DEGREE PROJECT IN COMPUTER SCIENCE AND ENGINEERING,
SECOND CYCLE, 30 CREDITS
STOCKHOLM, SWEDEN 2017

Analyzing KTH's course syllabuses from a pedagogical perspective

JOAKIM LINDBERG

KTH ROYAL INSTITUTE OF TECHNOLOGY
SCHOOL OF COMPUTER SCIENCE AND COMMUNICATION

Analyzing KTH's course syllabuses from a pedagogical perspective

JOAKIM LINDBERG
joakilin@kth.se

Master in Computer Science
Date: May 29, 2017
Supervisor: Stefan Nilsson
Examiner: Viggo Kann
Swedish title: Analys av KTH:s kursplaner ur ett pedagogiskt perspektiv
School of Computer Science and Communication

Abstract

Course syllabuses are used by students and teachers in various pedagogical situations, which makes it important that the syllabuses are well-designed from a pedagogical perspective. It is therefore useful to classify the syllabuses and identify areas in need of improvement. However, due to the large number of syllabuses it is infeasible to perform the classification manually.

The objective of this thesis is to analyze all current syllabuses at KTH according to how pedagogically useful they are. The analysis is enabled by defining what "pedagogically useful" means and by implementing a program which uses the definition to classify the syllabuses automatically.

The results show that syllabuses fulfill most specifications to a rather high degree, but there is room for improvement in the intended learning outcomes by increasing the usage of constructive alignment and verbs belonging to the higher classes of Bloom's taxonomy. The results also show that there are correlations of varying strength between how well the specifications are fulfilled and which school the course is given by, the cycle of the course and the number of credits. A remarkable result is that not all syllabuses comply with Swedish laws.

Sammanfattning

Kursplaner används av studenter och lärare i olika pedagogiska situationer, vilket gör det viktigt att kursplanerna är välutformade ur ett pedagogiskt perspektiv. Det är således användbart att klassificera kursplanerna och identifiera områden i behov av förbättring. På grund av det stora antalet kursplaner är det dock praktiskt omöjligt att utföra klassificeringen manuellt.

Målet med detta examensarbete är att analysera alla aktuella kursplaner vid KTH utifrån hur pedagogiskt användbara de är. Analysen möjliggörs genom att definiera vad ”pedagogiskt användbar” innebär och genom att implementera ett program som använder definitionen för att klassificera kursplanerna automatiskt.

Resultaten visar att kursplanerna uppfyller de flesta specifikationerna till en ganska hög grad, men det finns förbättringsmöjligheter i lärandemålen genom att öka användningen av konstruktiv länkning och verb tillhörande de högre klasserna i Blooms taxonomi. Resultaten visar också att det finns korrelationer av varierande styrka mellan hur väl specifikationerna uppfylls och vilken skola som ger kursen, kursens nivå och antalet högskolepoäng. Ett anmärkningsvärt resultat är att inte alla kursplaner följer svenska lagar.

List of abbreviations

API: Application programming interface

ESG: Standards and Guidelines for Quality Assurance in the European Higher Education Area

HTML: Hypertext Markup Language

ILO: Intended learning outcome

SUHF: The Association of Swedish Higher Education
Sveriges universitets- och högskoleförbund

XML: Extensible Markup Language

Each of the ten schools at KTH has a three-letter abbreviation:

ABE: School of Architecture and the Built Environment
Skolan för arkitektur och samhällsbyggnad

BIO: School of Biotechnology
Skolan för bioteknologi

CHE: School of Chemical Science and Engineering
Skolan för kemivetenskap

CSC: School of Computer Science and Communication
Skolan för datavetenskap och kommunikation

ECE: School of Education and Communication in Engineering Science
Skolan för teknikvetenskaplig kommunikation och lärande

EES: School of Electrical Engineering
Skolan för elektro- och systemteknik

ICT: School of Information and Communication Technology
Skolan för informations- och kommunikationsteknik

ITM: School of Industrial Engineering and Management
Skolan för industriell teknik och management

SCI: School of Engineering Sciences
Skolan för teknikvetenskap

STH: School of Technology and Health
Skolan för teknik och hälsa

Contents

Appendices

Chapter 1

Introduction

A course syllabus, or simply *syllabus*, is a summary outline of a course [1]. It is used by many parties in various pedagogical situations: prospective students use it to decide if they want to read the course, enrolled students use it to find information about the course during their studies and teachers use it to design the course activities.

Syllabuses are also important for quality assurance. By analyzing a syllabus from a pedagogical perspective it is possible to assess if the course design is of sufficient quality. By analyzing several syllabuses at once such as syllabuses for all courses in a program, it is possible to get a holistic view of the quality of the program and whether the program objectives are fulfilled. It is therefore of utmost importance that the syllabus is useful from a pedagogical perspective.

However, it is not completely obvious what it means that a syllabus is useful from a pedagogical perspective. To enable an analysis of the syllabuses it is therefore necessary to clearly define what "pedagogically useful" means in the context of syllabuses at KTH.

Furthermore, analyzing a large number of syllabuses can be a daunting task. At KTH there are around 2 000 active courses during the study year 2017/2018 which makes it time-consuming and repetitive to analyze all syllabuses manually. Also, if multiple persons perform the analysis there might be a discrepancy between how the analyses are performed which might make the results less reliable. To circumvent these issues, this thesis investigates a way to perform the analysis automatically using a computer program.

Such a program could also facilitate further research regarding syllabuses. Currently, there is ongoing research at the ECE school at KTH in which a list of courses having a connection to research, innovation or the labor market is needed, and this information could be constructed using the aforementioned program.

1.1 Objective

The objective of this thesis is to analyze all current course syllabuses at KTH according to how pedagogically useful they are. The analysis is enabled by first reviewing existing literature to define what "pedagogically useful" means in the context of course syllabuses at KTH, and then implementing a program which uses the definition to automatically classify the syllabuses.

1.2 Problem statement

The following three questions are answered in this thesis:

- How can "pedagogically useful" be defined in the context of course syllabuses at KTH?
- How pedagogically useful are course syllabuses at KTH according to the definition from the previous question?
- Are there any correlations between how pedagogically useful the course syllabuses are and the cycle of the course, which school the course is given by or the number of credits?

1.3 Problem definition

The problem can be divided into three parts:

1. **Define what it means that a syllabus is "pedagogically useful."** This is done by considering existing laws, regulations, recommendations and research regarding syllabuses. The definition consists of a number of variables that are well-defined and measurable.
2. **Classify the syllabuses at KTH according to how pedagogically useful they are.** Since there are around 2 000 active courses at KTH it is not feasible to do the classification manually. Thus, a

program is created that performs the classification automatically using the aforementioned definition.

3. **Analyze the classified syllabuses.** The purpose of the analysis is to answer the second and third questions of the problem statement.

1.4 Delimitations

A syllabus is mandated by law to correctly represent the course it is describing since it would otherwise be useless. For example, the course contents described in the syllabus should match the actual course contents. In this thesis, **we assume that syllabuses correctly represent courses**.

Courses at KTH belong to one of the following four levels: preparatory, first cycle (bachelor), second cycle (master) and third cycle (PhD). Third-cycle courses differ from the other courses by not having course offerings which makes it impossible to know when a third-cycle course was previously given or if it will be given again. Since the objective is to analyze the *current* syllabuses, **we choose to exclude third-cycle courses from the analysis**.

Syllabuses at KTH should be written in Swedish but are also translated to English. Since only the Swedish syllabus is legally binding, we will mainly analyze the Swedish version while only considering certain aspects of the English version that are relevant for fulfilling the specifications in Chapter 2.

1.5 Ethics and sustainability

As mentioned in Section 1.4 syllabuses are supposed to correctly represent the courses they are describing. A low-quality syllabus might therefore be a warning sign of a course being of low quality. Since all students should be entitled to high quality education it is crucial that the syllabuses are of high quality. It is therefore ethically sound to analyze syllabuses from a pedagogical perspective to identify weaknesses which can then be fixed.

A low-quality syllabus in which there is a lack of information or the objectives are poorly written will probably cause problems for students, teachers or other related parties in one way or another. This might lead to the need for correcting the syllabus or perhaps even creating a new one, which could be problematic if the problems occur during critical periods such as when the course is given. By analyzing the syllabuses and performing quality assurance in advance, it is possible to identify and fix weaknesses before they cause any problems which leads to more sustainable syllabuses and education.

Chapter 2

Background

To ensure that the definition of a "pedagogically useful" syllabus is meaningful and well-founded, it is necessary to consider laws, regulations, recommendations and research relevant for the design of syllabuses, which is presented in this chapter. Also, methods from natural language processing that are used for analyzing the syllabuses are presented, followed by a description of how syllabuses can be accessed at KTH.

2.1 Laws and regulations

KTH is obliged, like any other organization in Sweden, to follow Swedish laws and regulations. In this section, national laws and local regulations at KTH that describe requirements on the content and the language of syllabuses are presented.

2.1.1 Content

By Sections 14–15 in Chapter 6 of the *Higher Education Ordinance* (Högskoleförordningen) all first-cycle or second-cycle courses must have a syllabus which specifies

- which cycle the course belongs to,
- number of credits,
- objectives,
- specific eligibility requirements,

- how student performance is assessed and
- other regulations that might be needed. [2]

Local regulations at KTH require that syllabuses are constructed using a special form [3]. In addition to the six points in the Higher Education Ordinance mentioned above, the form contains fields for entering

- course name in Swedish and English,
- course code,
- main fields of study,
- department giving the course,
- when the syllabus is valid from,
- main course content,
- course literature,
- language of instruction,
- grading scale,
- examination modules in Ladok[1],
- requirements for obtaining a final grade and
- whether the number of examination opportunities to pass the course is limited. [4, 5]

2.1.2 Language

The language of government agencies is mandated to be Swedish by Section 10 of the *Language Act* (Språklagen) [6]. Since KTH is a government agency [7] the Language Act mandates that all course syllabuses at KTH must be available in Swedish. Also, the Swedish course syllabuses are legally binding [8].

In addition, syllabuses must be written in English according to local regulations at KTH [3]. The English versions are however not legally binding.

[1] The national student administration system for universities and university colleges in Sweden.

2.2 Recommendations

There are also non-mandatory recommendations that are useful when designing syllabuses. In this section, local recommendations from KTH followed by national recommendations from SUHF and European recommendations in the ESG are presented.

2.2.1 KTH

A working group at KTH has been given the task of developing a new structure for course information at KTH. The suggested structure consists of five parts, one of which is the syllabus. The syllabus should be written in Swedish and English, and the suggested syllabus content is

- course name,
- course code,
- number of credits,
- grading scale,
- which cycle the course belongs to,
- main fields of study,
- when the syllabus is valid from,
- specific eligibility requirements,
- main course content,
- objectives stated as constructively aligned ILOs (see Section 2.3.2),
- examination modules in Ladok,
- comments on the examination forms,
- other requirements for obtaining a final grade and
- ethical approach (e.g. no cheating or plagiarism). [9]

2.2.2 SUHF

The *Association of Swedish Higher Education* (Sveriges universitets- och högskoleförbund, SUHF) is an organization in which the members discuss and cooperate in questions regarding higher education [10]. The members consist of 37 Swedish universities and university colleges, KTH being one of them [11].

One of the activites at SUHF is to decide on and publish recommendations that the members should follow. There is a recommendation (REK 2011:1) regarding syllabuses which states that each syllabus should specify

- course name in Swedish and English,
- which cycle the course belongs to,
- number of credits,
- level of specialization according to SUHF recommendation dnr 08/025,[2]
- main fields of study (if any),
- objectives stated as ILOs (see Section 2.3.1),
- main course content,
- specific eligibility requirements,
- how student performance is assessed,
- grading scale,
- whether the course is split up in parts and if so how many credits each part consists of,
- whether the number of examination opportunities to pass the course is limited,
- when the course syllabus (or changes of it) is valid from and
- any transitional provisions or other regulations that might be needed. [12]

2.2.3 ESG

The *Standards and Guidelines for Quality Assurance in the European Higher Education Area* (ESG) is a set of standards regarding quality assurance in higher education, prepared by the *European Association for Quality*

[2]These levels are currently not used at KTH and will therefore be ignored in this thesis.

Assurance in Higher Education (ENQA) [13]. The ESG was created as a result of the *Bologna process*, which is a voluntary cooperation between 48 European countries with one of its main goals being to ensure high-quality learning and teaching [14].

Each of the 24 standards in the ESG has a guideline which gives further details on the standard such as why the standard is important and how it can be implemented. A standard which is relevant for syllabus design is Standard 1.3 titled *"Student-centred learning, teaching and assessment"*. It states that:

> *"Institutions should ensure that the programmes are delivered in a way that encourages students to take an active role in creating the learning process, and that the assessment of students reflects this approach."* [13]

The guideline for Standard 1.3 mentions, among other things, that the method of assessment should be published in advance and that the assessment should allow students to *"demonstrate the extent to which the intended learning outcomes have been achieved."* [13]

A method which facilitates the implementation of this standard is presented in Section 2.3.2.

2.3 Research on syllabus design

As in most other research situations, it is important to review existing research to ensure that our work is built upon a scientific foundation. In this section, three commonly used concepts in educational development are presented, namely outcomes-based learning, constructive alignment and Bloom's taxonomy. Finally, views on the optimal number of intended learning outcomes are presented.

2.3.1 Outcomes-based learning

Traditionally, the content of a syllabus has been centered on the course content and what the teacher will teach. An alternative approach is to instead use *intended learning outcomes* (ILOs) that describe what students will be able to do after successfully completing the course. This approach is called *outcomes-based learning* and is being increasingly adopted in higher education systems throughout the world. [15]

The motivation for using outcomes-based learning stems from the constructivist learning theory, which states that learners construct new knowledge through learning activities and using their current knowledge, instead of passively receiving information from a teacher. This has heavily influenced a commonly used method for implementing outcomes-based learning, namely constructive alignment. [16]

2.3.2 Constructive alignment

The core idea of *constructive alignment* is that the ILOs of a course should be aligned with the examination tasks. This is achieved by writing the ILOs using *active verbs* that describe what the student will be examined on. [16]

To illustrate this idea, assume that we want to write ILOs for a calculus course. An ILO which is not constructively aligned is that the student should *"show an understanding of Stokes' theorem."* This ILO is problematic since it is unclear what the student should do to demonstrate such an understanding, and it is also unclear what "understanding" actually means. A better ILO which is constructively aligned is that the student should be able to *"evaluate double integrals using Stokes' theorem."* This ILO makes it clear what the student is expected to do and be examined on since there is a clear action ("evaluate") followed by the object of the action ("double integrals").

However, simply using an active verb does not entail how complex the activity is. For example, is the verb "describe" simpler or more complex than the verb "analyze" from an educational point of view? To answer such questions systematically a model for classifying ILOs is needed, such as the SOLO[3] taxonomy [17], Marzano and Kendall's taxonomy [18] or Bloom's taxonomy, the last of which is used in this thesis and detailed below.

2.3.3 Bloom's taxonomy

Bloom's taxonomy is a model for classifying ILOs into different levels of complexity regarding educational behavior. It was developed by Bloom et al. and first published in 1956. [19]

There are three major categories or *domains* of the taxonomy that deal with different types of objectives: [19]

[3]Structure of observed learning outcome

Cognitive: Knowledge, intellectual abilities and intellectual skills.

Affective: Attitudes, interests and values.

Psychomotor: Physical skills.

The cognitive domain has been the most commonly used domain for developing and reviewing educational objectives and will also be the domain used in this thesis. It consists of six different classes that describe educational behavior: [19]

Knowledge: Remembering facts, methods, patterns etc.

Comprehension: Understanding existing knowledge.

Application: Applying existing knowledge to new situations.

Analysis: Breaking down information into parts to find relations between the parts.

Synthesis: Combining existing parts of knowledge to create something new.

Evaluation: Judging the value of using existing information or methods in given situations.

The classes are hierarchical with regards to educational behavior, where lower classes represent more simple behavior and higher classes represent more complex behavior. Knowledge is the lowest class and Evaluation is the highest class. For example, comparing two phenomena (Analysis) is seen as more complex than simply describing facts (Knowledge). [19]

To facilitate the design of ILOs in a certain class of the taxonomy, there are lists that map active verbs to their respective class. Such lists exist in many different languages, including English and Swedish (the relevant languages for this thesis).

While the typical use of Bloom's taxonomy is to design new ILOs, it can also be used as a tool to analyze existing information. Wong et al. analyze over 50 000 messages from a MOOC[4] discussion forum to identify the types of educational behavior occurring in the forum messages by classifying them according to the six classes of Bloom's

[4]Massive open online course.

taxonomy. The classification is done by using a list that maps active verbs to their respective class in the taxonomy and then searching for these verbs in the messages using a program. [20]

2.3.4 Optimal number of ILOs

There is no definitive answer to what the optimal number of ILOs is but many authors have similar suggestions.

- Biggs argues that it becomes increasingly difficult to align the learning activities with the examination as the number of ILOs increases. He states that having five or six ILOs per course is usually appropriate. [16]
- Kennedy reviews what other authors suggest and concludes that around six ILOs for a course is ideal while more than nine ILOs is too much. [15]
- The MIT Teaching and Learning Laboratory states that having at most six ILOs is considered to be best practice. [21]

In summary, having around six ILOs seems to be a good rule of thumb.

2.4 Natural language processing

To be able to answer the questions posed in Section 1.2 in a satisfactory manner, three methods from the field of natural language processing will be used: part-of-speech tagging, lemmatization and language identification. In this section, these three methods are presented together with a program called Granska that implements the first two methods.

2.4.1 Part-of-speech tagging

Words can be grouped into *parts of speech*. One of the first models for describing the parts of speech was developed over 2 000 years ago containing eight parts of speech such as nouns and verbs, and this model has been highly influential to this day. However, not all information about the word can be captured by these eight parts of speech. For example, both "I" and "my" are pronouns but "I" refers to a person while "my" indicates ownership. Thus, newer models use more detailed classifications and can have well over a hundred different parts of speech. [22]

Part-of-speech tagging is the process of marking each word and punctuation in a given text with its corresponding part of speech. There are two main types of part-of-speech tagging algorithms: rule-based taggers which use manually constructed rules to decide the part of speech, and stochastic taggers which calculate probabilities of words having a certain part of speech in a given context. [22]

2.4.2 Lemmatization

A word can either appear in its base form as a *lemma* or in various inflectional forms. For example, “plays”, “played” and “playing” are inflectional forms of the lemma “play”. Given an inflectional form of a word, how does one find its lemma?

A heuristic method for finding the lemma of a word is to cut off the last part of the word using language-specific rules, and this method is called *stemming* [23]. However, stemming can produce erroneous lemmas. For example, a stemming algorithm can suggest that the lemma of “creating” is “creat” whereas the correct lemma is “create”.

In contrast, *lemmatization* is the process of correctly mapping an inflectional form of a word to its lemma. Note that lemmatization is not a deterministic process since a word can have different lemmas depending on the context or on the part of speech. For example, both “find” and “found” are possible lemmas of “found” depending on if the meaning is “to locate” or “to establish”. [22]

2.4.3 Language identification

There are various approaches to identifying the language of a given text from a set of candidate languages.

One approach is to build statistical language models of candidate languages and calculating the probability that the given text has been generated by each of the language models. The language whose model maximizes this probability is chosen as the language of the given text. [24]

Another approach is to create vectors for the text and the candidate languages, where the vector components are the n-gram[5] frequencies

[5] An n-gram is a consecutive sequence of n tokens such as letters. For example, the

of the text or language. The distance between the text and a language can then be calculated using the cosine of the angle between the text vector and the language vector. The language whose vector minimizes this distance is chosen as the language of the given text. [25]

A similar approach by Cavnar and Trenkle, which will be used in this thesis, is to create ordered lists called *profiles* of n-grams ranked by their frequencies in the text and the candidate languages. These profiles are used to calculate the distance between the text and each language by comparing the rank for each n-gram in the language profile to its rank in the text profile. For example, if an n-gram has rank 3 in the language profile (i.e. the third most frequent n-gram in the language) and the same n-gram has rank 8 in the text profile, the n-gram is 5 ranks out of place. The distance is obtained by summing the out-of-place value of the n-grams in the language profile up to a certain rank, and the language with the lowest distance is chosen as the language of the given text. [26]

Cavnar and Trenkle tested the method on 3 478 texts in 9 different languages and found that for texts longer than 300 characters and using up to rank 300 in the language profiles, 99.8 % of the texts are classified correctly. [26]

2.4.4 Granska

Granska is a grammar checking system for the Swedish language that is able to detect errors in a Swedish text and suggest how to correct them. The basic workings of Granska is that it first tokenizes the text, tags each word with its part of speech and lemma, applies different rules to find structure and errors and finally presents suggestions on how to correct potential grammar and spelling errors. [27]

The Granska part-of-speech tagger is a hybrid of a rule-based tagger and a stochastic tagger [28]. It has shown to be one of the best Swedish part-of-speech taggers available with an reported accuracy of around 96–97 % [27, 29].

Granska has a web interface called *WebbGranska* [30]. WebbGranska can output the grammar checked text, along with all words lemmatized and part-of-speech tagged, either in a human-readable format or in an XML document. The XML format outputs one element for each word with its lemma and part of speech as attributes. For example, if the

---

possible letter trigrams (3-grams) of the word *course* are *cou*, *our*, *urs* and *rse*.

entered text is *"jag åt kakor"* (meaning "I ate cookies") a snippet of the output is:

```
...
<w no="0" tag="60" lemma="jag">jag</w>
<w no="1" tag="125" lemma="äta">åt</w>
<w no="2" tag="68" lemma="kaka">kakor</w>
...
```

The `lemma` attribute contains the lemma of the given word and the `tag` attribute contains a reference to the part of speech of the word. This reference can be used to look up the part of speech in a list of elements (which is also part of the output) with the following format:

```
...
<tag no="59" name="vb.inf.akt.aux"/>
<tag no="60" name="pn.utr.sin.def.sub"/>
<tag no="61" name="kn"/>
...
```

The XML document also contains information about suggestions for correcting spelling errors. For example, if we misspell the word *"kakor"* as *"kakorr"* Granska will give the two suggestions *"kakor"* and *"kakors"* in the following format (certain elements omitted to improve clarity):

```
...
<rule>stav1@stavning</rule>
<info>Misstänkt stavfel</info>
<suggestions>
<sugg>kakor</sugg>
<sugg>kakors</sugg>
</suggestions>
...
```

2.5 Syllabuses at KTH

Each course at KTH has a web page displaying course information (including the syllabus) in a human-readable format, and these pages are gathered in the *course and program directory*. In the directory there are multiple ways of finding a given course and its syllabus, such as using the search functionality to search for the course code, listing the courses given by each department or listing the courses that constitute a certain program. [31]

The information available in the course and program directory can also be accessed in an XML format through a REST[6] API provided by KTH. In addition, the API can be used to fetch information not directly available through the course and program directory such as all course offerings in a given term. All supported API requests can be found at [32].

[6]Representational state transfer

Chapter 3

Method

In this chapter, the construction of the definition and its variables are detailed, followed by explanations regarding how syllabuses are accessed, the needed preprocessing and the classification process.

3.1 Defining "pedagogically useful"

At a first glance, it is not completely obvious what the definition of a pedagogically useful syllabus should contain. The definition will probably contain different aspects depending on by whom the definition is constructed, especially considering regional differences. Thus, it would be beneficial to combine local, national and international aspects in the definition.

A good starting point is that syllabuses should comply with national laws since laws are mandatory to follow. Since this thesis is about syllabuses at KTH which is a Swedish university, we should look at Swedish laws. The relevant laws for syllabuses are the Higher Education Act and the Language Act described in Section 2.1. There are also local regulations at KTH described in Section 2.1 that we need to consider.

Furthermore, there are recommendations from both national and international organizations that work with topics regarding higher education such as syllabus content and quality assurance. Two such recommendations mentioned in Section 2.2 are SUHF's recommendations and the ESG, and both will be used in the definition to get a combination of national and international recommendations. Also, local recommendations from the working group at KTH described in

Section 2.2 will be used.

Another important aspect to consider is scientific research regarding how to design syllabuses. The relevant research mentioned in Section 2.3 to be used in the definition is that syllabuses should use an outcomes-based learning approach in form of constructive alignment, use active verbs preferably in the higher classes of Bloom's taxonomy and have around six ILOs. Some of these points overlap with the recommendations previously discussed but this is not a problem; on the contrary, it only shows that the recommendations are well-founded.

To summarize, we have identified the main areas to use in our definition for a pedagogically useful syllabus, namely laws, regulations, recommendations and research. In order to facilitate an automatic classification of syllabuses we need to express the definition in well-defined and measurable terms, which implies that a suitable structure for the definition is to use a set of variables. The variables are presented in Section 4.1 and the definition itself is presented in Section 4.2.

For some of the specifications in Chapter 2, a lack of information could be acceptable and not necessarily meaning that the syllabus is incomplete. In particular, if a course does not have any specific eligibility requirements or main fields of study the corresponding fields in the syllabus might be empty. For this reason we do not include the corresponding variables in the definition of pedagogically useful, but we still measure the variables and present the results since it is of interest to the project provider and might also be useful for further research.

3.2 Accessing the syllabuses

To access the syllabuses it is first necessary to know the course codes, and since we only want to analyze active courses we look at courses given during the upcoming study year of 2017/2018. We can fetch the course codes of these courses using the API mentioned in Section 2.5 by requesting the course offerings for the fall term of 2017 and the spring term of 2018.

We then need to decide on how to access the syllabuses. As mentioned in Section 2.5 the syllabuses can be accessed in a human-readable format using the course and program directory or in an XML format using the API. Since we want to access the syllabuses automatically using a program we are only interested in the data itself and not the presentation of it, which implies that using the API is the more appropriate choice.

Furthermore, using the API leads to lower bandwidth consumption due to the lack of superfluous HTML formatting and JavaScript code, which is good both for the client running the program and the web servers at KTH. For example, the web page for the course DD2448 is almost six times as big as the data extracted from the API (42.9 kB compared to 7.3 kB).

After performing initial testing, it was discovered that the API lacks information about the main fields of study for each course, which poses a problem since we are interested in that information. To investigate the cause, the team maintaining the API was contacted and they confirmed that the API indeed lacks this information. They offered a solution by sending an Excel document containing the main fields of study for each course, so we use this document to extract the information.

3.3 Preprocessing

The syllabuses need to be preprocessed before they can be classified and analyzed, and the needed preprocessing is explained in this section.

3.3.1 Parsing the XML document

Since the responses from the API are XML documents, they have to be parsed before they can be used. We choose to do this using the Java library *jsoup* [33] which can parse an XML document to a Document Object Model (DOM) tree structure. We can then access the wanted information in the syllabus by traversing the DOM tree and extracting the contents of text nodes.

The text nodes are encoded using HTML entities since they are meant to be presented in HTML on the KTH web pages. For example, the letter "ä" is encoded as the entity `ä` in the XML document. Thus, all HTML entities need to be decoded to their plain text representations before the text can be analyzed. We choose to do the decoding using the Java library *Apache Commons Lang* [34] which has methods for decoding HTML and XML entities.

Some syllabuses contain characters which were probably added by mistake by the syllabus authors or their text editors, such as invisible

Unicode control characters or remains of HTML tags, and we therefore remove these characters before performing any analysis. There are also other peculiarities such as the Swedish characters "å", "ä" and "ö" being written in an unorthodox way (e.g. writing "ä" as the concatenation of "a" and the Unicode character "Combining Diaresis" instead of a single character) so we perform normalization for these cases to ensure that all characters are represented in the same way.

3.3.2 Identifying the ILOs

The ILOs are located in a tag called `goals` in the XML document. Unfortunately, the XML structure of the `goals` tag content varies greatly between syllabuses which poses a difficulty for automatically identifying each ILO separately. Some examples of structures of the `goals` tag content are:

1. Begins with a paragraph similar to *"After a completed course, the student should be able to"* followed by a list containing ILOs. The ILO list structure varies and some examples are:
   - An HTML list with an `li` tag for each ILO (e.g. DD2448), possibly with sublists as well (e.g. DD1350).
   - A list with a `p` tag for each ILO beginning with a bullet character such as -, * or · (e.g. DD2476).
   - A list consisting of a single `p` tag in which the ILOs are separated using a `br` tag (e.g. AG1314).
   - A mix of the aforementioned structures (e.g. SK2753).
2. Same as in Point 1 but without the initial paragraph (e.g. EI1120).
3. Same as in Point 1 but also contains a paragraph explaining the general objectives of the course (e.g. SF1630).
4. Same as in Point 1 or Point 3 but after the ILO list there is a paragraph similar to *"in order to"* followed by another list explaining the purposes of fulfilling the ILOs (e.g. DD2300). The second list usually has the same structure as the first list.
5. Same as in Point 1 but after the ILO list there is a paragraph similar to *"For higher grades, the student should also be able to"* followed by another list containing ILOs (e.g. SF1626).

Since the structures are quite different from one another, we would need to process each structure using different methods which would be cumbersome. To avoid this, we convert the ILOs into a new structure to ensure that all syllabuses use the same structure which simplifies further processing.

The new structure consists of a long text string: All paragraphs are written in plain text (no XML tags) and occupy one line each, and all ILO paragraphs begin with a special character to separate them from regular paragraphs. This structure was chosen to facilitate the usage of Granska for performing natural language processing since the text to be processed is submitted as a text string (see Section 3.3.3).

3.3.3 Natural language processing

When analyzing syllabuses we are mainly interested in the occurrence of a word and not the form which the word is written in. For example, when classifying the ILOs using Bloom's taxonomy we want all forms of the word "combine" such as "combines" and "combining" to be mapped to the class Synthesis. We do this by lemmatizing all words before analyzing the syllabus to create an equivalence class for each word with the lemma being a representative for the class.

We also want to identify the active verbs in an ILO, which is done by part-of-speech tagging the ILO and selecting the words tagged as active verbs.

We choose to perform the lemmatization and part-of-speech tagging by using Granska, mainly due to its high accuracy rate for the part-of-speech tagging.

In addition, we choose to use Granska's suggestions for correcting spelling errors to possibly improve our results. In the case of multiple suggestions, we choose to use the first one.

Using Granska

Granska can be downloaded and run as a program or it can be run online on a server using the WebbGranska web interface. We choose to use WebbGranska since it is simple to use and does not require any installation.

WebbGranska uses GET requests for processing the text entered in the interface which means that the text also appears in the URL. For example, when entering the text *"jag åt kakor"* and selecting XML output the resulting URL is `http://skrutten.nada.kth.se/scrut.php?ruleset=svesve&text=jag+%E5t+kakor&url=&xmlout=on&x=Granska`. Note that the entered text is contained in the parameter `text` and that it is encoded, and according to the server description [35] the used encoding is ISO-8859-1. This means that we can lemmatize and part-of-speech tag a string by encoding it in ISO-8859-1, entering it as the `text` parameter, send a GET request using the URL and parse the resulting XML document using jsoup.

Before we construct the URL and send it to WebbGranska, we need to deal with a potential problem regarding the ambiguity of certain words causing them to be incorrectly tagged. To illustrate this problem, consider the following ILO from the syllabus of the course SF1626 (only one ILO is shown in the bullet list): [36]

> *"För högre betyg skall studenten dessutom kunna [...]*
>
> - *Lösa problem som kräver mer omfattande beräkningar i flera steg."*[1]

The ILO is divided into two parts, the first being a paragraph which applies to all ILOs in the bullet list and the second being the main content of the ILO. If considering both parts as being one sentence, it is clear that "lösa problem" means "to solve problems" which implies that "lösa" is a verb. However, by only considering the second part in isolation, "lösa problem" could mean "loose problems" which implies that "lösa" is an adjective. Indeed, Granska tags "lösa" as an adjective if only given the second part.

In general, this problem can occur if the ILOs are presented in a list and if each list item is considered to be a continuation of a paragraph preceding the list, such as in the example above.

One solution is to prepend the whole paragraph preceding the list to each ILO. This is however not particularly efficient since it introduces overhead on both the client side and the server side.

A similar but more efficient approach is to prepend the word "att" (meaning "to") to each ILO. This forces words that could be either verbs or other word classes to be interpreted as verbs which solves the problem. If "att" is prepended to a word that cannot be a verb there will be no difference since it will not be interpreted as a verb.

[1] English translation: *"For higher grades the student should also be able to [...] Solve problems that require more extensive computations in several steps."*

3.4 Classifying the syllabuses

This section motivates and describes the methods used to classify the syllabuses regarding language identification, Bloom's taxonomy, ESG, constructive alignment, number of ILOs, ethical approach, creating a score for each syllabus and grouping number of credits into ranges.

3.4.1 Language identification

To check that the Swedish syllabus is actually written in Swedish and not English, we use the method devised by Cavnar and Trenkle described in Section 2.4.3. The motivation for choosing this method is that it is quite straightforward to implement and that Cavnar and Trenkle found it to have near-perfect accuracy in identifying the language. The accuracy should in our case be even better since Cavnar and Trenkle used nine candidate languages for their tests while we only have Swedish and English, which also happen to have quite different n-gram frequencies. Also, simpler methods such as searching for words or letters that only occur in one language might not identify the language properly since there are many Swedish syllabuses which contain English words or phrases.

We will use the top 300 n-grams both for the language profiles and for the text profiles since Cavnar and Trenkle mention that the top 300 n-grams have a high correlation to the language [26]. Also, we choose to use trigrams since the trigram distributions in Swedish and English are very different which means that differentiating the two languages using trigrams should definitely be enough.

The process

The language identification process consists of the following four steps:

Step 1: Create the language profiles for Swedish and English. This is done by using large corpora containing Swedish and English texts and counting the occurrence of each trigram in all texts. Since this is a rather common approach, such trigram statistics have already been created and made available by other authors. We choose to use the statistics from [37, 38] which are compiled from the Leipzig Corpora Collection found at [39]. The reason for choosing these statistics is that they are based on big corpora with around 90 million characters in

Swedish and around 4.5 billion characters in English, and that they contain all trigrams whereas some other statistics only display the top 30–50 trigrams.

Step 2: Choose what parts of the syllabus to use for language identification. We decide on using four parts of the syllabus for language identification: objectives, content, eligibility and disposition. The reason for this choice is that other parts of the syllabus can make the identification less accurate since they might contain words from other languages (e.g. English literature in a Swedish syllabus).

Step 3: Create a text profile for each syllabus. We tokenize the four parts from Step 2 of the parsed and preprocessed syllabus into trigrams by reading each word, extracting one trigram (three letters) at a time and keeping track of how many times we observe each trigram. Finally, we create the text profile by ranking all trigrams by their frequency in the text with the most frequent trigram at the top.

Step 4: Compare the ranks in the language profiles with the ranks in the text profiles. For each trigram in the language profile, we search for the same trigram in the text profile and calculate the difference between the trigram's rank in the language profile and in the text profile. If the trigram is not located in the top 300 ranks of the text profile we assign the distance 301 to the trigram. Finally, we sum all the distances to obtain the total distance, and we identify the syllabus as being written in the language whose language profile has the shortest total distance to the text profile.

3.4.2 Choosing the model for classifying ILOs

As mentioned in Section 2.3.2 there are different models for classifying ILOs. However, since we need to perform the classification automatically it is necessary to have access to lists mapping verbs to classes, preferably several lists from different authors to ensure that our results are not biased. Not all relevant verbs might be contained in the lists, and it is therefore also necessary that the model has clearly defined classes to enable manual classification of specific verbs.

After looking at different models, it was found that most models do not have enough mapped verbs or have classes which are not suitable for automatic classification. However, a model that was found to fulfill the requirements is Bloom's taxonomy, and we therefore choose to use it for the classification.

3.4.3 Using Bloom's taxonomy

Since Bloom's taxonomy is a general model for classifying ILOs we first need to decide on how we want to use it to classify syllabuses. If we want to use it in our definition of pedagogically useful, we need to construct a measurable score and decide on which values are acceptable in the definition.

As explained in Section 2.3.3 the lower classes represent more simple behavior while higher classes represent more complex behavior. Since this thesis is about syllabuses in higher education, we believe that complex behavior is more desirable than simple behavior since higher education is supposed to build on the simpler knowledge acquired in secondary education. We therefore think it is appropriate to assign higher classes with higher scores and lower classes with lower scores, and the most natural way is to assign the scores 1 through 6 for each of the six classes and 0 if there is no class.

Thus, we choose to assign each ILO a score according to Table 3.1, after which we can calculate the *average Bloom's score* by simply summing the scores for all ILOs and dividing by the number of ILOs. An ILO might contain verbs from several classes in which case the highest class is chosen as the ILO score.

Table 3.1: The scores given to ILOs for each class in Bloom's taxonomy.

Highest class	**Score**
(no class)	0
Knowledge	1
Comprehension	2
Application	3
Analysis	4
Synthesis	5
Evaluation	6

We also need to decide on which values of the average Bloom's score are acceptable in the definition. As already discussed we believe that more complex behavior is desirable, but we cannot demand that all ILOs belong to the highest class since a course might begin with simple behavior and then continue on to complex behavior.

Since KTH is a technical and engineering university [40] and since applying existing knowledge to new situations is a central part of engineering, we believe that it is reasonable that syllabuses reach at least the Application class. Furthermore, the higher classes (Analysis, Synthesis and Evaluation) are also highly relevant for solving engineering problems. Considering this we believe that **an average Bloom's score of 3 or higher is desirable** and we therefore add such a check to the research part of our definition in Section 4.2.

Mapping between verbs and classes

To calculate the average Bloom's score we classify each ILO using lists by other authors that map Swedish verbs to their respective class in Bloom's taxonomy. The used lists are found in [41, 42, 43] and we combine them to form a new list containing a total of 133 unique verbs.

Since we want to perform the classification automatically, the mapping between verbs and classes must be unambiguous. However, some of the found verbs are mapped to multiple classes either by the same author or when combining lists by different authors which introduces ambiguity. We resolve this ambiguity by using the lowest class in which the verb occurs to ensure that we do not overestimate the class.

Expanding the lists

While trying to classify ILOs using the created list of verbs, it became clear that many relevant verbs in the ILOs are not covered by the created list. For example, the verb "implementera" (meaning "to implement") which is highly relevant in programming courses is not covered by the list. This might reduce the usefulness of the analysis since important verbs might be completely overlooked. To remedy this problem, the list of active verbs is expanded to map as many relevant verbs as possible, using two different methods.

The first method is to consult other lists mapping English verbs to classes in Bloom's taxonomy and use them by translating verbs between Swedish and English. The used lists are found in [44, 45, 46, 47, 48, 49].

The second method is to manually look at the ILOs and the contexts in which the verbs occur and trying to manually classify them using the definition of Bloom's taxonomy (see Section 2.3.3). It might be necessary to combine the two methods, especially when there does not exist a one-to-one mapping between a word in Swedish and English. Some of the English lists are accompanied by recommendations of verbs to avoid, which was useful for the manual classification.

Using these two methods, 99 new verbs were added. The final list contains 232 verbs and can be found in Appendix A.

3.4.4 ESG and constructive alignment

Measuring to which degree a syllabus complies with the ESG and constructive alignment cannot be done directly since they are not specified in measurable terms, and we therefore need to define a way to measure them.

The ESG specifies that students should be encouraged to *"take an active role in creating the learning process"* and that the assessment should allow students to *"demonstrate the extent to which the intended learning outcomes have been achieved."* This implies that the syllabus should contain ILOs using active verbs which describe the activities the student should participate in and be assessed on, which is exactly what constructive alignment is about. We can measure this by utilizing the mapping between verbs and classes in Bloom's taxonomy.

Thus, to comply with the ESG and constructive alignment, we say that **each ILO in a syllabus must contain at least one verb belonging to any class of Bloom's taxonomy**. Since we do not want a syllabus having no ILOs to be classified as being compliant to the ESG, we also check that **there is at least one ILO**.

3.4.5 Number of ILOs

As specified in Section 2.3.4 a syllabus should have around 6 ILOs. However, to make this notion measurable we need to specify an acceptable interval for the number of ILOs.

Biggs, Kennedy and the MIT Teaching and Learning Laboratory all mention that 6 ILOs is an acceptable value. Following the conclusions of Kennedy we also consider 7, 8 and 9 ILOs to be acceptable. MIT Teaching and Learning Laboratory states that at most 6 ILOs should be used which is not fully compatible with the conclusions of Kennedy, but we combine the two by also considering 3, 4 and 5 ILOs to be acceptable.

In total, we define the acceptable interval to be 3 through 9 ILOs.

3.4.6 Ethical approach

The KTH recommendations mention that a syllabus should include information about “ethical approach” since it is important for students to be aware of that education at KTH should be conducted without plagiarism, cheating and other irregularities [9]. We choose to measure this by looking in the whole syllabus for words containing “heder” (honor), “kodex” (codex), “fusk” (cheats) and “plagi” (covers words regarding plagiarism).

3.4.7 Aggregating the variables into a score

To find how well the syllabuses fulfill the definition, it is not enough to study the measured variables individually since the definition consists of multiple variables. Thus, we need to aggregate the variables and create a score for the whole syllabus.

To create the score, we first need to choose which weights the variables should have. The weights can either be uniform to make all variables equally important or non-uniform to make some variables more important than others. Using non-uniform weights is a good option when there is a clear ordering of importance of the variables, but this is not the case for our variables. For example, is it more important that a syllabus specifies a grading scale or the language of instruction? Such weights would be assigned rather arbitrary and different persons would probably assign the variables different weights. Therefore, we think it is more appropriate to use uniform weights for all variables.

If we have uniform weights, a natural way to create the score is to use an arithmetic mean. To do this, we consider each variable to be a binary variable assuming the value 1 or 0 depending on if the variable is fulfilled or not, which enables us to sum the number of fulfilled variables and divide by the number of variables to calculate the score.

More precisely, we define that

$$\text{score} = \frac{1}{n}\sum_{i=1}^{n} V_i$$

where n is the number of variables and

$$V_i = \begin{cases} 1 & \text{if variable } i \text{ is fulfilled} \\ 0 & \text{otherwise.} \end{cases}$$

This definition allows us to view the score as the percentage of fulfilled variables. For example, if the calculated score is 0.85 we know that 85 % of the variables are fulfilled. Since this is quite intuitive, we choose to illustrate the calculated scores as percentages in the results in Section 4.3.

In addition to the score for the definition, we also calculate similar scores for each specification from Chapter 2 which enables us to analyze them separately.

3.4.8 Number of credits ranges

While creating the results in Section 4.3.1 it was discovered that the amount of courses having each number of credits varies greatly (see Appendix B for detailed results). The courses having 6.0, 7.5, 9.0, 15.0 and 30.0 credits together constitute 85.5 % of all courses, while some number of credits are only used by a few (or no) courses which would lead to a single course making a big impact on the results. To reduce this problem, we choose to group the number of credits into ranges where 6.0, 7.5, 9.0, 15.0 and 30.0 credits are separated into different ranges. The ranges are specified in Table 3.2 along with the amount of courses contained in each range.

Table 3.2: The number of courses belonging to each number of credits range.

Credits range	#courses
1–5	150
6–7	457
7.5–8.5	851
9–14	197
15–22.5	141
30–45	179

Chapter 4

Results

The results presented in this section consist of the created variables, the definition itself (both the main definition and definitions for each specification) and the measured values needed for the analysis.

4.1 Variables

We created the following 21 measurable variables (regarding the Swedish syllabus unless otherwise stated):

V1: Which cycle the course belongs to

V2: Number of credits

V3: Objectives (regardless if they are stated as ILOs or not)

V4: Course name

V5: Course code

V6: Department giving the course

V7: When the syllabus is valid from

V8: Main course content

V9: Course literature

V10: Language of instruction

V11: Grading scale

V12: Examination modules in Ladok

V13: Comments on the examination forms

V14: Requirements for obtaining a final grade

V15: Words regarding ethical approach (see Section 3.4.6)

V16: Language of the syllabus

V17: Specific eligibility requirements

V18: Main fields of study

V19: Number of ILOs

V20: Number of ILOs belonging to any class of Bloom's taxonomy

V21: Average Bloom's score (see Section 3.4.3)

4.2 Definition of pedagogically useful

This section contains the main definition of pedagogically useful along with the definitions for each of the specifications in Chapter 2.

4.2.1 Main definition

We say that a syllabus is pedagogically useful if it fulfills the laws, regulations, recommendations and research specified in Chapter 2 in combination with the considerations in Section 3.4. This is equivalent to that our variables in Section 4.1 have the following values (in the Swedish syllabus unless otherwise stated):

- V1–V15 are non-empty
- V4 in the English syllabus is non-empty
- V16 is "Swedish"
- V16 in the English syllabus is "English"
- V19 is between 3 and 9 (inclusive)
- V20 equals V19
- V21 is at least 3

4.2.2 Definitions for each specification

The main definition is constructed with all specifications taken in mind, but we also want to study each specification separately. To fulfill each specification, the variables should have the following values (in the Swedish syllabus unless otherwise stated):

Swedish laws: V1–V3 and V12 are non-empty; V16 is "Swedish"

KTH regulations: V1–V12 and V14 are non-empty; V4 in the English syllabus is non-empty; V16 is "Swedish"; V16 in the English syllabus is "English"

KTH recommendations: V1–V5, V7–V8, V11–V15 are non-empty; V16 is "Swedish"; V16 in the English syllabus is "English"; V19 is greater than 0; V20 equals V19

SUHF recommendations: V1–V4, V7–V9, V11–V12 and V14 are non-empty; V4 in the English syllabus is non-empty; V16 is "Swedish"; V19 is greater than 0

ESG: V19 is greater than 0; V20 equals V19

Research: V19 is between 3 and 9 (inclusive); V20 equals V19; V21 is at least 3

4.3 Measured values

The measured values below are based on the syllabuses of all 1975 courses with at least one course offering during the study year 2017/2018 as of May 27, 2017.

Note that most of the figures below have y-axes that start above 0, with the purpose of increasing clarity.

4.3.1 Fulfilled variables

Table 4.1 shows the average percentage of variables fulfilled in the main definition and in each specification separately. Laws, SUHF recommendations and KTH regulations are the most fulfilled specifications and KTH recommendations are slightly less fulfilled, whereas the values for the research and ESG specifications are much lower. In total, 78.4 % of the main definition variables are on average fulfilled in a syllabus.

Table 4.1: The average percentage of specification variables fulfilled.

Specification	Fulfilled variables
Main definition	78.4 %
Laws	97.5 %
SUHF	92.7 %
KTH reg.	90.9 %
KTH rec.	82.8 %
ESG	63.2 %
Research	47.9 %

Fulfilled variables for each school

Figure 4.2 shows the average percentage of variables fulfilled in the main definition for each school. The maximum difference is 6.4 percentage points between CSC at 81.9 % and BIO at 75.5 %.

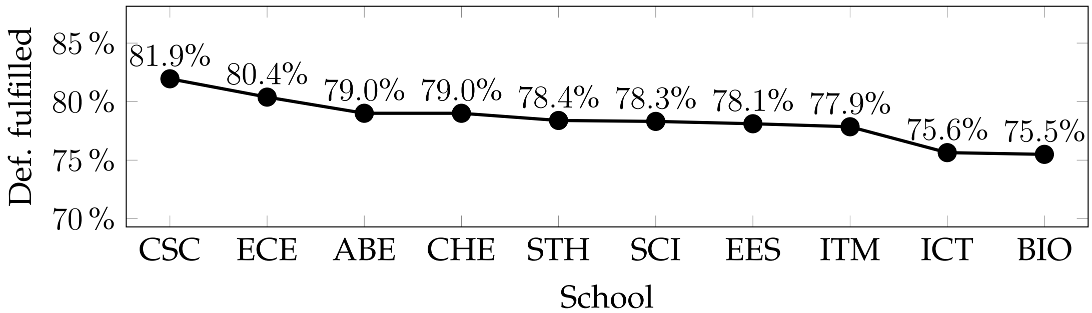

Figure 4.2: The average percentage of main definition variables fulfilled for each school.

Figure 4.3 shows the average percentage of variables fulfilled in each specification for each school (sorted as in Figure 4.2). The maximum differences between the schools are around 3–8 percentage points for laws, SUHF recommendations, KTH regulations and KTH recommendations, and around 15–21 percentage points for the ESG and research specifications.

Fulfilled variables for each cycle

Figure 4.4 shows the average percentage of variables fulfilled in the main definition for each cycle. There is a slight decrease in how well the

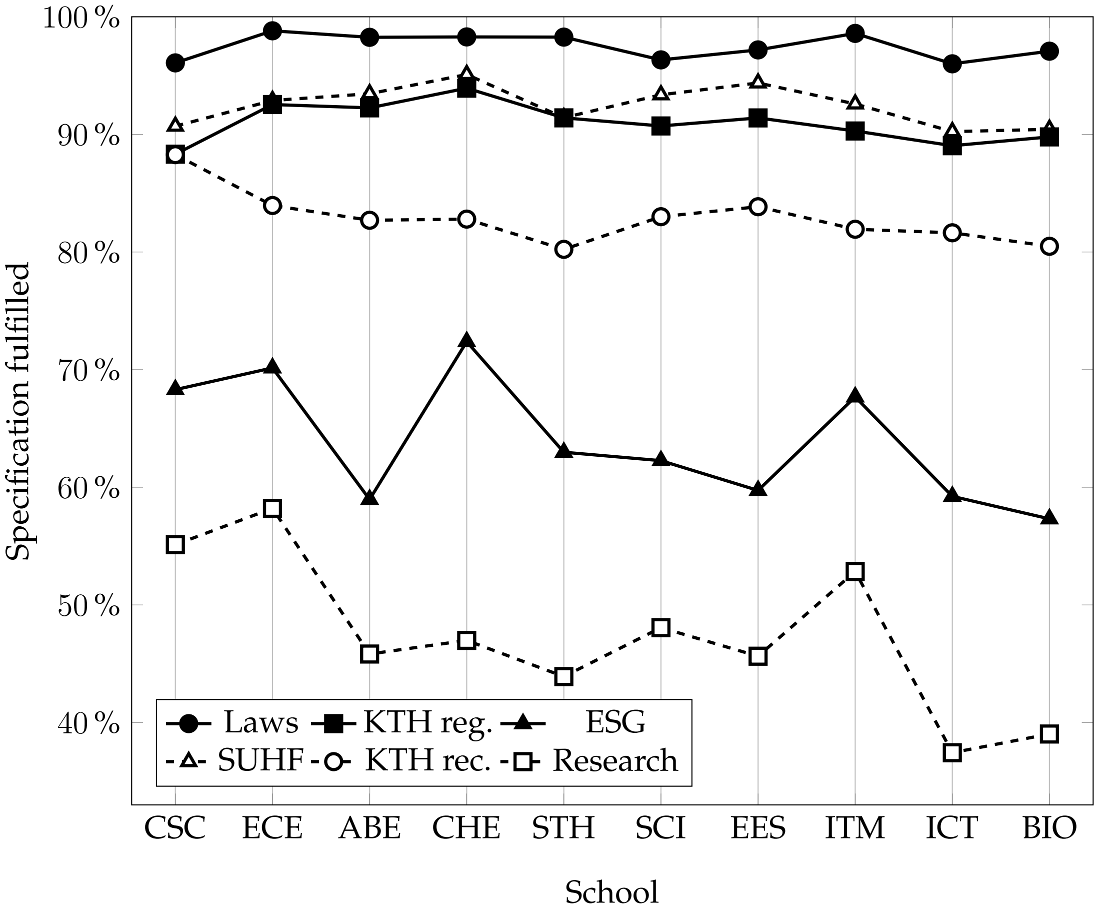

Figure 4.3: The average percentage of specification variables fulfilled for each school.

main definition is fulfilled as the cycle increases with the maximum difference being 5 percentage points between the preparatory and second cycles.

Figure 4.5 shows the average percentage of variables fulfilled in each specification for each cycle. There is a decreasing trend for laws, SUHF recommendations, KTH regulations and KTH recommendations as the cycle increases with maximum differences being around 1–8 percentage points. For the ESG and research specifications the first cycle shows higher values than the other two cycles. Note that the second cycle has the lowest values in all specifications.

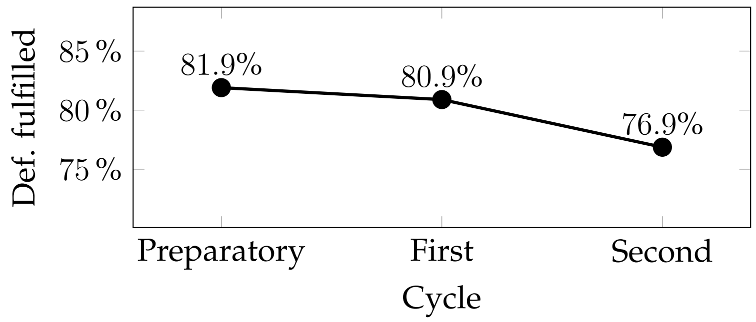

Figure 4.4: The average percentage of main definition variables fulfilled for each cycle.

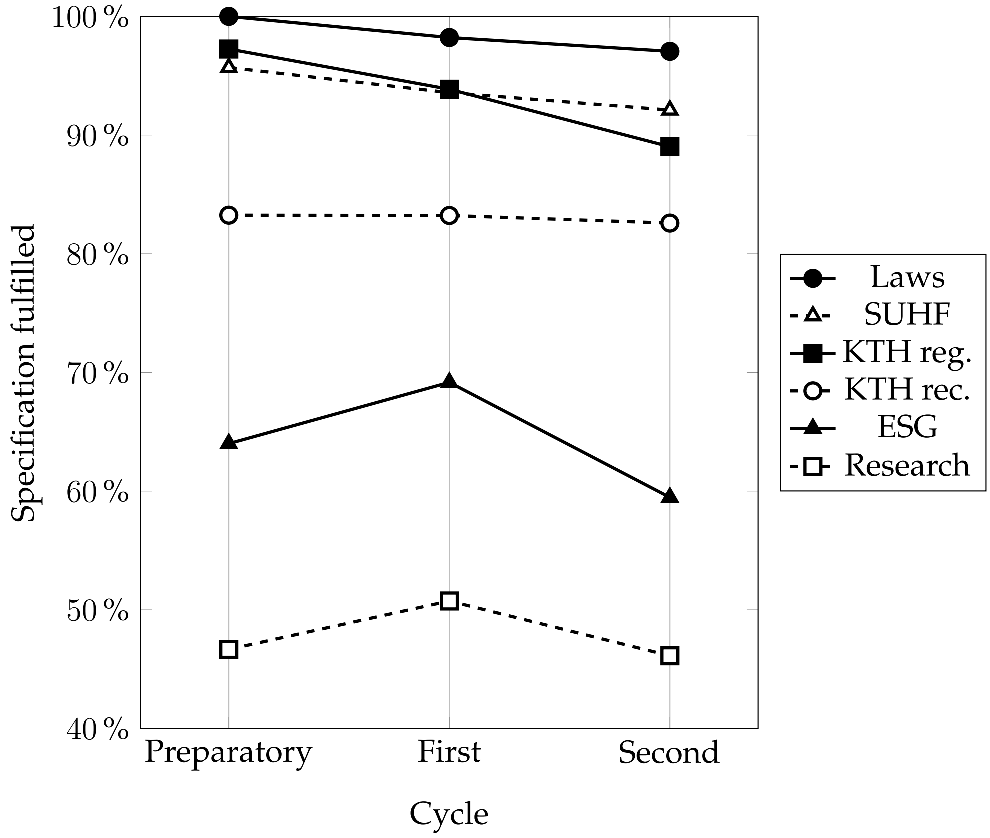

Figure 4.5: The average percentage of specification variables fulfilled for each cycle.

Fulfilled variables for each number of credits

In the following results, the number of credits are divided into ranges as described in Section 3.4.8.

Figure 4.6 shows the average percentage of variables fulfilled in the definition for each number of credits range. All ranges display similar values with the maximum difference being 1.9 percentage points.

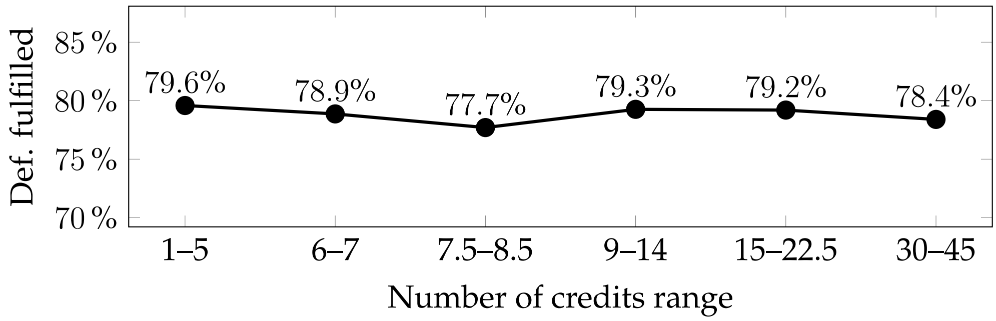

Figure 4.6: The average percentage of main definition variables fulfilled for each number of credits range.

Figure 4.7 shows the average percentage of variables fulfilled in each specification for each number of credits range. The maximum differences for laws, SUHF recommendations, KTH regulations and KTH recommendations are around 2–6 percentage points. The ESG and research specifications have maximum variations of around 12 percentage points with generally decreasing values as the number of credits increases.

4.3.2 Average Bloom's score

The overall average Bloom's score for all syllabuses is 2.426, which is between Comprehension and Application but slightly closer to Comprehension.

Average Bloom's score for each school

Figure 4.8 shows the average Bloom's score for each school. Six schools have values between 2.23–2.46 while ITM, CSC and ECE are noticeably higher at 2.69–2.76. BIO has a much lower score than all other schools at 1.83 which is 0.4 lower than the second lowest (ICT).

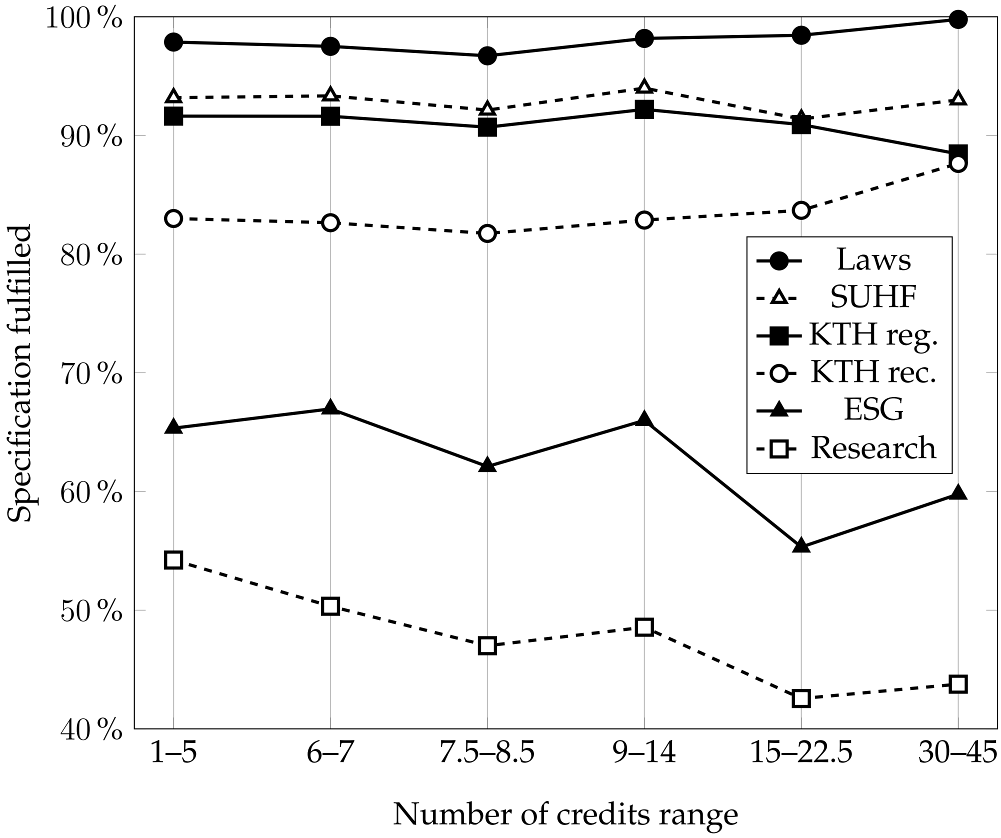

Figure 4.7: The average percentage of specification variables fulfilled for each number of credits range.

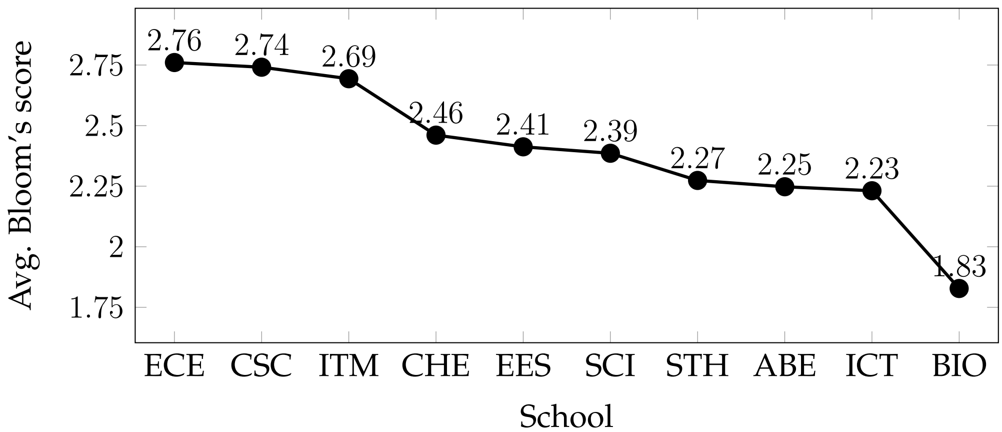

Figure 4.8: The average Bloom's score for each school.

Average Bloom's score for each cycle

Figure 4.9 shows the average Bloom's score for each cycle. The maximum difference is 0.31 between the first and second cycles.

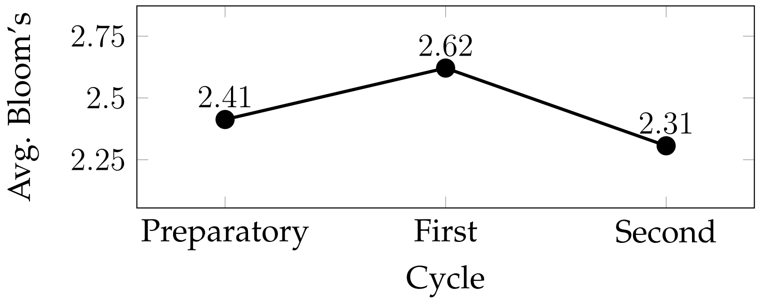

Figure 4.9: The average Bloom's score for each cycle.

Average Bloom's score for each number of credits range

Figure 4.10 shows the average Bloom's score for each number of credits range. 1–5, 9–14 and 15–22.5 credits have similar values at around 2.6 while 6–7 and 7.5–8.5 credits are noticeably lower at 2.43 and 2.21 respectively and 30–45 credits is much higher at 3.01.

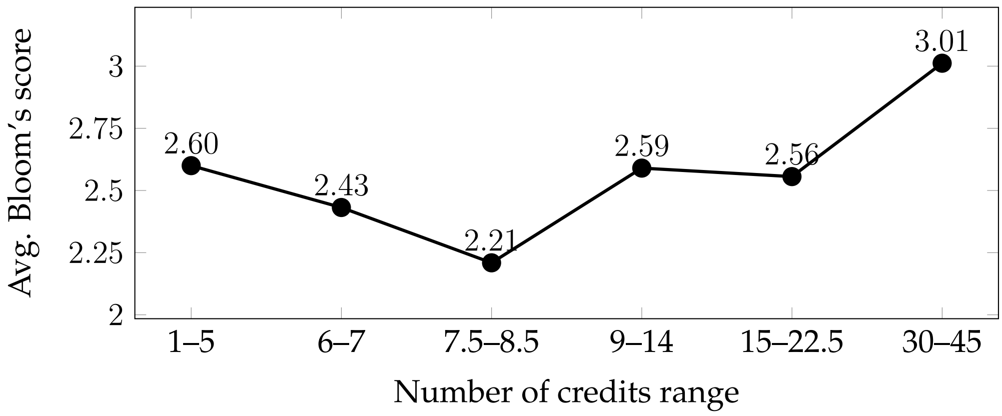

Figure 4.10: The average Bloom's score for each number of credits range.

4.3.3 Number of ILOs

Overall, the average number of ILOs in a syllabus is 6.35.

Number of ILOs for each school

Figure 4.11 shows the average number of ILOs for each school. The maximum difference is 3.48 between CHE at 8.86 and ABE at 5.38. All schools have averages between 3 and 9 ILOs (inclusive) which is accepted in the main definition, but CHE is very close to the upper boundary.

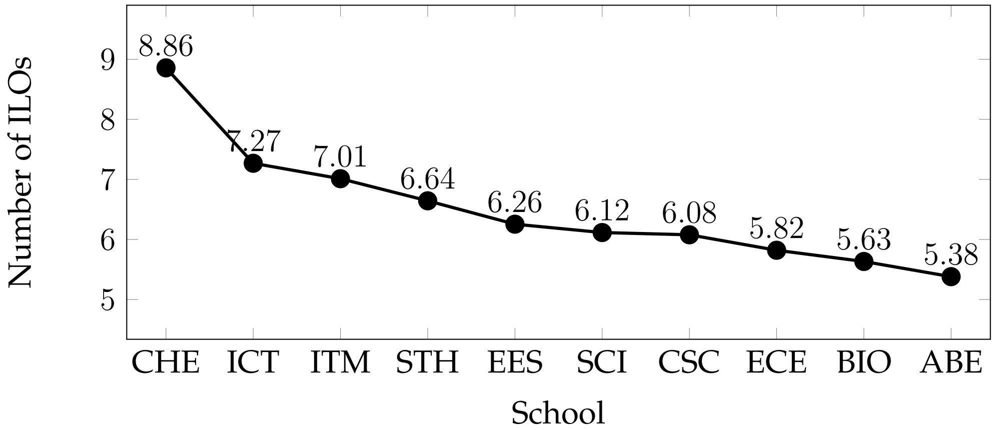

Figure 4.11: The average number of ILOs for each school.

Number of ILOs for each cycle

Figure 4.12 shows the average number of ILOs for each cycle. The first cycle clearly has the highest value at 7.61, being 2.01 higher than the second cycle and 2.49 higher than the preparatory cycle.

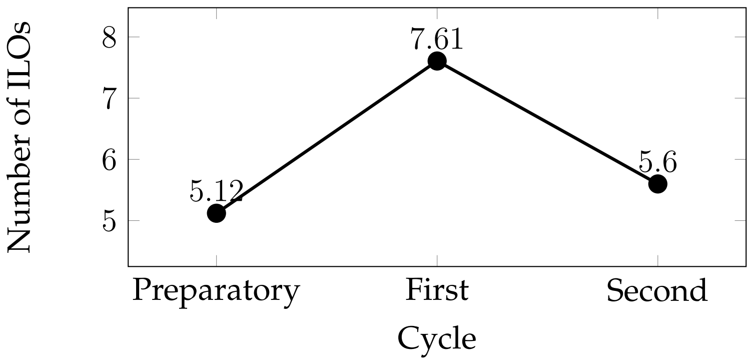

Figure 4.12: The average number of ILOs for each cycle.

Number of ILOs for each number of credits range

Figure 4.13 shows the average number of ILOs for each number of credits range. There is an increasing trend for the number of ILOs as the number of credits increases, since 1–5 credits have 4.67 ILOs, 6–8.5 credits have 5.97–6.43 ILOs and 9 or more credits have 7.11–7.65 ILOs.

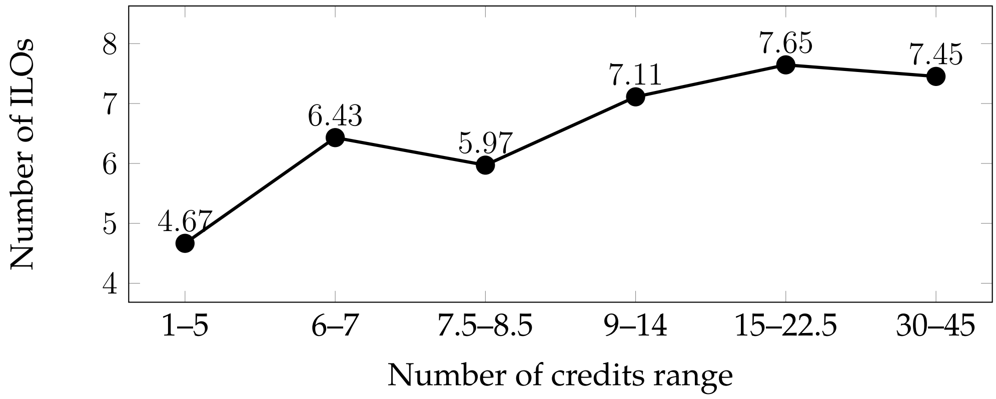

Figure 4.13: The average number of ILOs for each number of credits range.

4.3.4 Further measurements

Appendix C contains further measurements which are of interest to the project provider but not part of any definition and not needed to answer the problem statements of this thesis.

Chapter 5

Discussion

In this chapter, the results from Chapter 4 will be interpreted and analyzed, the used method will be criticized, possibilities for further work will be discussed and the questions in the problem statement will be answered.

5.1 Fulfilled variables

On average, 78.4 % of the main definition variables are fulfilled in a syllabus. We consider this to be a rather good value, but there is nevertheless room for improvement. In general, the syllabuses are good at providing the needed information since the four most fulfilled specifications mainly regard the existence of information, while they are less good at having well-written ILOs since the ESG and research specifications are much less fulfilled. The biggest improvements could therefore be made by ensuring that all ILOs are constructively aligned by using active verbs, preferably from the higher classes of Bloom's taxonomy. To facilitate this, it might be necessary to inform syllabus authors of how to use constructive alignment and Bloom's taxonomy for syllabus design since they might not be familiar with those methods.

5.1.1 School

There is a slight correlation between the school and how well the main definition is fulfilled, but it is not especially strong since the difference is only a few percentage points.

Stronger correlations can be found between schools and fulfilling the ESG and research specifications since the variations are quite large. ECE has the highest value for the research specification being around 21 percentage points higher than the lowest scoring school (ICT). Since ECE is conducting research in the field of educational development, one theory for the high value is that they have more knowledge on syllabus design which reflects on their syllabuses.

5.1.2 Cycle

In general, both the main definition and the specifications become slightly less fulfilled as the cycle increases. This is quite interesting and also unexpected since one of the hypotheses was that they would become more fulfilled as the cycle increases.

Note that the values for the preparatory cycle might not be fully comparable to the other cycles since there are only 25 courses in the preparatory cycle, compared to 741 courses in the first cycle and 1209 courses in the second cycle. However, this is not a big problem for the overall results since the observed negative correlation exists even if the preparatory cycle is completely ignored.

5.1.3 Number of credits

The number of credits does not have a notable influence on how well the main definition is fulfilled, but there is a decreasing trend for the research and ESG specifications. This trend is mainly due to the requirement that all ILOs should be constructively aligned, which makes it harder for syllabuses having more ILOs to fulfill the specification. We could have been more lenient and required that, say, at least 80 % of the ILOs are constructively aligned instead of 100 %, but we do not believe that this is a good idea since we want to ensure that all ILOs are examinable.

5.2 Average Bloom's score

The overall average Bloom's score is quite low at 2.426 which is above the Comprehension class. An average of at least 3 was expected, but the lower average is probably due to many syllabuses not being written with constructive alignment in mind and thus scoring very low which brings down the average.

The average Bloom's score varies quite much between schools with the maximum difference being almost a whole class.

The second cycle surprisingly has the lowest average Bloom's score. One of the hypotheses was that the average Bloom's score would increase as the cycle increases since the courses should require more advanced educational behavior, but this is not reflected in the syllabus. This could mean that second cycle courses do not require more advanced educational behavior than first cycle courses, in which case the results should be seen as a warning sign of the second cycle courses being too simple. However, another reason could be that constructive alignment is not applied to the same extent when writing second cycle syllabuses as when writing first cycle syllabuses, perhaps because first cycle courses have more students and it is therefore considered more important to write the syllabuses properly.

30–45 credits has by far the highest average Bloom's score, probably due to most courses in that credit range being degree project courses which have a bigger focus on skills located in the three highest classes of Bloom's taxonomy. Interestingly, 1–5 credits have a higher score than 6–7 and 7.5–8.5 credits.

5.3 Number of ILOs

The number of ILOs varies greatly between the schools so there is definitely a correlation between the school and the number of ILOs. Notably, CHE has a much higher value than the other schools.

There is also a rather big difference for the cycles as the first cycle is clearly higher than the other two cycles. One theory is that first cycle courses are introductory courses and thus cover many different topics shallowly needing more ILOs, whereas the second cycle courses are more specialized and focus on fewer topics in more depth needing fewer ILOs.

Furthermore, there is a general positive correlation between the number of credits and the number of ILOs even if the values are not monotonically increasing.

5.4 Laws

A remarkable result is that Swedish laws are not fulfilled by all syllabuses. Even though 97.5 % is a high value it is not sufficient since laws are mandatory to follow and the value should therefore be nothing less than 100 %.

The main reason for not fulfilling the laws is that the Swedish syllabus is in fact written in English in 110 syllabuses. This should be fixed by translating the English texts to Swedish in the current syllabuses. It should also be prevented from happening in the future by ensuring that the language is correct when approving the syllabus, perhaps with the aid of a program which gives a warning if the wrong language is used.

5.5 SUHF

One part of the SUHF recommendations is that courses should be classified according to the levels of specialization in SUHF recommendation dnr 08/025. These levels are currently not used at KTH which means that no syllabuses fulfill the SUHF recommendations, which is why the levels had to be ignored in this thesis to get useful results.

It is quite strange that KTH does not use the suggested levels despite being members of SUHF and despite that many other Swedish universities and university colleges use them. The purpose of the levels is to facilitate mobility between universities but this cannot be achieved unless all universities actually use the levels.

5.6 Method criticism

In this section, we criticize the used method by discussing its possible weak points.

5.6.1 Adding own verbs

As mentioned in Section 3.4.3 there are many relevant verbs that occur in the ILOs but do not occur in any existing list mapping verbs to classes in Bloom's taxonomy, which is why we had to expand the existing lists.

One potential issue is that it is not always clear-cut which class in Bloom's taxonomy a certain verb belongs to. This introduces ambiguity when manually classifying the verbs which could lead to the same verb being classified differently by different persons, and this was indeed observed in the existing lists where some verbs are classified as belonging to several classes, sometimes even by the same author.

Despite this, we believe that the ambiguity does not pose a serious threat for the accuracy of the results, especially since we choose the lowest class in case of multiple possible classes. We might in some cases be too restrictive and underestimate the class, but we believe that it is better than to be generous and overestimate the class.

The ambiguity is an inherent part of the mapping between verbs and classes of Bloom's taxonomy and cannot be fully avoided. This is mainly because a verb can have different meanings depending on the context in which it occurs, and this meaning is hard to capture when considering the verb in isolation. One improvement for the method could therefore be to take the context of the verbs into account, but this would probably be extremely difficult since it would be necessary to map the combination of the verb and its context to a class of Bloom's taxonomy.

5.6.2 Identifying the ILO structures

As explained in Section 3.3.2 there is a multitude of ILO structures in the syllabuses. To ensure that each structure is correctly parsed, it is necessary to identify the structure by finding a syllabus in which the structure occurs and then implementing a method for parsing it. The identification of structures was made using manual checks and regular expressions to cover as many structures as possible. Nevertheless, it is possible that some structures were incorrectly parsed since it is infeasible to verify the parsing of each structure manually, but we believe that this problem is relatively small since most structures are somewhat similar to each other and parsing one structure should therefore also enable parsing of other structures.

Apart from having made the work of this thesis easier, we believe that having a standardized ILO structure would be beneficial for all involved parties since it would improve readability, simplify the design of constructively aligned ILOs and simplify quality assurance. A standardized structure could be to start with the paragraph *"After a* *completed course, the student should be able to"* followed by an HTML `ul` list with an `li` element for each ILO. To enforce this structure, the initial paragraph should not be possible to delete and there should be separate text boxes for each ILO (with the possibility to add more text boxes).

5.6.3 Subjectivity

The process of creating measurable variables, combining them into a definition and classifying the syllabuses contains many subjective aspects. Unfortunately, this subjectivity cannot be avoided since it is an inherent part of the problem. It is possible that the process could have been done differently and thus getting other results, but since we used existing specifications and tried to stay as close to them as possible we believe that the difference would not be large enough to affect the results drastically.

5.7 Further work

Finding explanations for some of the results is difficult since we are able to observe the values but we cannot know for sure what caused them. We have tried to give as qualified explanations as possible using the available information, but it might be interesting to further investigate the causes. Nevertheless, we believe that the results are more useful than the causes since the results make it possible to see where the weak spots are and perform actions to improve them.

After the results showing that laws are not fulfilled in all syllabuses, one might wonder why there are syllabuses not fulfilling the laws. This is indeed a justified question and we believe that it should be investigated further. One possible reason is that there is a lack of knowledge about the laws, either by the syllabus authors or by the ones approving the syllabuses, which could be improved by informing the involved parties on the matter.

As mentioned in Section 1.4 we assume that syllabuses correctly represent courses, but in practice there might be discrepancies between the course and its syllabus. It would therefore be valuable to see to what extent courses actually correspond to their syllabuses. However, this is probably not possible do to automatically which means that most of the work has to be done manually and it might therefore be infeasible to cover all courses.

Syllabuses at KTH usually have two versions, one legally binding Swedish version and one English version. They are usually constructed by writing the syllabus in one language and then translating it to the other language. The translation is done using an automatic translation system called Convertus Linguistic Translator [50] followed by a manual review with possible corrections before it is approved. However, this system has only been used for the last few years which means that older syllabuses are not affected. Furthermore, it is possible to manually bypass the system completely. An interesting question to investigate would therefore be to what extent the Swedish syllabuses correspond to the English syllabuses.

Finally, the method of this thesis could be adapted to analyze syllabuses in other higher education institutions in Sweden or other countries. The adaptation would be rather straightforward since the variables and definitions are modular and can easily be replaced. For Swedish higher education institutions the KTH regulations and recommendations should be replaced, and for higher education institutions in other countries the Swedish laws should also be replaced. The remaining specifications are relevant for all higher education institutions, and they could therefore be used to compare syllabuses between different higher education institutions in the same country or internationally, which could be very useful to promote further discussion regarding quality assurance.

5.8 Conclusions

We conclude this thesis by summarizing the results and discussions to answer the three questions from the problem statement in Section 1.2.

5.8.1 Defining "pedagogically useful"

We defined "pedagogically useful" by first looking at relevant specifications such as laws, regulations, recommendations and research, then constructing measurable variables which correspond to the specifications and finally combining the variables to create the definition. The definition is found in Section 4.2.

5.8.2 How pedagogically useful the syllabuses are

On average, 78.4 % of the main definition is fulfilled in a syllabus. There are big differences between how well each specification is fulfilled which implies that there is room for improvement, especially by increasing the usage of constructive alignment and the higher classes of Bloom's taxonomy. Remarkably, not all syllabuses comply with Swedish laws.

5.8.3 Correlations

There are correlations, albeit not especially strong, between how well the main definition is fulfilled and the school, cycle and number of credits. Stronger correlations are however found when considering each specification separately.

Bibliography

[1] merriam-webster.com. *Syllabus [Online]*. Available at: https://www.merriam-webster.com/dictionary/syllabus [Accessed 2017-03-02].

[2] SFS 1993:100. *Högskoleförordning*. Stockholm: Ministry of Education and Research.

[3] *Anvisning om inrättande av kurser på grundnivå och avancerad nivå [Online]*. Available at: https://intra.kth.se/polopoly_fs/1.661126!/inr%C3%A4ttande_kurs.pdf [Accessed 2017-03-30].

[4] *Kursplan [Online]*. Available at: https://intra.kth.se/polopoly_fs/1.299449!/Menu/general/column-content/attachment/UT-KU_sv.pdf [Accessed 2017-04-11].

[5] *Kommentarer till mall för kursplan [Online]*. Available at: https://intra.kth.se/polopoly_fs/1.299216!/Menu/general/column-content/attachment/KursplanInfo_sv.pdf [Accessed 2017-04-11].

[6] SFS 2009:600. *Språklag*. Stockholm: Ministry of Culture.

[7] Statistics Sweden. *Det allmänna myndighetsregistret - Statliga förvaltningsmyndigheter [Online]*. Available at: http://www.myndighetsregistret.scb.se/Myndighet.aspx [Accessed 2017-02-18].

[8] Högskoleverket. *Rättssäker examination*. 2008.

[9] Magnus Andersson, Veine Haglund, Viggo Kann, Björn Kjellgren, Ida Naimi-Akbar, and Lasse Wingård. *Förslag till struktur för kursinformation på KTH*, 2017. Unpublished report.

[10] *Om SUHF [Online]*. Available at: `http://www.suhf.se/om-oss` [Accessed 2017-02-20].

[11] *Medlemmar [Online]*. Available at: `http://www.suhf.se/om-oss/medlemmar` [Accessed 2017-05-21].

[12] SUHF. *Rekommendationer om kursplaner, utbildningsplaner och betygssystem (REK 2011:1)*, 2011.

[13] *Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG)*. Brussels, Belgium, 2015.

[14] Publications Office of the European Union. *The Bologna process: setting up the European higher education area*, 2015.

[15] Declan Kennedy. *Writing and Using Learning Outcomes: A Practical Guide*. 2006.

[16] John B. Biggs. *Teaching for Quality Learning at University*. Second edition, 2003.

[17] John B. Biggs and Kevin F. Collis. *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome)*. 1982.

[18] Robert J. Marzano and John S. Kendall. *The New Taxonomy of Educational Objectives*. Second edition, 2006.

[19] Benjamin S. Bloom, Max D. Engelhart, Edward J. Furst, Walker H. Hill, and David R Krathwohl. *Taxonomy of Educational Objectives: The Classification of Educational Goals*. 1956.

[20] Jian-Syuan Wong, Bart Pursel, Anna Divinsky, and Bernard J. Jansen. Analyzing MOOC discussion forum messages to identify cognitive learning information exchanges. *Proceedings of the Association for Information Science and Technology*, 52(1):1–10, 2015.

[21] MIT Teaching and Learning Laboratory. *Intended Learning Outcomes [Online]*. Available at: `http://tll.mit.edu/help/intended-learning-outcomes` [Accessed 2017-02-25].

[22] Daniel Jurafsky and James H. Martin. *Speech and Language Processing: An introduction to natural language processing*. Second edition, 2008.

[23] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to information retrieval*. Cambridge: Cambridge university press, 2008.

[24] Ted Dunning. *Statistical Identification of Language*. Computing Research Laboratory, New Mexico State University, 1994.

[25] Marc Damashek. Gauging Similarity with n-Grams: Language-Independent Categorization of Text. *Science*, 267:843–848, 1995.

[26] William B. Cavnar and John M. Trenkle. N-Gram-Based Text Categorization. In *Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*, pages 161–175, 1994.

[27] Johan Carlberger, Rickard Domeij, Viggo Kann, and Ola Knutsson. The development and performance of a grammar checker for Swedish: A language engineering perspective. *Natural Language Engineering*, 1(1), 2004.

[28] Johan Carlberger and Viggo Kann. Implementing an efficient part-of-speech tagger. *Software Practice and Experience*, 29(9):815–832, 1999.

[29] Jonas Sjöbergh. Combining POS-taggers for improved accuracy on Swedish text. *Proceedings of NoDaLiDa*, 2003.

[30] *WebbGranska 1.1 [Online]*. Available at: `http://skrutten.nada.kth.se/scrut.php?ruleset=svesve` [Accessed 2017-03-04].

[31] *KTH | Kurser inom program [Online]*. Available at: `http://www.kth.se/student/kurser/kurser-inom-program` [Accessed 2017-03-10].

[32] *KTH | API för kurs- och programinformation [Online]*. Available at: `http://www.kth.se/api/kopps/v1/` [Accessed 2017-03-10].

[33] *jsoup Java HTML Parser, with best of DOM, CSS, and jquery [Online]*. Available at: `https://jsoup.org/` [Accessed 2017-03-16].

[34] *Apache Commons Lang [Online]*. Available at: `https://commons.apache.org/proper/commons-lang/` [Accessed 2017-03-16].

[35] *Tools, Human Language Technology Group, KTH CSC [Online]*. Available at: http://www.csc.kth.se/tcs/humanlang/tools.html [Accessed 2017-03-19].

[36] *KTH | SF1626 Flervariabelanalys 7,5 hp [Online]*. Available at: https://www.kth.se/student/kurser/kurs/SF1626 [Accessed 2017-05-06].

[37] *Practical Cryptography – Swedish Letter Frequencies [Online]*. Available at: http://practicalcryptography.com/cryptanalysis/letter-frequencies-various-languages/swedish-letter-frequencies/ [Accessed 2017-03-19].

[38] *Practical Cryptography – English Letter Frequencies [Online]*. Available at: http://practicalcryptography.com/cryptanalysis/letter-frequencies-various-languages/english-letter-frequencies/ [Accessed 2017-03-19].

[39] *Leipzig Corpora Collection Download Page [Online]*. Available at: http://wortschatz.uni-leipzig.de/en/download/ [Accessed 2017-03-19].

[40] *KTH | An innovative European technical university [Online]*. Available at: https://www.kth.se/en/om/fakta [Accessed 2017-05-15].

[41] kursutveckling.se. *Aktiva verb [Online]*. Available at: http://www.kursutveckling.se/kursverkstaden/docs/Aktiva_verb_1.pdf [Accessed 2017-05-15].

[42] *Blooms taxonomi för lärandemål: Kognitiva mål [Online]*. Available at: http://www.ark.lu.se/uploads/media/Grund_foer_bedoemning_av_examinationer.pdf [Accessed 2017-05-15].

[43] Tova Stenlund. *Modellprövning: Empirisk prövning av teoretiska modeller för beskrivning av kursplan för förarutbildning*. Department of Educational Measurement, Umeå University, 2006.

[44] *Action Words for Bloom's Taxonomy [Online]*. Center for University Teaching, Learning, and Assessment, University of West Florida.

Available at: https://www.cte.cornell.edu/documents/Assessment%20-%20Blooms%20Taxonomy%20Action%20Verbs.pdf [Accessed 2017-05-16].

[45] *Bloom's Taxonomy Action Verbs [Online]*. Available at: http://www.fresnostate.edu/academics/oie/documents/assesments/Blooms%20Level.pdf [Accessed 2017-05-16].

[46] *Bloom's Taxonomy Verb List [Online]*. Available at: https://adp.uni.edu/documents/bloomverbscognitiveaffectivepsychomotor.pdf [Accessed 2017-05-16].

[47] *Bloom's Taxonomy Verbs [Online]*. Available at: https://prd-medweb-cdn.s3.amazonaws.com/documents/contacthours/files/blooms.pdf [Accessed 2017-05-16].

[48] *Bloom's Taxonomy of Measurable Verbs [Online]*. Available at: https://www.pdc.edu/wp-content/uploads/2012/01/Bloom-Taxonomy_2012.pdf [Accessed 2017-05-16].

[49] *blooms_taxonomy_verb_list.pdf [Online]*. Available at: http://www.schoolcraft.edu/pdfs/oca/blooms_taxonomy_verb_list.pdf [Accessed 2017-05-16].

[50] *Technology | Convertus AB [Online]*. Available at: https://www.convertus.se/en/technology [Accessed 2017-05-16].

Appendix A

List of used verbs

The following 232 Swedish verbs are used in this thesis, divided into which class they are mapped to in Bloom's taxonomy:

Knowledge: ange, beskriva, citera, definiera, dokumentera, hitta, identifiera, komma ihåg, känna igen, lista, lokalisera, markera, matcha, memorera, namnge, nämna, recitera, redogöra, rubricera, uppge, upprepa, välja, återberätta, återge

Comprehension: abstrahera, använda, approximera, diskutera, exemplifiera, extrapolera, förklara, försvara, förtydliga, förutse, förutspå, förutsäga, ge exempel, generalisera, illustrera, inordna, interpolera, jämföra, karakterisera, karaktärisera, kategorisera, klargöra, klassificera, kontrastera, konvertera, kvantisera, omskriva, omsätta, omvandla, partitionera, prediktera, påvisa, rapportera, referera, renodla, representera, rita, sammanfatta, skilja, skissa, skriva om, summera, tolka, utveckla, åskådliggöra, översätta

Application: addera, anpassa, applicera, bereda, beräkna, bestämma, bevisa, bygga, byta, certifiera, datorbearbeta, demonstrera, derivera, dimensionera, dimensionsanalysera, estimera, experimentera, faktorisera, fastslå, fastställa, fullfölja, förbereda, fördela, förenkla, förverkliga, förändra, genomföra, georeferera, härleda, implementera, installera, integrera, inverstransformera, invertera, kompilera, konfigurera, kvadratkomplettera, linjerisera, lösa, manipulera, modifiera, monteringsanpassa, multiplicera, måttsätta, normalisera, omforma, omkonstruera, organisera, producera, programmera, projektera, redigera, relatera, segmentera, simulera, skatta, skissera, testa, tidsintegrera, tillämpa, transformera, upprätta, upptäcka, utföra, verkställa, ändra, återimplementera, överföra, överslagsberäkna

Analysis: analysera, bearbeta, bryta ned, bryta ner, debattera, dekonstruera, dela in, dela upp, differentiera, fokusera, granska, hänföra, ifrågasätta, införliva, korrelera, kvantifiera, ordna, peka på, peka ut, prioritera, problematisera, rangordna, separera, sortera, strukturera, systematisera, särskilja, tillräkna, tillskriva, undersöka, urskilja, utforska

Synthesis: alstra, antaga, designa, detaljutforma, dra slutsats, formge, formulera, frambringa, framkalla, framställa, föreslå, författa, generera, kombinera, kommunicera, konstruera, kreera, modellera, planera, planlägga, presentera, rekonstruera, sammanställa, skapa, skriva, studieplanera, ställa samman, ställa upp, syntetisera, ta fram, tidsplanera, tillverka, uppföra, utarbeta, utforma

Evaluation: argumentera, avgöra, bedöma, evaluera, komma fram till, kontrollera, kritisera, motivera, mäta, observera, reflektera, rekommendera, resonera, rimlighetsbedöma, samordna, spekulera, stödja, ta ställning, teoretisera, uppskatta, utvärdera, validera, verifiera, värdera, övervaka, överväga

Appendix B

Number of credits distribution

Table B.1 shows the amount of courses having each number of credits. Note that the courses having 6.0, 7.5, 9.0, 15.0 and 30.0 credits together constitute 85.5 % of all courses.

Table B.1: The amount of courses having each number of credits.

#credits	#courses	#credits	#courses
1.0	1	10.5	7
1.5	24	11.0	7
2.0	3	12.0	34
3.0	71	13.0	1
4.0	16	13.5	3
4.5	13	14.0	1
5.0	22	15.0	125
6.0	437	16.0	3
6.5	4	18.0	5
7.0	16	19.0	1
7.5	816	20.0	4
8.0	32	21.0	1
8.5	3	22.5	2
9.0	133	30.0	178
9.5	1	45.0	1
10.0	10		

Appendix C

Further measurements

This appendix contains measurements regarding variables V16–V20 which are of interest to the project provider. The reason for including them in an appendix instead of in Section 4.3 is that they are not part of any definition and not needed to answer the problem statement of this thesis.

C.1 Language of the syllabus

As explained in Section 3.4.1 we identify the language of the syllabus based on the text contents of the following four fields in the syllabus: objectives, content, eligibility and disposition.

Of the Swedish syllabuses, 94.4 % are written in Swedish, 5.6 % are written in English and one syllabus lacks text in the four fields.

Of the English syllabuses, 94.3 % are written in English and 5.7 % lack text in the four fields. No English syllabuses are written purely in Swedish, although there are English syllabuses mixing Swedish and English but predominantly English (e.g. MJ2659) which using the implemented algorithm is classified as English.

C.2 Specific eligibility requirements

In total, 91.1 % of the syllabuses contain specific eligibility requirements.

Figure C.1 shows the percentage of syllabuses containing specific eligibility requirements for each cycle. The preparatory and first cycles have similar values at around 88 % while the second cycle is around 5 percentage points higher at 93.1 %. This is not completely unexpected since second cycle courses often build upon first cycle courses.

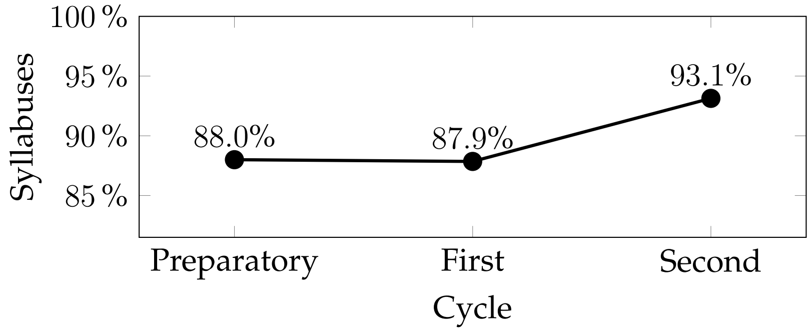

Figure C.1: The percentage of syllabuses containing specific eligibility requirements for each cycle.

C.3 Main fields of study

Table C.2 shows the distribution of how many fields of study all courses belong to for each cycle and in total. Interestingly, no preparatory cycle courses belong to a field of study. In total, most courses have zero, one or two fields of study while only very few courses have three or four fields of study. On average, a course belongs to 1.09 fields of study.

Table C.2: The distribution of how many fields of study all courses belong to for each cycle and in total.

Fields of study	Preparatory	First	Second	Total
0	100 %	5.3 %	11.3 %	10.2 %
1	0 %	77.5 %	71.9 %	73.1 %
2	0 %	17.3 %	13.4 %	14.7 %
3	0 %	0 %	3.2 %	2.0 %
4	0 %	0 %	0.2 %	0.1 %

C.4 Relation between V19 and V20

Of all 12 534 ILOs, 82.0 % contain at least one verb in Bloom's taxonomy. However, only 39.8 % of the syllabuses fulfill the requirement that V19 equals V20 (all ILOs in the syllabus contain at least one verb in Bloom's taxonomy) while in the remaining 60.2 % of the syllabuses V19 is greater than V20.

C.5 Highest Bloom's taxonomy class

Table C.3 shows the distribution of the highest class in Bloom's taxonomy in all 12 534 ILOs. Application and Comprehension are the most used classes at 18.7–17.6 % followed by Synthesis, Evaluation and Knowledge at 13.7–12.5 % and finally Analysis at 6.9 %. Note that 18.0 % of the ILOs do not belong to any class.

Table C.3: The distribution of the highest class in Bloom's taxonomy in all ILOs.

Highest class	ILOs
(no class)	18.0 %
Knowledge	12.5 %
Comprehension	17.6 %
Application	18.7 %
Analysis	6.9 %
Synthesis	13.7 %
Evaluation	12.6 %