



A Semi-automated Approach to Categorise Learning Outcomes into Digital Literacy or Computer Science

Stefan Pasterk^(✉), Max Kesselbacher, and Andreas Bollin

Department of Informatics Didactics, Alpen-Adria-Universität Klagenfurt,
Klagenfurt, Austria

{stefan.pasterk,max.kesselbacher,
andreas.bollin}@aau.at

Abstract. Computer science related curricula, standards and frameworks are designed and implemented in many countries to incorporate informatics education in schools, already starting with kindergarten and primary education. A recurring point of discussion addresses the focus of those educational models concerning the different fields of computer science - the topics related to the scientific subject of computer science, or digital literacy (the set of skills and competencies needed in everyday life in the digital age). In this paper, we present a semi-automated approach to categorise learning outcomes of computer science related curricula into one of those two categories. Categorisation is performed with linguistic metrics computed for nouns and verbs of representative curricula of each category. The categorisation is compared against classifications of nine experts of computer science teaching and research. The results show a matching categorisation for 70% of all learning outcomes and 90% of learning outcomes uniformly classified by the experts.

Keywords: Curriculum · Computer science · Digital literacy · Natural language processing · Primary education

1 Introduction

The incorporation of topics, skills and competencies related to computer science and computer literacy in primary education is currently in focus worldwide [1]. Curricula, standards and frameworks related to computer science are designed and implemented in many countries. The developed curricula differ in many aspects. A possible distinguishing factor of comparison is the focus between computer science and digital literacy categories. The former typically includes topics related to the scientific subject of computer science, while the latter includes skills and competencies needed in everyday life in the digital age. Experts are still in discussion about how the two terms should be correctly classified, which of the two should be focused, and where to draw the distinguishing line between the two categories regarding the formulation of learning outcomes. One of the problems that arises from those open discussion points is that the number of formulations and learning outcomes overwhelms researchers and curricula developers who seek to determine the focus of a curriculum. Curricula often include

between dozens and hundreds of formulations and learning outcomes, and manually classifying them is a tedious work [2]. This paper describes a semi-automated approach to categorise learning outcome formulations into computer science or digital literacy categories. Natural language processing (NLP) techniques [3, 4] are applied to analyse learning outcomes and extract categorisation features of representative curricula for both categories, building dictionaries of verbs and nouns with their respective fraction of occurrence in learning outcome descriptions. The approach is evaluated by categorising four computer science related curricula and comparing the results to a classification of experts of computer science research and teaching. The results show that it is possible to determine the focus of a curriculum with the NLP-based categorisation approach.

The remainder of the paper is structured as follows. Section 2 presents related work and contrasts the presented approach. Section 3 covers the educational models used for analysis and evaluation. Section 4 presents the experimental setup and the results. Section 5 discusses the results, an application and possible implications. Section 6 summarises the contribution of the paper.

2 Related Work

With the recency of digital technology emergence, the concepts of computer science are in the focus of researchers worldwide, especially for primary education. Most of the resulting articles concerning computer science related curricula focus on one single curriculum and describe this, possibly new, approach in a detailed way. A few other publications analyse and compare different curricula for either primary or secondary education, although most curricula combine those two levels. The article from Barendsen et al. [1] focuses on computer science concepts in K-9 education (from kindergarten to school level 9) and considers curricula from England, Italy and the United States (US). To analyse the curricula, the learning outcomes of the documents are grouped into knowledge categories with the help of open coding. The occurrences and distribution of the codes within the knowledge categories are calculated and presented to compare the curricula. With the goal of designing a primary school curriculum for computer science and programming, Duncan and Bell [5], in a first step, compared different related curricula. For this purpose, they chose the main English-language curricula for the primary school level, the Computer Science Teachers Association (CSTA) K-12 Computer Science Standards [6], the England computing curriculum, and the Australian Digital Technologies curriculum [7]. To identify possible key ideas and to show similarities as well as differences, the elements of the curricula were categorised into six content themes [5].

An overview of the global situation of K-12 education is given by Hubwieser et al. [8]. They use articles that discuss the situations in different countries as a corpus. Following the steps for qualitative text analysis, the corpus is categorised using the tool MaxQDA. They collected 249 competence statements and analysed knowledge elements like ‘Algorithm’ regarding the verbs used in combinations with them; as we will see later, this was a step that was also relevant for our work. The statements of the ‘Goals’ category were manually preprocessed and collected into

content categories. Afterwards, they compared those new categories and showed which were covered in which countries [8]. The authors used a manual qualitative analysis approach to extract, categorise and summarise text passages with different topic foci from research texts. In this paper, we present an approach for semi-automatic extraction and categorisation of learning outcome descriptions from curricula documents. Instead of a categorisation considering computer science topics, we focus on the comparison of learning outcomes regarding computer science and digital literacy categories.

3 International Educational Models

Different educational models vary in organisational circumstances, learning goals, topics and teaching methods [9]. With a high number of educational models, the number of basic pedagogical approaches used also rises. Some of them are based on learning objectives or statements. Most of them differ in formulation, details and volume. In this contribution, the umbrella term ‘learning outcome’ is used to collect all the statements, and the following definition is used: “Learning outcomes are statements of what the individual knows, understands and is able to do on completion of a learning process [10].” This definition suffices for the purpose of this contribution as the focus is on the used words and word combinations, not the structure or the volume.

3.1 Selected Curricula, Educational Standards, and Competency Models

Following the related work, two of the main English-language educational models for computer science in primary education, the CSTA Computer Science Standards from 2011 and the Australian national curriculum for Digital Technology [5] are selected for this contribution. As a recent update, the new CSTA Computer Science Standards from 2017 [11] and due to locality, the curriculum 21 from Switzerland [12] were added. The selected curricula, educational standards and competency models are briefly described.

CSTA K-12 Computer Science Standards (2011). The CSTA K-12 Computer Science Standards from 2011 [6] are well known and often referenced in relevant literature [1, 5, 9]. They start with the kindergarten and last until the twelfth grade. A combination of the levels K-3 and 3–6 covers an age range comparable to primary education. These levels include 45 standards, 16 for levels K-3 and 29 for levels 3–6.

Australian Curriculum (AC). As part of the learning area ‘Technologies’, the subject Digital Technologies was presented in Australia in 2013 [7]. It is an obligatory subject from the first school year called Foundation (F) until the eighth year. The ninth and tenth year is elective. The learning outcomes are described for each level, representing two school years. That means levels F-2, 3–4 and 5–6 cover the age range of primary education. For this range, 22 learning outcomes can be found, six of them belong to level F-2, seven to 3–4, and nine to 5–6 [13].

Curriculum from Switzerland (21). In Switzerland, the new curriculum for primary and lower secondary education called ‘Lehrplan 21 (curriculum 21)’ was presented and

established in 2014 by 21 of 26 cantons with the possibility of individual adaptations [12]. It includes the subject ‘Medien und Informatik (Media and Informatics)’ from the first school year on. The levels of this curriculum are represented by ‘cycles’ containing three to four school grades. For primary education, it contains overall 44 competence levels, 14 for cycle 1 and 30 for cycle 2.

CSTA K-12 Computer Science Standards (2017). The reworked CSTA K-12 Computer Science Standards were presented in 2016 as a draft version and published in 2017. They differ from the older version in a lot of aspects, such as the levelling system and the strands. Considering primary education, the level 1A (age range from five to seven years), and level 1B (age range from eight to eleven years), are of interest. It contains 39 standards for primary education, 18 in level 1A and 21 in level 1B [11].

3.2 Categorisation of Learning Outcomes

The categorisation of the learning outcomes is an often-applied method to compare educational models [1, 5]. In most cases, the categories represent areas of interest, like ‘Algorithms’. This contribution looks at two more general categories to identify the focus of the selected educational models: ‘computer science’ and ‘digital literacy’.

Considering the different terminology used in computer science related educational models, it is necessary to clarify and define the term ‘computer science’ (CS) for this contribution. In English-language countries ‘computer science’ is a common term, especially in the US and Australia. In Europe, the term ‘informatics’ is frequently used. For this contribution, we use these terms synonymously, following the definition from the UNESCO/IFIP Curriculum 2000 [14]: “The science dealing with the design, realization, evaluation, use, and maintenance of information processing systems, including hardware, software, organizational and human aspects, and the industrial, commercial, governmental and political implications of these.” This contribution builds on this definition of computer science and uses the abbreviation CS.

The terms ‘digital literacy’ and ‘digital competence’ can be used synonymously. In the ‘DIGCOMP Framework for Developing and Understanding Digital Competence’ in Europe [15], ‘digital competence’ is defined as “the confident, critical and creative use of [information and communication technologies] ICT to achieve goals related to work, employability, learning, leisure, inclusion and/or participation in society. Digital Competence is a transversal key competence which enables acquiring other key competences (e.g. language, mathematics, learning to learn, cultural awareness)”. In the following sections, this contribution will use this definition of digital literacy and refer to it as DL.

4 The Experiment

This contribution presents a semi-automated approach to categorise learning outcomes of different educational models with the aim of gaining information about their foci. To evaluate our approach, in a first step, experts were asked to categorise the learning outcomes into CS and DL using a questionnaire. The process and first results of this

step have already been described by Pasterk and Bollin [2] and are summarised and extended in Sect. 5.1. In a second step, a categorisation with the help of natural language processing based on linguistic features is applied on the same learning outcomes. This process, the results and a comparison to the results from the experts' categorisation are presented in Sect. 5.2.

4.1 Learning Outcomes Classified by Experts

As described by Pasterk and Bollin [2], a group of nine experts, consisting of four computer science teachers and five researchers in the field of computer science education, participated in a survey to categorise the learning outcomes of three selected educational models. To get a larger basis for the evaluation of the semi-automated approach, the survey was repeated with the same group of experts and with two additional educational models, the CSTA computer science standards from 2011 and from 2017. Every expert completed a questionnaire including all learning outcomes of the selected models for primary education in a random order and had to choose one of the following categories: 'CS', 'DL', 'Both' or 'None'. Further, they were asked to describe their strategy for the categorisation process.

Experts' Strategy. Considering the answers of the experts regarding their strategy, seven out of the nine experts referred to the definitions of CS or DL. Six experts used keywords that they assigned to either CS or DL. Finding keywords or key terms was the way to categorise for two experts. Two other experts focused on the topics of the learning outcomes and the combined objectives, which were often defined by keywords. Eight out of nine experts took keywords into account during categorisation.

Results of Classification. First results have already been presented by Pasterk and Bollin [2] and are summarised and extended in Table 1. The added results for the CSTA curriculum from 2011 and from 2017 can also be found in Table 1. The general categories CS and DL were determined by majority votes. Because of the possibility to choose 'Both', this method can lead to undecided learning outcomes. However, this concerned only a few learning outcomes, as can be seen in Table 1. Additionally, the learning outcomes where there was a strong agreement between the experts are included in Table 1. For those, at most, a single expert disagrees with the common classification. Overall, the inter-rater agreement value (Fleiss's kappa) is 0.43 and shows a 'fair to good' agreement, following the interpretation guidelines of Fleiss [16].

Table 1. Summary of the results from the experts' categorisation

	AUS	21	CSTA 11	CSTA 17
Number of LOs	22	44	43	39
CS	10	10	10	20
CS (Strong Agreement)	7	3	0	9
DL	11	32	32	18
DL (Strong Agreement)	2	19	16	5
Undecided	1	2	1	1

Discussion. The results of the experts' categorisation show that the selected educational models can be grouped into the two types 'focus on digital literacy' and 'balanced orientation'. As can be seen in Table 1, the Australian curriculum and the CSTA standards from 2017 have nearly a uniform distribution between CS and DL. Whereas, more than two-thirds of the learning outcomes from the curriculum 21 from Switzerland and the CSTA standards from 2011 were categorised into DL. Following the majority of the experts, those two educational models focus on DL.

4.2 Categorisation by Linguistic Features

We now present an automated categorisation approach based on linguistic features to assign a category, either CS or DL, to each learning outcome of the four analysed curricula. The categorisation results are evaluated against the expert classification.

Linguistic Processing for Analysis. The analysed curricula are available as portable document format (PDF) documents. Manual preprocessing was done by extracting the texts of the learning outcomes. The extraction is implemented in Python. The process of extraction of the linguistic features included the following basic techniques: normalisation of words to improve comparability (lowercase, lemmatising); stop word removal with a list of English stop words; and word tokenising to produce term lists. Each learning outcome text constitutes a single element, called document, in the analysis. Tagging is applied with a trained part-of-speech tagger [17] to extend the words with part of speech categories. The learning outcome text is tagged in full sentence form - 'the students will be able to' is added at the beginning. Tags are grouped; one of the following categories is assigned to each word: noun, verb, adjective, adverb, and other. After tagging, the sentence start is removed, and the tag category and lemmatised words are stored as term list for each learning outcome.

Categorisation Process. For the categorisation, linguistic frequency measures of curricula representative for CS and DL education are computed, following the same linguistic processing, and stored in two dictionaries. The 'Computer Science Curricula 2013' (AIE) [18], created by a cooperation of the Association for Computing Machinery (ACM) and the Institute of Electrical and Electronic Engineers (IEEE) members, contains a set of curriculum guidelines for undergraduate CS programmes and is used to build the dictionary for CS. For DL, the dictionary is built from three curricula. 'The Digital Competence Framework for Citizens' (Dig) [19] designed by the European Union in 2017, 'British Columbia Digital Literacy Framework' (BC) [20] from the Province of British Columbia, and 'Digi.Komp' (DK) [21] from the Austrian initiative for digital competencies and informatics education.

The considered linguistic features include term frequency (TF), term frequency over inverse document frequency (TF-IDF) and document frequency (DF) [3, 4]. The metrics TF and TF-IDF performed poorly for the categorisation because of size differences in the dictionaries. For categorisation, the DF value is used. In context, this value describes in which fraction of learning outcomes a term occurs.

For each learning outcome to be categorised, the sum of the DF values of the occurring tagged terms in the two dictionaries was computed. Insights from the experts' strategies suggested that content terms (nouns), cognitive activity terms (verbs) and

their combinations should be considered. In this contribution, individual terms tagged as nouns and verbs are counted. The highest value determines the category. When both sums are within 10% of the highest value, a third category ‘undecided’ is assigned. Figure 1 shows an example of this process.

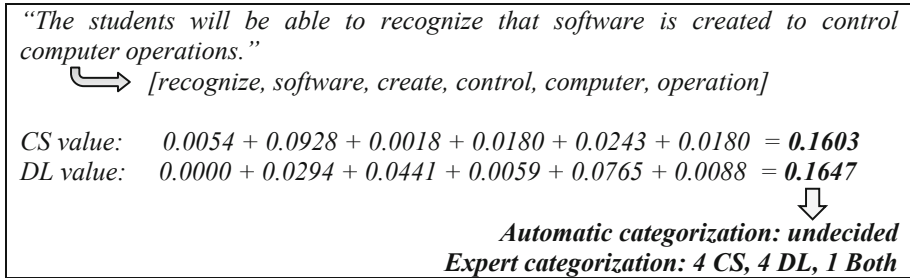


Fig. 1. Example for learning outcome processing from CSTA standards from 2011 [6]

Comparing Categorisation Results to Experts’ Classification. The automated categorisation results are compared to the expert classifications in two ways. First, the categorisation is compared against the expert classification regarding all learning outcomes of the curricula. The results show the categorisation performance for a wide range of learning topics. Second, the categorisation is separately compared against the classification of learning outcomes for which the experts showed a strong agreement.

Table 2 summarises results of all possible combinations of different sets of the curricula used for building the dictionary for DL categorisation. The results show the fraction of matching categorisations. E.g., row five shows the approach achieves a match with expert ratings for 74% of all learning outcomes of CSTA 11 using BC and DK for the DL dictionary. For the categorisation of uniformly classified learning outcomes, this configuration matches in 94% of the DL learning outcomes, and 94% of all those learning outcomes. CSTA 11 does not contain uniformly classified CS learning outcomes, indicated with ‘—’. No single dictionary performs best for the categorisation of all analysed curricula. The best overall categorisation scores are achieved by DK for a single curriculum dictionary, with scores in the range .64 to .75 and a mean score of .70, and by the combination of BC and DK for a multi-curriculum dictionary, with scores in the range .68 to .74 and a mean score of .70. Notably, these two dictionaries perform best for two different sets of analysed curricula.

Regarding categorisation of uniformly classified learning outcomes, again no single dictionary performs best. Measured with the sum of matching categorisation scores, the dictionary built with BC performs best for categorising CS learning outcomes and overall uniformly classified learning outcomes, with mean scores of .81 and .90, respectively. The dictionary built with all three curricula performs best for categorising DL learning outcomes, with a mean score of .99, mismatching one learning outcome.

Table 2. Comparison of results

^a Curricula for DL dict.	AUS, all Agreement			21, all Agreement			CSTA 11, all Agreement			CSTA 17, all Agreement		
BC			.73			.70			.65			.64
	1.0	1.0	1.0	.86	.67	.89	.88	—	.88	.86	.78	1.0
Dig			.73			.57			.67			.67
	.89	1.0	.50	.64	.67	.63	1.0	—	1.0	.71	.56	1.0
DK			.73			.75			.70			.64
	.89	.86	1.0	.91	.67	.95	.94	—	.94	.79	.78	.80
BC, Dig			.73			.64			.72			.64
	.89	1.0	.50	.73	.34	.79	1.0		1.0	.71	.56	1.0
BC, DK			.68			.70			.74			.69
	.89	.86	1.0	.81	.67	.84	.94	—	.94	.86	.78	1.0
Dig, DK			.73			.70			.67			.64
	.89	.86	1.0	.77	.67	.79	1.0	—	1.0	.71	.71	1.0
Dig, BC, DK			.68			.75			.72			.67
	.89	.86	1.0	.91	.67	.95	1.0	—	1.0	.79	.67	1.0

^aThe header denotes analysed curricula. Results rounded to two decimal places use the format [All CS DL] to show overall relative match of categorisation and relative match for competencies with strong expert agreement. Best scores per column and category are marked bold

5 Discussion

With the help of the experts’ categorisation, it was possible to identify the foci of the selected educational models. For automated categorisation, the best performing sets of dictionaries with an accuracy of 70% are AIE for CS and DK for DL or the combination of DK and BC for DL. To identify the foci of the educational models, the numbers and fractions of categorised learning outcomes are presented in Table 3.

Table 3. Application of categorisation with two different dictionaries

		AUS		21		CSTA 11		CSTA 17	
CS	Experts	10	.45	10	.23	10	.23	20	.51
	AIE/DK	7	.32	7	.16	8	.19	19	.49
	AIE/BC, DK	8	.36	7	.16	9	.21	17	.44
DL	Experts	11	.50	32	.73	32	.75	18	.46
	AIE/DK	15	.68	35	.80	31	.72	15	.39
	AIE/BC, DK	9	.41	34	.77	31	.72	17	.44
Undecided	Experts	1	.05	2	.04	1	.02	1	.03
	AIE/DK	0	.00	10	.23	4	.09	5	.12
	AIE/BC, DK	5	.23	7	.16	3	.07	5	.12

As can be seen for the Australian curriculum (AUS), the dictionary based on AIE/DK tends to identify a focus in DL (.68 compared to .32 in CS). Following the results of the dictionary based on AIE/BC, DK this curriculum is balanced (.41 for DL and .36 for CS). This balanced view corresponds to the results from the experts' choices. For the curriculum 21 from Switzerland, a clear focus on DL can be identified with both dictionaries having similar results (.77–.80 in favour of DL). This result corresponds to the experts' categorisation. A similar situation can be seen for the CSTA standards from 2011 where a focus for DL is visible (.72 in favour of DL). Here again, the results from the experts also indicate a focus on DL. Following the experts' results, the CSTA standards from 2017 tend to be balanced, which is also reflected by the semi-automated generated results (.39–.49 for CS and .44 for DL). Summarising, the semi-automated categorisation matches the experts' opinions in the identification of the focus for the majority of the analysed educational models. In three cases, both dictionaries of the semi-automated approach identified the same foci as the experts did. In one case, only the results from the dictionary based on AIE/BC, DK corresponded with those of the experts. An important conclusion is that the quality of categorisation is highly dependent on the curricula used for building the categorisation dictionary.

The semi-automated approach presented in this contribution shows a few threats to validity. The translation of the learning outcomes from the German-language curriculum can lead to the use of different terms. This can result in a lower frequency of important terms. Because all of the experts were from Austria, it can also be the case that they are biased by the local, well-known 'digikomp (DK)' competency model which was chosen to build the dictionaries. This could be a factor resulting in the higher performance of the dictionary based on DK. Another threat can be that expert categorisation can arise from misinterpretation, invalidating the comparison.

6 Conclusion and Future Work

Educational models are designed and implemented on different levels in many countries. These models include national curricula, workgroup recommendations, competency frameworks and guidelines. There is ongoing discussion about whether this newly implemented education trend should focus on topics related to the scientific subject of computer science or the development of skills and competencies needed in everyday life in the digital age. In this paper, we present an approach to semi-automatically categorise learning outcomes of computer science related curricula into computer science or digital literacy categories. For each of the categories, a dictionary of noun and verb terms of curricula representative for the category was built. The value of relative frequency of each term in all learning outcomes of the dictionary was used as a categorisation metric. The categorisation was applied to four computer science related curricula, and results were compared against classifications from nine experts of computer science teaching and research. The best performing dictionaries achieved a matching categorisation of 70% of all learning outcomes of the analysed curricula. Furthermore, for learning outcomes which were uniformly classified by the experts (at most one expert disagreed), the best performing dictionary achieved a matching categorisation of 90% of those learning outcomes. The results suggested that the focus of a

curriculum regarding the two categories, computer science and digital literacy, can be identified with the application of the approach. Our goal for future work is an automatic classification of computer science related curricula and the individual learning outcomes regarding different categories. Going forward, we intend to take into account the verb and noun phrases for categorisation, following the general strategy of the experts. Additionally, we want to compare our approach of categorisation with machine learning classification and with sets of additional linguistic features. We also plan to evaluate them against a larger set of expert ratings, including additional experts.

Acknowledgments. The authors would like to thank Günther Fliedl for his feedback on the natural language processing methods and on preliminary studies.

References

1. Barendsen, E., et al.: Concepts in K-9 computer science education. In: ITiCSE-WGR 2015, pp. 85–116 ACM, New York (2015)
2. Pasterk, S., Bollin, A.: Digital literacy or computer science: where do information technology related primary education models focus on? In: Jakab, F. (ed.) Proceedings of the 2017 ICETA, pp. 1–7. IEEE, Stary Smokovec, Slovakia (2017)
3. Han, J., Kamber, M.: Data Mining: Concepts and Techniques, 2nd edn. Morgan Kaufmann, San Francisco (2006)
4. Bird, S., Klein, E., Loper, E.: Natural Language Processing with Python, 1st edn. O'Reilly Media, Sebastopol (2009)
5. Duncan, C., Bell, T.: A pilot computer science and programming course for primary school students. In: Proceedings of the 2015 WiPSCE, pp. 39–48. ACM, New York (2015)
6. Seehorn, D., et al.: CSTA K-12 Computer Science Standards: Revised 2011. Technical Report. ACM, New York (2011)
7. Falkner, K., Vivian, R., Falkner, N.: The Australian digital technologies curriculum: Challenge and opportunity. In: Proceedings of the Sixteenth Australasian Computing Education Conference, Australian Computer Society, vol. 148, pp. 3–12 (2014)
8. Hubwieser, P., et al.: A global snapshot of computer science education in K-12 Schools. In: ITiCSE-WGR 2015, pp. 65–83. ACM, New York (2015)
9. Hubwieser, P., et al.: Computer science/informatics in secondary education. In: Adams, L., Jurgens, J. (eds.) ITiCSE-WGR 2011, pp. 19–38. ACM, New York (2011)
10. ECTS: Users' Guide (2015). https://ec.europa.eu/education/ects/users-guide/docs/ects-users-guide_en.pdf
11. CSTA: Standards (2017). <http://www.csteachers.org>
12. Lehrplan 21: Lehrplan 21 (2014). <http://v-ef.lehrplan.ch>
13. Australian Curriculum: Digital technologies in focus (2013). <http://www.australiancurriculum.edu.au/>
14. Anderson, J., van Weert, T., Duchâteau, C. (eds.): Information and Communication Technology in Secondary Education: A Curriculum for Schools and Programme of Teacher Development. UNESCO, Paris, France (2002)
15. Vuorikari, R., Punie, Y., Carretero Gomez, S., Van den Brande, G.: DigComp 2.0: The Digital Competence Framework for Citizens. Update Phase 1: The Conceptual Reference Model. Publication Office of the European Union, Luxembourg (2016)
16. Fleiss, J.L.: Statistical Methods for Rates and Proportions, 2nd edn. Wiley, New York (1981)

17. Toutanova, K., Klein, D., Manning, C., Singer, Y.: Feature-rich part-of-speech tagging with a cyclic dependency network. In: Proceedings of HLT-NAACL, Association for Computational Linguistics, Stroudsburg, pp. 173–180 (2003)
18. Joint Task Force on Computing Curricula, Association for Computing Machinery (ACM) and IEEE Computer Society: Computer Science Curricula 2013: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science. ACM, New York (2013)
19. Ferrari, A.: DIGCOMP: a framework for developing and understanding digital competence in Europe. IPTS, Seville, Spain (2013)
20. British Columbia: Digital Literacy Framework (2013). <https://www2.gov.bc.ca/gov/content/education-training/k-12/teach/teaching-tools/digital-literacy>
21. Digi.komp: Die Initiative (2011). <https://www.digikomp.at/>