BRIDGING THE GAP BETWEEN INDUSTRY SKILL DEMAND AND UNIVERSITY SKILL PROVISION

Patrick Föll
University of Wuerzburg, patrick.foell@uni-wuerzburg.de

Marco Röder
University of Würzburg, marco.roeder@uni-wuerzburg.de

Frédéric Thiesse
University of Wuerzburg, frederic.thiesse@uni-wuerzburg.de

Follow this and additional works at: https://aisel.aisnet.org/ecis2021_rp

Recommended Citation
Föll, Patrick; Röder, Marco; and Thiesse, Frédéric, "BRIDGING THE GAP BETWEEN INDUSTRY SKILL DEMAND AND UNIVERSITY SKILL PROVISION" (2021). ECIS 2021 Research Papers. 96.
https://aisel.aisnet.org/ecis2021_rp/96

This material is brought to you by the ECIS 2021 Proceedings at AIS Electronic Library (AISeL). It has been accepted for inclusion in ECIS 2021 Research Papers by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
BRIDGING THE GAP BETWEEN INDUSTRY SKILL DEMAND AND UNIVERSITY SKILL PROVISION

Abstract

Today’s working environments are subject to dynamic changes due to the proliferation of digital technologies and systems. This phenomenon poses a challenge for universities and other institutions of higher education, which are expected to adapt their course offerings to the rapidly changing demands in the labor market. The complexity of the task and the importance of speed pose an opportunity for automated data-driven methods with which the contents of study curricula can be compared, assessed, and, if necessary, adapted to the world of work. Owing to the lack of established solutions, this study presents a procedural methodology to create artifacts for analyzing, evaluating, and comparing curricula as well as job postings using topic modeling. In addition, we demonstrate the practical applicability of the methodology by the example of the IS discipline and present empirical results from the analysis of IS-related study programs in Germany.

Keywords: IS Curriculum Design, Job Skill Requirements, Topic Modeling.

1 Introduction

Digitalization offers immense opportunities as well as new challenges and uncertainties. Hardly a week goes by without the working world’s digitalization being highlighted in major (business) magazines and daily newspapers. This level of attention on the part of the mass media reflects the fact that digital transformation has brought about a profound change in companies and the labor market. While these changes have already been substantial over the past years, COVID-19 further increases the speed and reach of digital transformation in organizations. According to recent studies (Baig et al., 2020; Laberge et al., 2020), the pandemic often brought down the time needed to adopt IT innovations, which would have taken years in the past, to just a few months. A side-effect of this development, however, is that employee’s skills that are in demand today may be obsolete tomorrow, thus furthering the skills mismatch that already exists in the labor market. According to Hoteit et al. (2020), nearly 1.5 billion jobs may be affected of which nearly 30% require an entirely updated skill set. Students as future employees have to cope with this dynamic, which makes lifelong learning and up-to-date skills even more critical. Against this backdrop, universities and other institutions of higher education as providers of the required skills and knowledge face the challenge of adjusting their curricula to the rapidly evolving skill demand in the labor market. In the worst case, students may be taught skills that are no longer in demand. It is hence crucial that these have tools at their disposal that enable them to compare their curricular contents to the changing demands of the working world, draw conclusions from these analyses and, if necessary, take action. This holds, among others, for the case of Information Systems (IS) curricula, which are said to adapt to changes in the “real world” slowly (Granger et al., 2007; Hirschheim and Klein, 2003; McGann et al., 2007). Automated data-driven tools that support the process of curriculum design would be of interest for at least three interest groups: (i) academic management could assess their curriculum and further enhance it; (ii) companies can target those universities that most closely fit their required skills in their recruitment of graduates; and (iii) students can benefit from the overview to help them select a university, based on their desires and interests in terms of the field of study. However, such tools...
are currently not available or only to an insufficient extent with regard to the necessary implementation effort, repeatability, or transferability. Furthermore, owing to the lack of tools, there is a lack of empirical studies matching the labor market’s skill needs with the skills taught by universities.

In this paper, we present a novel procedural methodology based on text mining methods and aligned to the elements of a design theory (Gregor and Jones, 2007) to overcome the lack of appropriate tools (i.e., artifacts) for the analysis and comparison of study program curricula and the corresponding demands in the job market. After outlining our research agenda in an earlier paper (Föll and Thiesse, 2017), the following sections present our procedure and practical results of the artifact implementation by the example of the IS discipline. After a review of the literature and a description of our research methodology, we describe our procedure along the principles of a design theory put forth by Gregor and Jones (2007) for better comprehensibility and transferability. Subsequently, we demonstrate the suitability of our procedural methodology through the design of a corresponding artifact using the IS discipline as an example, before concluding our paper.

2 Related Work

The research objective of aligning (IS) curricula with industry skill expectations shows overlap with three different research fields: (i) university and (ii) labor-market research as well as (iii) IS education research. IS and related disciplines are subject to continuous change and constant adaptations because content, technologies, and skills can change rapidly (Lee et al., 1995; Gallivan et al., 2004). These changes can take the form of, for example, new hardware, platforms (e.g., smartphones) or new programming languages. The changes associated with digitalization—such as the increase in large data sets and their processing and evaluation, or innovative business models that manifest themselves in novel business processes in organizations—also confront the discipline with fresh challenges. This generates new needs and requirements for IS graduates, which should be reflected in the curricula to prepare graduates for this ever-changing environment. In order to guarantee successful curricular integration, employee-skill research is initially required to identify and evaluate changing needs and requirements. The task of curriculum research is to take these findings into account and incorporate them into the curricula’s adjustment or redesign. Thus, both topics interact with each other and are interdependent. The related work in these fields is divided along two axes.

On the one hand, there are studies pertaining to the guidelines for curriculum design of IS courses (Davis et al., 1997; Topi et al., 2010; Jung and Lehrer, 2017) and the evaluation of how they are used in reality (Yang, 2012; Dwyer and Knapp, 2004). These studies mainly applied traditional data-gathering methods, such as (expert) interviews or questionnaires, to discover the curricula needs and generate guidelines. The preferred methods for evaluating these guidelines and their uses are data-driven approaches via content analysis (Stefanidis et al., 2013; White, 2005).

On the other hand, there are studies on business skill expectations in IS-related job categories. They cover the skill requirements in a field of work characterized by innovation and technological change, which are often sources of competitive advantage (Gallivan et al., 2004; Nelson et al., 2007). A majority of these studies are empirical; they gather their information via traditional questionnaires (Noll and Wikins, 2002; Davis, 2003) and interviews (Nettleton et al., 2008; Simon et al., 2007) or through data collection in printed (Todd et al., 1995; Maier et al., 2002) or online job postings (Litecky et al., 2010; Debertoli et al., 2014).

Some studies also try to combine the IS curriculum with the job market requirements to determine the correlation between the skills supplied by universities and the skills demanded by industry (Litecky et al., 2004). Through this combination, the studies aim to facilitate more appropriate offerings at universities to better prepare the students for the job market’s expectations. To achieve this goal, various methods, such as interviews (Trauth et al., 1993), questionnaires (Lee et al., 1995) or data-driven approaches (Stefanidis, 2014), are used.

However, the examined literature largely lacks approaches that combine IS curriculum design with the skill expectations of the labor market and are (i) repeatable and reproducible, (ii) based on quantitative
data, (iii) up-to-date, and (iv) capable of enabling curriculum designers to react quickly to the constantly changing environment – this is the gap that we address in the following sections.

3 Methodology

For the structure and description of our procedural methodology, we align with the principles of a design theory put forth by Gregor and Jones (2007), shown in Figure 1. The design theory framework enables the communication independently of a specific implementation (Peffers et al., 2018). Therefore, we expect to obtain better comparability to other proposals as well as an easier implementation and transferability to other disciplines or fields of application.

According to Gregor and Jones (2007), one of the basic ideas behind developing the design theory approach was to enable a structured transfer of theoretical design knowledge to other researchers. For this purpose, they identified eight components that interact as shown in Figure 2. The component Purpose & Scope describes the meta-requirements of the depicted system or artifact and its scope. Constructs explain the important terms (entities of interest) in relation to the artifact. The Principles of Form and Function describe the architecture and functions of the artifact. Artifact Mutability depicts the extent to which changes to the artifact are taken into account in the design of the theory. Testable Propositions are hypotheses about the system, which enable an evaluation. Justificatory Knowledge is equivalent to knowledge from the natural, social, and design sciences—it forms the foundation of the theory. The Principles of Implementation describe the prerequisites that must be fulfilled to implement the theory in a specific context. The Expository Instantiation corresponds to the implementation and realization of the artifact.

4 Research Approach

In this section, we present our procedural methodology in more detail. This includes the Purpose and Scope (i.e., problems and meta requirements to be addressed) and the Principles of Implementation. Also, Testable Propositions are developed to check whether the implementation of the procedure (e.g., artifact) can fulfill the set goals and principles.

4.1 Purpose and Scope

In order to achieve the meta-requirements of the procedural methodology, we have to initially consider the problems raised by the formulation of our research objective of overcoming (i) the lack of tools and (ii) the lack of analyses for matching the labor market’s skill needs with the skills taught by universities.
Regarding our procedure, the issues raised can be condensed into four domains. Each domain is derived from the previous one. The primary and initial problem is the already mentioned (P1) lack of adequate literature, methods, and findings regarding the research question raised. Suppose we want to remedy this issue by using a new, quantitative procedure. In that case, the next problem area arises—that of the (P2) (mass) data to be examined. These data are typically unstructured, heterogeneous, and cannot be analyzed easily. In addition, the challenge emerges in analyzing and evaluating the data with reasonable effort (P3). Finally, we are confronted with the question of how the obtained findings can be evaluated (P4). The four problem fields yield requirements that are summarized under the concept of Meta Requirements conceived by Gregor und Jones (2007). For our procedural methodology, the meta requirements derived in Figure 2 result from the problem definitions above. They are explained in more detail in the following.

**Figure 2.** Identified problems (P) and corresponding meta requirements (MR)

Given the lack of prior research, a sufficiently large database (MR1.1) is required to address the research issue. To this end, all relevant curricula and the most important platforms of appropriate job postings are to be collected. Depending on the respective study goals, smaller samples may be sufficient. Also, the collected data set has to be up-to-date (MR1.2), i.e. the curricula and job postings that are relevant at the time of implementation should be used. As these data are unstructured and heterogeneous, requirements regarding the quality of the data (MR2.1) are specified—they must be transformed into a structured form. With regard to both the curricula and the job postings, not all elements of the data are relevant. We must extract the relevant information from the complete data (MR2.2). As soon as the relevant data are available in a structured form, they are to be analyzed with reasonable effort. Due to the considerable amount of data involved, this is only possible by automating the evaluation and analysis process (MR3.1). For this step, we recommend the supervision of experts (MR3.2) to ensure the evaluation’s coherence and thus the results based on it. Once the results are available, the issue of assessing and evaluating the findings arises. Here, testable propositions regarding the methodology (MR4.1) can be formulated and stakeholders from academia and industry (MR4.2) can be involved by using interviews and surveys to evaluate and discuss the results.

To address the challenges and requirements mentioned and thus provide the foundation for a viable solution, the interaction of different Constructs, in combination with recommendations based on Design Principles, is necessary (Gregor and Jones, 2007). In the following, we present the constructs and design principles that build the framing of our procedural methodology.

### 4.2 Constructs

The three constructs used to compare the industry skill expectations towards graduates with the skills imparted during a study program are shown in Figure 3. We briefly explain them in the following.

**Figure 3.** Constructs
Module manuals: These contain information about the competences and skills imparted in a course or lecture. They should be available in a standardized form so that they can be compared with each other, and comply with certain minimum requirements regarding their content (e.g. module descriptions, number of credits). In Germany, accredited study programs generally take this form. For European or international study programs, only those study programs should be considered which are structured according to comparable procedures.

Job postings: These can be found on (online) job platforms and contain information about the skills required for a job and its areas of responsibility. The focus should be on job postings that are suitable for job starters. Graduates tend to apply for such positions as they do generally not have any prior work experience. The inclusion of job postings for experienced professionals would thus hinder the determination of the skills expected from graduates.

Discipline-specific framework(s): In order to compare the above constructs (i.e., both data sources), it is useful to include a third construct, if available: discipline-specific frameworks. They can help to categorize and aggregate the recognized skills and job postings. Thus, individual competencies can be assigned to hard or soft skills or a job posting can be assigned to a specific job profile.

4.3 Design Principles

The framing of the procedural methodology results from the interaction of the mentioned constructs with specific design principles (DP). These principles must be adhered to ensure a successful artifact implementation, Figure 4 illustrates the interaction.

**Figure 4. Design principles**

The first design principle refers to the focus of the procedure. This should be placed on a single discipline since otherwise, no discipline-specific frameworks can be applied and the separation of skills and job postings will be more difficult. The focus on one discipline does not exclude the possibility of comparing the obtained results to other disciplines – **DP1:** The focus of the module manual and job posting selection should be on one discipline.

The module manuals should follow a standardized structure to enable comparability of the contents – **DP2:** Module manuals of standardized study programs should be considered.

The job postings should address job starters. This can be done through specialized job platforms for graduates or through job postings on general platforms that are identified as suitable for job starters – **DP3:** Only job postings should be considered that are addressed to graduates and do not require prior job experience.

Since both module manuals (despite standardized structure) and job postings (coming from various platforms and companies) are unstructured and heterogeneous, they must be transformed into a structured
form for further processing – **DP4**: The data heterogeneity of the module manuals and job postings must be eliminated. Methods have to be chosen that transform them into a structured form.

As we want to analyze the data with reasonable effort, the analysis and evaluation process should be automated to the greatest possible extent – **DP5**: The analysis and the evaluation process are to be automated.

Finally, to support the comparison of both constructs, it is useful to include skill and job profile frameworks or standards (if available for the considered discipline) in the comparison – **DP6**: Competence and job profile standards or frameworks should be integrated into the comparison of constructs.

### 4.4 Testable Propositions

The design principles result in testable propositions (TP) (Gregor and Jones, 2007), which can be used to verify the implementation of an artifact based on the proposed procedural methodology. Table 1 shows these propositions together with the design principle they are derived from.

<table>
<thead>
<tr>
<th>Origin</th>
<th>Testable propositions</th>
</tr>
</thead>
<tbody>
<tr>
<td>DP1</td>
<td>TP1: If only one discipline is considered, the results of the approach will be better than if different disciplines are considered at the same time.</td>
</tr>
<tr>
<td>DP2</td>
<td>TP2: If standardized module manuals are considered, the implementation of the approach is more likely to be successful than if non-standardized module manuals are used.</td>
</tr>
<tr>
<td>DP3</td>
<td>TP3: If job postings for job starters are considered, the quality of the results will be better compared to regular job postings.</td>
</tr>
<tr>
<td>DP4</td>
<td>TP4: If the underlying data are converted into a structured form, the quality of the results will be improved compared to the unstructured form.</td>
</tr>
<tr>
<td>DP5</td>
<td>TP5: If the analysis and evaluation process is automated, the total effort is reduced compared to non-automated approaches.</td>
</tr>
<tr>
<td>DP6</td>
<td>TP6: If existing skill and job profile standards and frameworks are included in the analysis, this facilitates classification in comparison to a procedure without frameworks.</td>
</tr>
</tbody>
</table>

Table 1. Testable propositions

TP1 results from DP1 and the presumption that a consideration of different disciplines would lead to evaluation problems. The separation of the individual competences is more difficult and the application of discipline-specific frameworks is not possible. TP2 results from the requirement, derived from DP2, that certain minimum standards must be set for the module manuals. Thus, for example, descriptions of the contents of the modules are mandatory for the implementation. In the case of non-standardized module manuals, these minimum standards are not automatically guaranteed. Where TP2 focuses on the module manuals, TP3 is about the job postings. It is derived from DP3. The focus here is on jobs for job starters, which can increase the significance and thus the quality of the results since these are the jobs that are the most relevant for the considered group of graduates. TP4 is derived from DP4 and thus makes propositions about the structured form of the data. The organization into such a structured form is necessary to increase the quality. Only then can the relevant parts of the respective documents be considered and later grouped for analysis. TP5 is the derivative of DP5 and refers to the automation of the analysis and evaluation process. Due to the large amount of data required for the approach, automation of this process is of great relevance in order to keep the effort to an acceptable level. TP6 finally results from DP6 and considers the integration of discipline-specific frameworks in the comparison of the two data sets. Such an integration can facilitate the classification of the results obtained from the analysis and the subsequent evaluation.

After outlining all relevant components of our procedural methodology, the following section outlines the design of an artifact along the described design principles for the IS discipline. Thereby we describe the implementation of the design principles which also provide a way of solving the problems (P1-P4) presented initially.
5 Implementation for the IS Discipline

The design components described before result in the procedural methodology visualized in Figure 5, the steps of which we explain in the following subsections by using the example of the IS discipline. The first three subsections describe the text mining part of the procedure. After providing insights into the results obtained from the comparison, the last subsection is devoted to evaluation and validation.

Figure 5. Conceptual procedure

5.1 Data Collection, Filtering and Preparation

The first step is the collection of module manuals and job postings, as these are the chosen source of competences imparted during university studies and the skills expected by the industry. To yield a suitable database for further processing, the procedure is structured as follows—using the example of IS and taking DP1 to DP4 into account: We start with collecting the module manuals of IS study programs (DP1) in Germany from the respective university pages. Since accredited (DP2) German-language study programs are the aim of the analysis, the module manuals first pass a number of filtering steps (e.g., removal of non-accredited, non-German-language study programs) until the final number of manuals is determined. From these, we extract the relevant parts for further processing (e.g., module description and title) and save them (DP4). After collecting the module manuals, we turn our focus towards the collection of job postings. Due to the large number of available job postings and platforms in IS, we have to conduct a preselection amongst the platforms. This is necessary because the job postings are collected by web scraping. Each platform added to the selection requires additional development effort. After completing the collection, the job postings and the module manuals previously collected pass various filtering steps. These are e.g. the exclusion of incomplete or other-language job postings as well as duplicates. In addition, only jobs suitable for job starters, who do not require professional experience (DP3), are considered. We then extract the relevant parts from the job postings (e.g., task description and job requirement profile) and save them (DP4) for further processing.

5.2 Data Preprocessing

The data preprocessing occurs as preparation for the analysis and evaluation components of the procedure, where text mining techniques are used. The methods of text mining are selected due to the automation requirements of the analysis and evaluation process (DP5). In data preprocessing, several steps must be performed for the data sources. Thereby, the procedure is similar for both the relevant module parts collected from the module manuals and the job postings’ relevant parts. Preprocessing includes removing hyphenation and punctuation marks as well as the transformation of umlauts or other language-specific characters and capital letters. Furthermore, 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 and removal of the non-relevant ones), are performed.

### 5.3 Data Analysis via Topic Modeling

As we consider unstructured data, we rely on text mining methods to process the data for the next step. Integrating our task definition into the framework by Miner (2012) led us to the field of concept extraction. The text mining methods therein are summarized under the term **topic modeling**. The basic idea behind topic modeling is that every document consists of latent topics characterized by specific word allocations. All topic modeling methods try to attain the identification of topics through unsupervised learning. One of the most common topic modeling methods is the Latent Dirichlet Allocation (LDA) method by Blei et al. (2003). LDA is a Bayesian model in which each word in a document is assigned to one or more topics. It includes dirichlet priors for the topic-word distribution, which means a “distribution over distributions” to prevent overfitting. LDA, in a first-round, divides all words from the corpus randomly into a predetermined number of topics. Each word is then checked to see if it matches the rest of the words in the topic. The approach assumes that all other words have been correctly divided into topics. If similar words are included in this topic, the word remains in this topic. If it does not fit, it is assigned to the topic to which it most likely fits. This process is repeated for each word until all words are assigned to the topic representing the best match. We chose LDA because it is a generic model, that allows it to be used on the different types of documents we will consider, as they vary from university module descriptions to job advertisements (Lee et al., 2010; Alghamdi and Alfalqi, 2015). For the data analysis using Topic Modeling (DP5), different techniques are conceivable, i.e., depending on the available data, utilized frameworks: (i) The separate viewing of the data sets and thus the training of two topic models—one for each data set—followed by a merge via a framework. (ii) The training of a topic model for one of the two data sets and the subsequent addition of the documents of the second data set as new documents to the first model. (iii) The joint consideration of both data sets and thus the training of a joint topic model. For our artifact implementation for the IS discipline, we tested all three of them. Ultimately, we chose the third option to compare the two data sources due to the outstanding results (cf. section 5.4).

As the topics resulting from LDA have to be labeled by human experts, for the purpose of grouping the topics and job advertisements we recommend the integration of one or more discipline-specific frameworks that define the relevant job competencies as well as possible job categories (e.g. “software developer” or “project manager” (DP6). However, if such frameworks are not available, they can be created (with additional effort) from the topics and the data. For grouping the university modules, we first recommend the aggregation into the prior module manuals so that they can be investigated based on their original assignments. The resulting module manuals can then be subdivided according to characteristics such as the type of university (e.g. university of applied sciences or university) or type of degree (e.g. bachelor’s or master’s). Once the groupings have been completed in the desired form and level of detail, the data can be analyzed and evaluated according to the defined objectives.

### 5.4 Findings

To illustrate the kind of results that can be achieved using the presented procedural methodology, we outline some findings from our research analyzing universities’ coverage of the industry skill expectations for IS graduates. In 2017 we collected the module manuals of 188 study programs in Germany with standardized module manuals. Also we collected the job postings from four platforms (get-in-it, Absolventa, Monster, and Stepstone) which revealed an initial job posting corpus of 12,875 records. Our analysis is based on 93 German IS study programs (3,752 modules) and 6,848 job postings for IS graduates after filtering and preprocessing. As we trained one topic model out of both data sets, the overall document corpus contained 10,600 documents.
Topic Number Determination

We determined suitable topic numbers for the underlying dataset via methods developed by Cao (2009) and Griffiths (2004). These methods are implemented in the R package LDAtuning. The method formulated by Cao adaptively selects the best LDA model based on density, i.e., optimizes towards minimum. Griffiths’ method devised uses the Gibbs sampling algorithm to evaluate the consequences of changing the number of topics in different runs, i.e., optimizing towards the maximum. Put simply, the overall approach in analyzing the methods’ results to find extrema. Since one method is optimized to minimum and the other to maximum, to be able to compare them, we standardized and normalized both to the range between 0 and 1. Given the top aggregated values of both methods (cf., Table 2), we chose 300 as topic number parameter for topic modeling.

<table>
<thead>
<tr>
<th>Value</th>
<th>0.9979</th>
<th>0.9977</th>
<th>0.9975</th>
<th>0.9974</th>
<th>0.9973</th>
<th>0.9973</th>
<th>0.9968</th>
<th>0.9967</th>
</tr>
</thead>
<tbody>
<tr>
<td># Topics</td>
<td>300</td>
<td>303</td>
<td>322</td>
<td>299</td>
<td>194</td>
<td>210</td>
<td>263</td>
<td>282</td>
</tr>
</tbody>
</table>

Table 2. Methods’ Result Excerpt: Best Aggregated Values

Topic modeling, labeling and framework design

After performing LDA topic modeling with MALLET (McCallum, 2002), the resulting 300 topics were labeled by three experts independently. Each of them checked the top 9 words/n-grams of the topic. We used the top 9 words of each topic because cognitive science literature identified 7±2 as a proper upper size for artifacts browsed by humans (Miller, 1956). We aggregated the topics by using three frameworks: (i) the Guidelines for Education in Business and Information Systems Engineering at Tertiary Institutions (Jung and Lehrer, 2017); (ii) the European e-Competence Framework (CEN, 2014); and (iii) the Skills Framework for the Information Age (SFIA Foundation, 2015). We describe the framework design process in detail in Föll et al. (2018). These frameworks reflect a European point of view on which topics (e.g. databases) constitute the main areas of the IS discipline (e.g. information systems).

![Figure 6. Skill Framework for Combined Data Sets](image)

Adding our received topics to a recombination of these frameworks, we developed the framework depicted in Figure 6 in order to represent the specific skill distribution for the combination of industry skill demand and university skill provision. The highest level of abstraction is divided into hard and soft skills. Hard skills are skills or competences that can be tested, while soft skills are primarily interpersonal skills. Soft skills are divided into social and personal skills. The main sections are divided into further subitems, with the related topics located at the lowest level. The number in brackets indicates the number of topics assigned to a given block.

The job postings are grouped by aligning with the European Professional Profiles (EPP) framework (CEN, 2012). This framework divides the whole IS/ICT (information & communications technology) job market into six profile families (business management, technical management, design, development,
service & operation and support) and consists of 23 IS job profiles (cf., Figure 7). The university modules were aggregated to their original module manuals and then grouped according to university type and degree. In the following, we (i) present an overview of how industry’s demand for competencies distributes across the individual EPP at a high aggregation level of the topic framework. (ii) We show insights into the universities’ coverage of the labor market requirements for IS graduates.

**Figure 7. European Professional Profiles**

**Industry demand for competencies**

The overview of industry’s demand for competencies reveals this demand’s distribution according to the EPP’s job profiles. As Figure 8 shows, the demand’s focus is in the management topic aggregation. In 13 profiles it has the largest share. Where management does not represent the most vital characteristic, it nevertheless has a large share in the profile’s topic distribution. Followed by the topic aggregations for operations (largest share in four profiles), development (three profiles) and consulting (one profile). On average, the topic aggregations with the lowest proportion account for only two percent of the distribution, so 98 percent of the requirements depicted in the job postings are dedicated to other topics. The fact that those aggregations are primarily social and personal competences may be surprising at first. One possible reason for this would be that only a small proportion explicitly demands scientific working skills, which are part of employers’ aggregations. Another reason may be that, for example, transfer and presentation abilities are also implicitly assumed and thus hardly mentioned in the descriptions. Here, the part of the procedural methodology comprising interviews and surveys with academic and industry experts can contribute valuable insights into the extent to which the assumption corresponds to reality.

**Figure 8. Industry skill demand of individual job profiles**
Universities’ coverage of the labor market requirements for IS graduates

Figure 9 depicts the universities’ coverage of industry’s skill demand for IS graduates in general. It shows the values of the two university types “university for applied sciences” (UAS) and “university” (UNI), divided into the degree types “bachelor” and “combined bachelor and master”. To enable comparison with the job profiles, we used a method based on the best values’ average to aggregate the modules. We did this to compensate for the different structures of the two document groups, i.e. the specialization of the job profiles (high percentage for few topic aggregations, e.g. programming for profile developer) and the general orientation of the module manuals (small percentage for all topic aggregations, as a module manual covers all topics to some degree). Coverage is indicated if the values depicted in Figure 9 correspond with adequate-high positive values compared to the share of the corresponding topic aggregation on the demand side (cf. Figure 8).

Figure 9. Industry skill demand covered by universities of applied sciences and universities

With regard to the requirements of the labor market, for studies at universities of applied sciences, the consecutive master offers added value to the bachelor’s program in only four of the 20 topic aggregations. Concerning the universities, this picture is changing somewhat. Compared to the bachelor’s degree, the consecutive master offers added value in nine out of 20 topic aggregations. Overall, however, the bachelor’s degree for both types of higher education institutions is better suited to the labor market requirements, as the chart illustrates. The aim of the Bologna process—to create a vocational qualification with the bachelor’s degree and to design the master for deepening scientific knowledge—seems to have been taken into account in the design of the curricula. The results thus indicate that a bachelor’s degree prepares one sufficiently for the demands of professional life. Having presented these brief insights into the possible range of analysis and results, we proceed with the possibilities of evaluating the text mining procedure.

5.5 Evaluation and Validation

To evaluate the results achieved via text mining, we propose a procedure based on surveys and interviews in feedback with the topic modeling process. The effectiveness of the procedural methodology can be tested with the hypotheses (DP1 - DP6) previously established. For the IS discipline, we would structure the evaluation and validation as follows. Initially, we conduct a survey with experts from science and industry. The stakeholders’ involvement seems to be useful as they have an inherent interest in the results (Petrova and Claxton, 2005). As experts from academia, we consider professors of the examined study programs. As experts from industry, we choose managers and recruiters from enterprises in our job postings data set. In a first step, both groups receive a questionnaire based on Trauth et
al. (1993) and Noll and Wilkins (2002). The competences available in the questionnaire should be identical to those identified in the topic modeling. However, it is possible to add additional competences. Industry experts should assess the importance of the different IS competences for their businesses. The academic experts are also asked to explain the importance of skills for IS graduates and indicate to what extent the skills are integrated into their modules. The advantage of this approach is that adding additional competencies makes it possible to evaluate the fitting accuracy of the topic model and the hypotheses set out in the testable propositions section. Suppose many competences are added, which means that they have not been captured during topic modeling. In that case, the model-specific parameters have to be adjusted in order to achieve a better alignment. Finally, the overall results from surveys and topic modeling will be presented to the stakeholders in order to interpret the findings of the comparison between the universities’ supply and the industry skill expectations. Aligning with Bogner et al. (2014), we would perform this step as theoretical expert interviews.

6 Conclusion

As a result of digitalization across organisations, value chains and entire industries, the working world is undergoing profound change. Competences that are in demand today may become obsolete tomorrow. Even universities cannot resist this change. Therefore, it is important to have instruments available with which the contents of the curricula can be compared, assessed and, if necessary, adapted to the changing requirements of the working world. Due to these needs and the lack of appropriate approaches, we present a novel procedural methodology for collecting, analyzing, evaluating and comparing curricula and job postings using topic modeling. We designed the procedure to be applied to a variety of disciplines, job markets, countries, and time periods. To the best of our knowledge, our methodology is the first to allow determining and comparing the contents of study programs and job market requirements, regardless of the considered discipline, with a high degree of automation through the use of combined topic modeling. Thereby our proposed methodology provides several advantages compared to the traditional curriculum revision process. In particular, our procedure can shorten the curriculum revision process by automating the very time-consuming manual evaluation of module handbooks. After an initial expert evaluation, it can be re-executed on a regular base (e.g., once a year) to monitor changing skill demands in a specific domain. Thus our procedural methodology combines the quantitative aspects of the curriculum revision process with the qualitative aspects whereas the quantitative part becomes automatable and replicable.

We demonstrated the practical application of our approach by the example of IS-related study programs, but the procedural methodology itself is generalizable and can be applied to arbitrary disciplines, subject areas, or periods. Different standardization procedures for the considered curricula should be taken into account in specific applications. However, we do not expect this issue should to cause problems with regard to the data required. The same holds for other sources module manuals or job postings. Only documents written in uncommon languages might turn out to be problematic regarding the implementation of different code packages, as the latter may not be available for these languages. Therefore, these packages would first have to be developed, which entails major efforts. Depending on the research goal and the desired results, the topic modeling step itself is highly flexible due to the different model training possibilities (cf. section 5.3). Thus, a broad application of the procedural methodology to other disciplines should be feasible without major modifications.

As regards topic modeling in general, the implementation of our procedure for the IS discipline showcases that even a high number of topics (i.e. 300) is suitable for such a kind of analysis. Consequently, our findings are in contrast to the intuitive expectation that only a small number of topics (approx. 10 to 50) lead to human-readable and -interpretable results (Debortoli et al., 2016). Therefore, we argue that topic modeling with a high topic number should not be generally excluded, since the increased plurality of topics can provide meaningful insights.

With regard to the IS discipline, the results based on a large data set of IS curricula and job postings for IS graduates provided insights into the industry demand for specific competencies and the universities’
coverage of the labor market requirements for IS graduates. Our results showed that the examined universities fulfill most of the overall job market requirements. Possible gaps in the curricula may exist to some extent in the topic aggregations consulting and operations. However, for the operations aggregation, the question arises as to whether we really see a gap or rather a deliberate focus by the universities. A subsequent survey among experts should reveal the answer and validate the other findings. Academic management can, however, already use the results and the underlying procedure when updating the IS curricula and incorporate them into the curricular design process. For practitioners, the results provide insights into the coverage of their requirements by the various universities and degrees. This enables them to identify graduates from universities who are most likely to fulfill their requirements. The results can also be used to dispel concerns about bachelor’s graduates. Based on our results and data, we conclude that these fulfill most of the skill requirements that are expected from graduates in the job market.

The limitations of our procedural methodology 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 (i.e., LDA) results from the parameters used for the topic modeling, such as the number of topics. In the first place, for the underlying data, we use job postings for our analysis. It is recognized in the literature that job postings can be seen as a surrogate for actual industry skill demand. However, we have to be aware that these postings may not always present a full and accurate picture of the skills sought by employers, who may use unrealistic or exaggerated vocabulary in the descriptions of the jobs to be occupied. Second, we use module manuals of accredited study programs. Although the module descriptions are standardized due to their associated institutions’ accreditation, they differ in their extent. For example, the descriptions of some modules comprise only a few lines, while others have several pages. We acknowledge that such biases regarding the job postings and module descriptions may exist in our present or following research. Nevertheless, due to the large amount of data the procedure is based on, we are confident that the effects have only minimal influence on the results. An advantage of processing such a broad range of job postings and module descriptions over other research methods, such as interviews, is the minimization of distortion caused by, for example, different contextual backgrounds (Debertoli et al., 2014). One more point 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 content taught during the lectures. Nevertheless, our daily lecturing experience gives us confidence that there is a close correspondence in this respect.

Overall, the practical demonstration of our procedural methodology showed convincing results. This includes the text mining approach itself as well as its specific implementation for the IS discipline and its transferability towards other application fields. Our next step will be a detailed analysis and presentation of the findings with regard to the IS discipline, which here were discussed only briefly. As we presented our procedural methodology in detail and made our code public available (www.bitbucket.org/tmhelper/tmh), we now encourage researchers to transfer the procedure to other disciplines, labor markets or time periods to reveal their curricular alignment with industry skill expectations. In particular, we believe that our procedural methodology can be used to answer questions regarding (i) alignments of educational offerings with the labor market (or vice versa) over time or (ii) the influence of the regional context (specifically regional companies) on the design and the profile of university curricula.
References


Föll et al. /Exploring Skill Demand and Provision


