We also see a larger share of the topic of *entrepreneurship and innovation*, which was only marginally present in the bachelor’s programs, as well as an increased emphasis on the topics of *management accounting* and *human resources and organization*. *Economics* and *accounting*, on the other hand, are less represented in the master’s programs. The outliers determined for the economic skills area are listed in selected form in Table 9. Examples of outliers are the bachelor’s and master’s study programs of the distance-learning University of Hagen, along with the master’s study program of the Technical University of Darmstadt, for their emphasis on the topic of *fundamentals of economics*. The latter also has a very strong focus on the topic of *investment and finance*.

#### 4.2.3 Computer science

We depict the results for the topics of the *computer science* skill area in Fig. 13. The bachelor’s programs focus on the transfer of programming knowledge; the other topics are

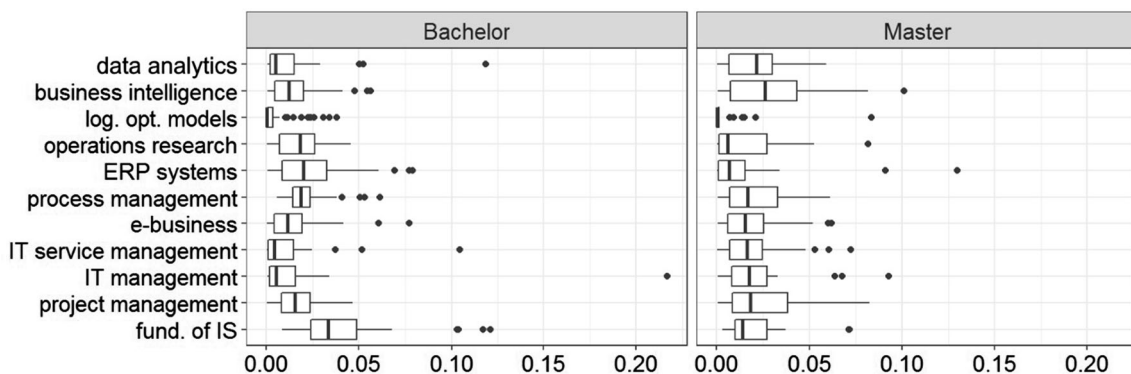


Fig. 11 Topic distribution in the IS skill area by type of degree

Table 8 Excerpt of study programs focusing on IS topics

Study content	University (university type)	Degree	Share
Fundamentals of information systems	Siegen (Uni)	Bachelor	0.121
	Duisburg-Essen (Uni)	Bachelor	0.104
	Bamberg (Uni)	Master	0.072
IT management	Albstadt-Sigmaringen (UAS)	Bachelor	0.217
IT service management	Albstadt-Sigmaringen (UAS)	Bachelor	0.104
Data analytics	Hanover (UAS)	Bachelor	0.118
ERP systems	Ludwigshafen (UAS)	Master	0.130
	Brandenburg (UAS)	Master	0.091
Logistical optimization models	Munich (Uni) (Bundeswehr)	Master	0.083

distributed evenly. Overall, a broad spectrum of topics seems to be covered. Only the topics of *theoretical computer science* and *information and communication technology* are rarely represented. In the master’s programs, we see the focus shifting away from *programming and testing* towards more general *software development*, *mobile applications*, and *distributed systems*. The outliers at the topic level are listed in selected form in Table 10. One of these is, for example, the bachelor’s study program at the Ruhr West UAS, which shows higher-than-average values for the topic of *information and communication technology*. The University of Braunschweig’s bachelor’s program

offers a focus on the topic of algorithms and data structures, while the Münster UAS’s bachelor’s program emphasizes the topic of *web programming*. For its master’s program, the Ravensburg-Weingarten UAS foregrounds the topic of *software development*.

4.2.4 Adjacent Fields of Knowledge

Figure 14 presents the results for the AK skill area, which has the smallest curricular share. The largest share in the bachelor’s programs is represented by the topic *foundations of mathematics*, followed by *statistics* and *law*. Other

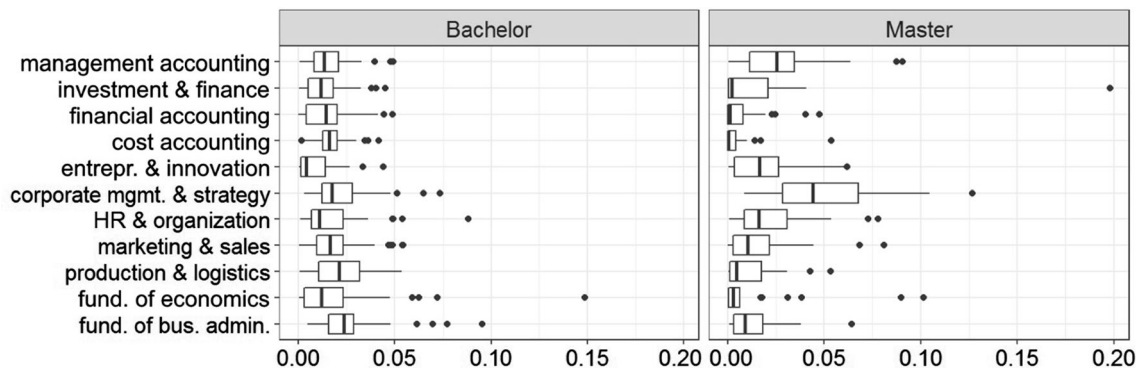


Fig. 12 Topic distribution in the business and economics skill area by type of degree

topics, e.g., from the field of engineering sciences, are only included in the curricula of certain individual universities, but not by the broad spectrum of institutions. In the master's programs, the adjacent fields of knowledge continue to lose importance: *mathematics* hardly plays a role anymore; the few contents from this skill area are concerned with the topics of *statistics* and *law*. The other topics – *physics*, *engineering sciences*, and *medicine* – are still offered by only a few universities. We see outliers, depicted in Table 11, for example in the form of the master's program of the Bundeswehr University or the bachelor's programs at the universities of Stuttgart, Clausthal, and Hildesheim. They have higher shares for the topic of *physics*, *engineering sciences*, and *medicine*. The Aalen UAS, on the other hand, places a comparatively large emphasis on the topic of *law* in its master's program.

#### 4.2.5 Application Knowledge and Soft Skills

In Fig. 15 we show the results for the AS skill area. Already strongly pronounced in the bachelor's programs, the *research methods* topic sees a further increase in the master's programs. This reflects the stronger scientific orientation of these programs. In the bachelor's programs, we see a larger share of the topic of *professional practice*. In contrast, the topics on *presentation techniques*, *language and culture*, *analysis and evaluation competences*, as well as *technology and implementation competences*, are only imparted to a limited extent. However, we observe a greater focus on topics related to *case studies*, *project work*, and *self-reliance and transfer competence*. Examples of outliers in Table 12 are the master's programs at the Berlin UAS and the Leipzig UAS, both of which place great emphasis on *professional practice*. Stralsund UAS, on the other hand, focuses on *case studies* in both its bachelor's and master's programs.

#### 4.3 Curricular Orientation Towards Business and Economics or Computer Science

In order to investigate whether the study programs are more oriented towards business and economics or computer science, we determine the difference sum from the shares of the skill areas of (1) business and economics and

(2) computer science for each study program. Resulting positive values indicate a stronger business orientation, while negative values signal an orientation towards computer science. Figure 16 shows the results as a scatter diagram. The X-axis depicts the difference sums, while the Y-axis shows the corresponding share of the study programs of the IS skill area. The scatter diagram is subdivided by type of degree; UAS are marked as points, universities as triangles. It can be seen from the figure that for bachelor's degrees, a large proportion of the study programs have an emphasis on computer science, especially as far as the universities of applied sciences are concerned, and there is a larger proportion of IS represented. For the master's programs, it is evident that both the dispersion of values and the orientation towards business are increasing. We show outliers with a difference sum of 0.2 or more in Table 13. For example, the bachelor's and master's programs of the distance-learning University of Hagen have a very pronounced business orientation, as does the master's program of the Technical University of Darmstadt. On the other hand, the University of Nuremberg, for example, places a focus on computer science in both the bachelor's and master's programs.

## 5 Conclusion

Due to the large number of different IS study programs, it is hardly possible for academic management, companies, or prospective students (not to mention politicians and university sponsors) to obtain and maintain an overview of the various study programs and their educational contents. However, this overview can be crucial for the aforementioned interest groups for various reasons. For starters, it can be used for the further development and improvement of curricula, the targeted search for suitable graduates, the selection of the most appropriate university location based on one's interests, or the management of grants. Due to these practical needs and the lack of corresponding methodological approaches, we presented and implemented a text mining procedure, founded on topic modeling techniques, for the evaluation and comparison of study programs at IHEs. As we have shown by the example of IS-related programs, text mining methods offer a

**Table 9** Excerpt of study programs focusing on business and economics topics

Study content	University (university type)	Degree	Share
Fundamentals of economics	Hagen (Uni)	Bachelor	0.149
	Hagen (Uni)	Master	0.101
	Darmstadt (Uni)	Master	0.090
Investment and finance	Darmstadt (Uni)	Master	0.198
Cost accounting	Bad Honnef-Bonn (UAS)	Master	0.054

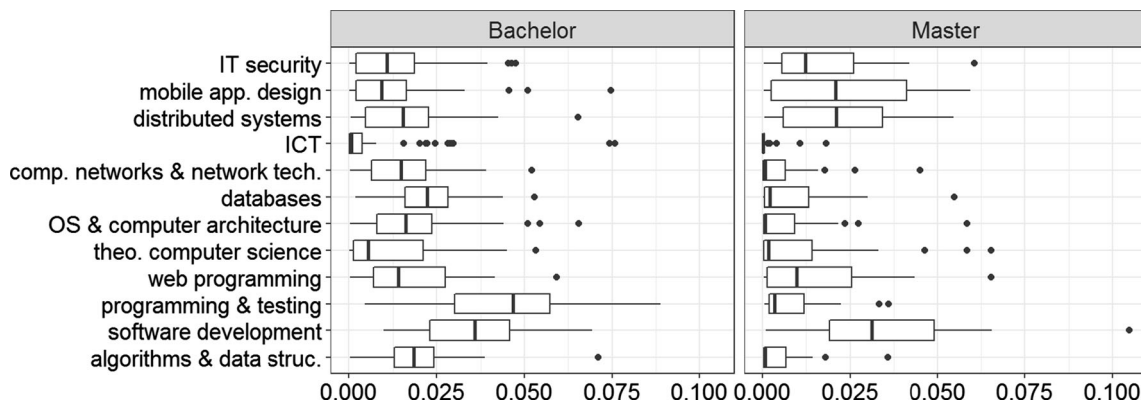


Fig. 13 Topic distribution in the computer science skill area by type of degree

Table 10 Examples of study programs focusing on computer science topics

Study content	University (university type)	Degree	Share
Theoretical computer science	Clausthal (Uni)	Bachelor	0.053
Information and communication technology	Ruhr West (UAS)	Bachelor	0.076
Algorithms and data structures	Braunschweig (Uni)	Bachelor	0.071
Web programming	Münster (UAS)	Bachelor	0.059
Distributed systems	Lower Rhine (UAS)	Bachelor	0.065
Mobile applications design	Lower Rhine (UAS)	Bachelor	0.075
Software development	Ravensburg-Weingarten (UAS)	Master	0.105

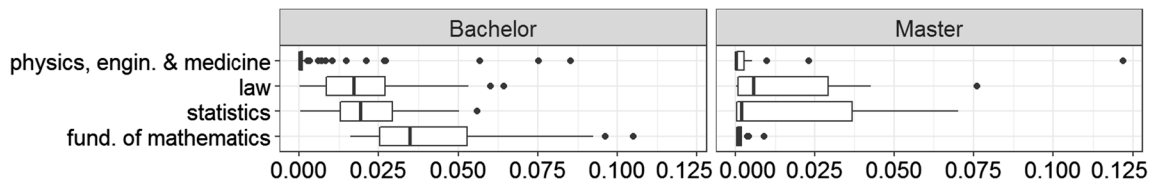


Fig. 14 Topic distribution in the AK skill area by type of degree

Table 11 Examples of study programs focusing on AK topics

Study content	University (university type)	Degree	Share
Physics, engineering sciences, and medicine	Munich (Uni) (Bundeswehr)	Master	0.122
	Stuttgart (Uni)	Bachelor	0.085
	Clausthal (Uni)	Bachelor	0.075
	Hildesheim (Uni)	Bachelor	0.057
Law	Aalen (UAS)	Master	0.076

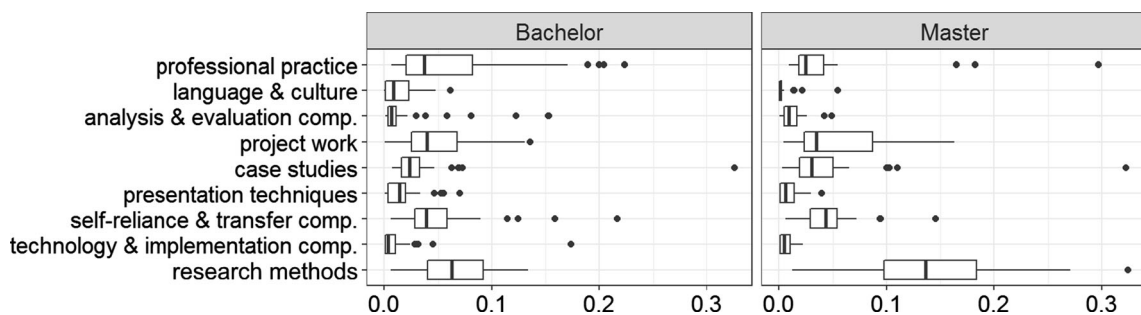


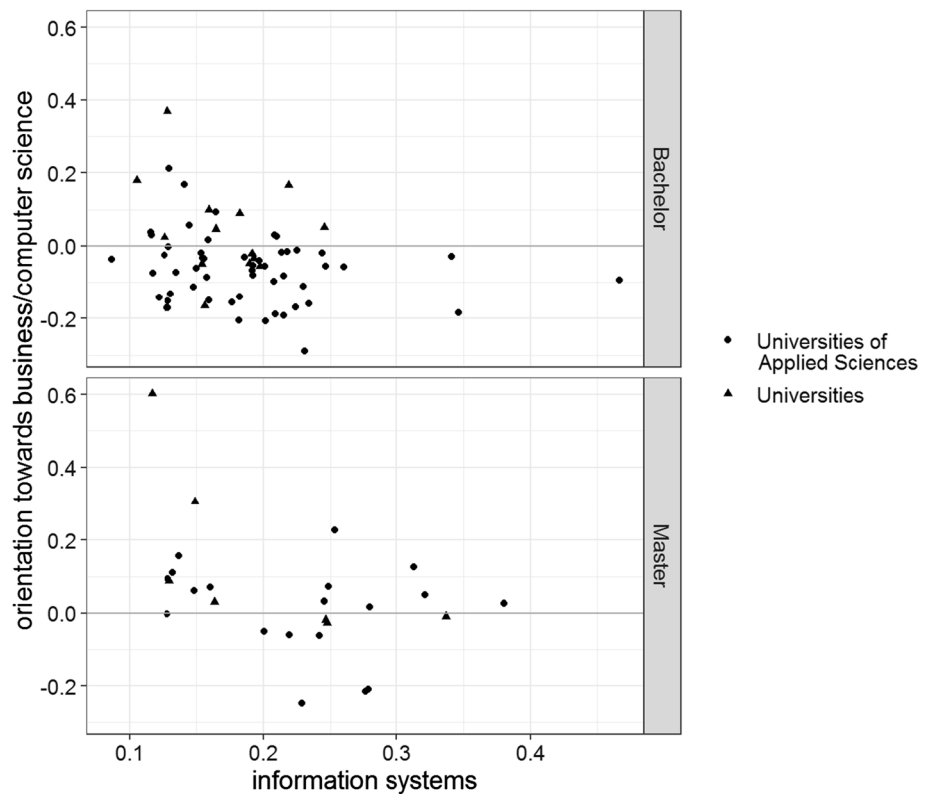
Fig. 15 Topic distribution in the AS skill area by type of degree



**Table 12** Examples of study programs focusing on AS topics

Study content	University (university type)	Degree	Share
Case studies	Stralsund (UAS)	Bachelor	0.326
	Stralsund (UAS)	Master	0.322
Research methods	Elmshorn (UAS)	Master	0.325
Professional practice	Berlin HTW (UAS)	Master	0.297
	Leipzig (UAS)	Master	0.182
	Ostfalia (UAS)	Master	0.165

**Fig. 16** Topic distribution in the IS skill area by type of degree



suitable toolset for conducting curricula analysis. The proposed procedure offers several advantages over traditional qualitative content analysis. These include, in particular, the partial automation of analysis steps, which cannot – and should not – completely replace the human data analyst, but nevertheless makes the processing of large amounts of unstructured raw data significantly more efficient. In the case of single isolated analyses, this increased efficiency may not play a major role; in the case of frequently recurring evaluations across a wide variety of study programs, however, this results in a substantial scaling advantage over purely manual procedures. Another advantage can also be seen in the easier repeatability or better reproducibility of the results compared to manual methods. Nevertheless, more research is needed to confirm our conjecture that the proposed procedure can be applied to arbitrary curricula from other academic disciplines. To

provide a starting point for this type of future research, we provide our code at [www.bitbucket.org/tmhelper/tmh](http://www.bitbucket.org/tmhelper/tmh). We also hope that our work will encourage academic stakeholders to think about standardizing the provision of module manuals to reduce the effort necessary to exploit them for research purposes. One central database where all accredited module manuals are available using clear formatting rules, or at least one central storage place for clearly structured PDFs, would reduce the necessary manual extraction effort and facilitate future studies.

As regards the field of IS in particular, the results, based on a data set of more than 90 study programs and 3700 modules, provided novel insights into the differences and similarities among the programs examined. Our analysis revealed the distribution of the individual skill areas and topics for bachelor’s and master’s programs, broken down into universities and UAS. For example, we identified

**Table 13** Examples of study programs focusing on IS topics

Focus	Bachelor's study programs (share)	Master's study programs (share)
Business and economics	Hagen (Uni) (0.368)	Darmstadt (Uni) (0.601)
	Pforzheim (UAS) (0.212)	Hagen (Uni) (0.306)
		Lower Rhine (UAS) (0.227)
Computer science	Nuremberg (UAS) (– 0.289)	Nuremberg (UAS) (– 0.248)
	Westphalian (UAS) (– 0.206)	Brandenburg (UAS) (– 0.215)
	Bremerhaven (UAS) (– 0.205)	Dortmund (UAS) (– 0.209)

outliers that may be of particular interest to prospective students who can incorporate the results into their decision concerning where and what to study. As expected, the biggest differences were found between the two types of degree. Where the bachelor's programs focus on the acquisition of programming knowledge and subject-specific fundamentals, for the master's programs this changes into a more holistic view of software development, corporate strategy, or data analytics. The UAS are, in line with their mandate, more practice-oriented, but not to the extent that one might have expected. Universities, as well, seem to have integrated the practical relevance of the various fields of study into their curricula, for example, in the form of project work or internships. Academic management could use the results and the underlying methodological approach when updating the IS curricula and incorporate the findings into the curricular design process. For practitioners, the results provide insights into the fulfillment of their expectations as employers by the various universities and degree programs. These insights enable them to identify graduates from universities who are most likely to satisfy their search criteria. Moreover, university sponsors can utilize the results in their grant processes.

Limitations of our approach result from the utilized text mining method and the underlying data. As with all topic modeling approaches, the limitation of our chosen text mining method (LDA) results from the parameters used for the topic modeling, such as the number of topics. For the underlying data we use module manuals of accredited study programs. Although the module descriptions are standardized due to the accreditation of their associated institutions, they differ in their extent. For example, the descriptions of some modules comprise only a few lines, while others occupy several pages. We acknowledge that certain biases regarding the module descriptions may exist in our research. But due to the large amount of data the approach is based on, we are confident that the effects have only minimal, if any, influence on the results. In comparison to other research methods, such as interviews, an advantage of processing such a broad range of module descriptions is the minimization of distortion caused by, for example, different contextual backgrounds (DeBortoli et al. 2014). One more issue regarding the module manuals is

that, as we could not identify any work comparing the contents of manuals with the content actually taught, we cannot exclude the possibility that module manuals do not reflect the material taught during the lectures. Nevertheless, our daily lecturing experience gives us confidence that there is a close correspondence in this respect. Regarding duplicate documents amongst the modules due to universities offering the same modules in different study programs, we are confident that they do not bias the results. Only few study programs offer courses in two study programs; thus, following Schofield et al. (2017), those few duplicates should not noticeably influence the topic modeling. If they had in our case, we would have seen it during the labeling of the topic models with  $K < 47$ . The problem with many duplicates would be that LDA would designate exclusive topics to these documents at the loss of other topics in the corpus. We did not receive more meaningful topics for topic models with  $K < 47$ , so that we are quite sure that our topics cover the content of the corpus.

Overall, we are convinced that our approach has delivered promising initial results that offer multiple opportunities for further research. These include primarily the text mining procedure itself, as well as its application to the broad spectrum of IS study programs in Germany. After examining the specific case of IS curricula in this paper, more design research is needed, aiming at a generalizable IT artifact that is tested for transferability to other fields of application. Second, further research should compare the curricula data with IS industry skill expectations, as outlined in Föll and Thiesse (2017). We are optimistic that this research avenue will yield interesting insights regarding possible gaps between employer expectations and demand in the job market on the one side, and the curricular offerings and universities' supply of graduates on the other side. The findings of this comparison might be assessed by experts from industry and academia in the form of surveys and interviews in order to verify the results and derive implications for future IS curricula. Third, it would be interesting to repeat the present study on a regular basis or to compare it to historical data and/or former studies in order to track the curricular changes over time.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Accreditation Council (2013) Regeln für die Akkreditierung von Studiengängen und für die Systemakkreditierung. [http://akkreditierungsrat.de/fileadmin/Seiteninhalte/AR/Beschluesse/AR\\_Regeln\\_Studiengaenge\\_aktuell.pdf](http://akkreditierungsrat.de/fileadmin/Seiteninhalte/AR/Beschluesse/AR_Regeln_Studiengaenge_aktuell.pdf). Accessed 16 Jan 2017
- Accreditation Council (2017) Entwicklung des Qualifikationsrahmens. <http://www.akkreditierungsrat.de/index.php?id=175>. Accessed 18 Jan 2017
- ACQUIN (2014) Leitfaden für Verfahren der Programmakkreditierung. <https://www.acquin.org/wp-content/uploads/2014/06/LeitfadenProgrammakkreditierung.pdf>. Accessed 25 Jan 2017
- Alghamdi R, Alfalqi K (2015) A survey of topic modeling in text mining. *Int J Adv Comput Sci Appl* 6(1):147–153
- AQAS (2014) Leitfaden zur Erstellung eines Akkreditierungsantrags. <http://www.aqas.de/downloads/AQAS-Leitfaden.pdf>. Accessed 25 Jan 2017
- Athey S, Plotnicki J (1998) The evaluation of job opportunities for IT professionals. *J Comput Inf Syst* 38(3):71–88. <https://doi.org/10.1080/08874417.1998.11647343>
- Benamati JH, Ozdemir ZD, Smith HJ (2010) Aligning undergraduate IS curricula with industry needs. *Commun ACM* 53(3):152–156. <https://doi.org/10.1145/1666420.1666458>
- Benbasat I, Dexter AS, Mantha RW (1980) Impact of organizational maturity on information system skill needs. *MIS Q* 4(1):21–34. <https://doi.org/10.2307/248865>
- Blei DM (2012) Probabilistic topic models. *Commun ACM* 55(4):77–84. <https://doi.org/10.1145/2133806.2133826>
- Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. *J Mach Learn Res* 3:993–1022
- Cao J, Xia T, Li J, Zhang Y, Tang S (2009) A density-based method for adaptive LDA model selection. *Neurocomput* 72(7–9):1775–1781. <https://doi.org/10.1016/j.neucom.2008.06.011>
- Chang CL, Chen V, Klein G, Jiang JJ (2011) Information system personnel career anchor changes leading to career changes. *Eur J Inf Syst* 20(1):103–117. <https://doi.org/10.1057/ejis.2010.54>
- Cheney PH, Hale DP, Kasper GM (1989) Information systems professionals: skills for the 1990s. In: Proceedings of the 22nd annual Hawaii international conference 4, pp 331–336. <https://doi.org/10.1109/HICSS.1989.48140>
- Chrysler E, van Auken S (2002) Entry level value versus career value of MIS courses: faculty expectations versus alumni perceptions. *J Comput Inf Syst* 42(3):38–43. <https://doi.org/10.1080/08874417.2002.11647501>
- Davis DC (2003) Job titles, tasks, and experiences of information systems and technologies graduates from a Midwestern university. *J Inf Syst Educ* 14(1):59–68
- Davis GB, Gorgone JT, Couger JD, Feinstein DL, Longenecker HE Jr (1997) Model curriculum and guidelines for undergraduate degree programs in information systems. *Assoc Inf Technol Prof*. <https://doi.org/10.1145/2593308>
- Debortoli S, Müller O, vom Brocke J (2014) Comparing business intelligence and big data skills. *Bus Inf Syst Eng* 6(5):289–300. <https://doi.org/10.1007/s12599-014-0344-2>
- Debortoli S, Müller O, Junglas I, vom Brocke J (2016) Text mining for information systems researchers: an annotated topic modeling tutorial. *Commun Assoc Inf Syst* 39(1):7. <https://doi.org/10.17705/ICAIS.03907>
- Dickey JM (1983) Multiple hypergeometric functions: probabilistic interpretations and statistical uses. *J Am Stat Assoc* 78(383):628–637. <https://doi.org/10.1080/01621459.1983.10478022>
- DiMaggio P, Nag M, Blei D (2013) Exploiting affinities between topic modeling and the sociological perspective on culture: application to newspaper coverage of U.S. government arts funding. *Poetics* 41(6):570–606. <https://doi.org/10.1016/j.poetic.2013.08.004>
- Downey JP, McMurtrey ME, Zeltmann SM (2008) Mapping the MIS curriculum based on critical skills of new graduates: an empirical examination of IT professionals. *J Inf Syst Educ* 19(3):351–363
- Dwyer C, Knapp CA (2004) How useful is IS 2002? A case study applying the model. *J Inf Syst Educ* 15(4):409–416
- Ehie IC (2002) Developing a management information systems (MIS) curriculum: perspectives from MIS practitioners. *J Educ Bus* 77(3):151–158. <https://doi.org/10.1080/08832320209599064>
- Fang X, Lee S, Lee TE, Huang W (2004) Critical factors affecting job offers for new MIS graduates. *J Inf Syst Educ* 15(2):189–203
- Föll P, Thiesse F (2017) Aligning IS curriculum with industry skill expectations: a text mining approach. In: Proceedings of the 25th European conference on information systems (ECIS), Guimarães
- Föll P, Hauser M, Thiesse F (2018) Identifying the skills expected of IS graduates by industry: a text mining approach. In: Proceedings of the international conference on information systems, San Francisco
- Foundation for the Promotion of the German Rectors' Conference (2016) Hochschulkompass Hochschulen. <https://www.hochschulkompass.de/hochschulen/download.html>. Accessed 27 Nov 2016
- Gallivan MJ, Truex DP, Kvasny L (2004) Changing patterns in IT skill sets 1988–2003: a content analysis of classified advertising. *SIGMIS Database* 35(3):64–87. <https://doi.org/10.1145/1017114.1017121>
- German Council of Science and Humanities (2000) Empfehlungen zur Akkreditierung privater Hochschulen. <http://www.wissenschaftsrat.de/download/archiv/4419-00.pdf>. Accessed 12 Nov 2016
- German Rectors' Conference (2016a) Studiengangsuche. <http://www.hochschulkompass.de/studium/suche/erweiterte-suche.html>. Accessed 23 Sept 2016
- German Rectors' Conference (2016b) Hochschulen. <https://www.hochschulkompass.de/hochschulen/download.html>. Accessed 30 Mar 2017
- German Rectors' Conference (2017) Das Hochschulsystem in Deutschland. [https://www.hrk.de/uploads/media/Das\\_Hochschulsystem\\_In\\_DEUTSCHLAND.pdf](https://www.hrk.de/uploads/media/Das_Hochschulsystem_In_DEUTSCHLAND.pdf). Accessed 30 Mar 2017
- Gorgone JT (2006) Information systems accreditation: preparation, process, and standards. *Commun Assoc Inf Syst* 17(19):391–403. <https://doi.org/10.17705/ICAIS.01719>

- Hickson J (2000) First steps to an IT career. *Comput Bull* 42(1):28–29. <https://doi.org/10.1093/combul/42.1.28>
- Igbaria M, Greenhaus JH, Parasuraman S (1991) Career orientations of MIS employees: an empirical analysis. *MIS Q* 15(2):151–169. <https://doi.org/10.2307/249376>
- Jung R, Lehrer C (2017) Guidelines for education in business and information systems engineering at tertiary institutions. *Bus Inf Syst Eng* 59:189–203. <https://doi.org/10.1007/s12599-017-0473-5>
- Kurbel K, Brenner W, Chamoni P, Frank U, Mertens P, Roithmayr F (2009) Studienführer Wirtschaftsinformatik 2009 | 2010. Studieninhalte – Anwendungsfelder – Berufsbilder. Universitäten in Deutschland | Österreich | Schweiz. Gabler, Wiesbaden. <https://doi.org/10.1007/978-3-8349-8057-1>
- Lee K, Mirchandani D (2010) Dynamics of the importance of IS/IT skills. *J Comput Inf Syst* 50(4):67–78. <https://doi.org/10.1080/08874417.2010.11645432>
- Lee DMS, Trauth EM, Farwell D (1995) Critical skills and knowledge requirements of IS professionals: a joint academic/industry investigation. *MIS Q* 19(3):313–340. <https://doi.org/10.2307/249598>
- Lee S, Song J, Kim Y (2010) An empirical comparison of four text mining methods. *J Comput Inf Syst* 51(1):1–10. <https://doi.org/10.1080/08874417.2010.11645444>
- Leitheiser RL (1992) MIS skills for the 1990s: a survey of MIS managers' perceptions. *J Manag Inf Syst* 9(1):69–91. <https://doi.org/10.1080/07421222.1992.11517948>
- Lifer JD, Parsons K, Miller RE (2009) A comparison of information systems programs at AACSB and ACBSP schools in relation to IS 2002 model curricula. *J Inf Syst Educ* 20(4):469–476
- Litecky CR, Arnett KP, Prabhakar B (2004) The paradox of soft skills versus technical skills in IS hiring. *J Comput Inf Syst* 45(1):69–76. <https://doi.org/10.1080/08874417.2004.11645818>
- Litecky C, Aken A, Ahmad A, Nelson HJ (2010) Mining for computing jobs. *IEEE Softw* 27(1):78–85. <https://doi.org/10.1109/MS.2009.150>
- Maier JL, Greer T, Clark WJ (2002) The management information systems (MIS) job market late 1970s-late 1990s. *J Comput Inf Syst* 42(4):44–49. <https://doi.org/10.1080/08874417.2002.11647052>
- McCallum AK (2002) Mallet: a machine learning for language toolkit. <http://mallet.cs.umass.edu>. Accessed 28 Jan 2021
- Mimno D, Wallach H, Talley E, Leenders M, McCallum A (2011) Optimizing semantic coherence in topic models. In: Proceedings of the 2011 conference on empirical methods in natural language processing, pp 262–272
- Miner G, Elder IVJ, Hill T, Delen D, Nisbet R (2012) Practical text mining and statistical analysis for non-structured text data applications. Academic Press, Cambridge. <https://doi.org/10.1016/C2010-0-66188-8>
- Müller O, Schmiedel T, Gorbacheva E, vom Brocke J (2016) Towards a typology of business process management professionals: identifying patterns of competences through latent semantic analysis. *Enterp Inf Syst* 10(1):50–80. <https://doi.org/10.1080/17517575.2014.923514>
- Nelson HJ, Ahmad A, Martin NL, Litecky CR (2007) A comparative study of IT/IS job skills and job definitions. In: Proceedings of the 2007 ACM SIGMIS CPR conference on computer personnel research: the global information technology workforce, pp 168–170. <https://doi.org/10.1145/1235000.1235038>
- Nettleton SC, Litchfield AJ, Taylor TL (2008) Engaging professional societies in developing work-ready graduates. In: Proceedings of the 31st HERDSA conference, pp 241–251
- Noll CL, Wilkins M (2002) Critical skills of IS professionals: a model for curriculum development. *J Inf Technol Educ* 1(3):143–154
- Schmiedel T, Müller O, vom Brocke J (2018) Topic modeling as a strategy of inquiry in organizational re-search: a tutorial with an application example on organizational culture. *Organ Res Methods* 22(4):941–968. <https://doi.org/10.1177/1094428118773858>
- Schofield A, Magnusson M, Thompson L, Mimno D (2017) Understanding text pre-processing for latent Dirichlet allocation. In: Proceedings of the 15th conference of the European chapter of the association for computational linguistics 2, pp 432–436
- Simon JC, Kaiser KM, Beath C, Goles T, Gallagher K (2007) Information technology workforce skills: does size matter? *Inf Syst Manag* 24(4):345–359. <https://doi.org/10.1080/10580530701586102>
- Standing Conference of the Ministers of Education and Cultural Affairs (2015) Das Bildungswesen in der Bundesrepublik Deutschland 2013/2014: Darstellung der Kompetenzen, Strukturen und bildungspolitischen Entwicklungen für den Informationsaustausch in Europa. [https://www.kmk.org/fileadmin/Dateien/pdf/Eurydice/Bildungswesen-dt-pdfs/dossier\\_de\\_ebook.pdf](https://www.kmk.org/fileadmin/Dateien/pdf/Eurydice/Bildungswesen-dt-pdfs/dossier_de_ebook.pdf). Accessed 3 May 2017
- Stefanidis A (2014) Undergraduate information systems in the UK: analysing curriculum skills and industry needs. Doctoral dissertation. <https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.594456>. Accessed 28 Jan 2021
- Stefanidis A, Fitzgerald G, Counsell S (2013) IS curriculum career tracks: a UK study. *Educ Train* 55(3):220–233. <https://doi.org/10.1108/00400911311309297>
- Steyvers M, Griffiths T (2007) Probabilistic topic models. *Handb Latent Semant Anal* 427(7):424–440
- Stifterverband (2013) Methodik: Der Hochschulvergleich. <http://www.stifterverband.com/gruendungsradar/2013/hochschulbefragung/methodik/hochschulvergleich/index.html>. Accessed 28 Oct 2016
- Tang J, Meng Z, Nguyen X, Mei Q, Zhang M (2014) Understanding the limiting factors of topic modeling via posterior contraction analysis. *Int Conf Mach Learn* 32(1):190–198
- Todd PA, McKeen JD, Gallupe RB (1995) The evolution of IS job skills: a content analysis of IS job advertisements from 1970 to 1990. *MIS Q* 19(1):1–27. <https://doi.org/10.2307/249709>
- Topi H, Valacich JS, Wright RT, Kaiser K, Nunamaker JF Jr, Sipior JC, de Vreede GJ (2010) IS 2010: curriculum guidelines for undergraduate degree programs in information systems. *Commun Assoc Inf Syst* 26(18):359–428
- Topi H, Karsten H, Brown SA, Carvalho JA, Donnellan B, Shen J, Tan BCY, Thouin M (2017) MSIS 2016 global competency model for graduate degree programs in information systems. *Commun Assoc Inf Syst*. <https://doi.org/10.17705/1CAIS.04018>
- Trauth EM, Farwell DW, Lee D (1993) The IS expectation gap: industry expectations versus academic preparation. *MIS Q* 17(3):293–307. <https://doi.org/10.2307/249773>
- White GW (2005) Business information courses in LIS programs: a content analysis. *J Bis Finance Libr* 10(2):3–15. [https://doi.org/10.1300/J109v10n02\\_02](https://doi.org/10.1300/J109v10n02_02)
- White KB, Lederer AL (1987) Panel 6 – Connecting: strategies for bridging the gap from education to industry and government. In: Proceedings of the international conference on information systems, pp 470 ff
- Williams CL, Pomykalski JJ (2006) Comparing current IS curricula to the IS 2002 model curriculum. *Inf Syst Educ J* 4(76):3–10
- Yang SC (2012) The master's program in information systems (IS): a survey of core curriculums of US institutions. *J Educ Bus* 87(4):206–213. <https://doi.org/10.1080/08832323.2011.591847>
- ZEVA (2016) Die Programmakkreditierung – Leitfaden zur Akkreditierung von Studiengängen. [http://www.zeva.org/fileadmin/Downloads/Leitfaden/Leitfaden\\_Programmakkreditierung.pdf](http://www.zeva.org/fileadmin/Downloads/Leitfaden/Leitfaden_Programmakkreditierung.pdf). Accessed 25 Jan 2017