Proposal for a standard thesis template

Gerald Q. Maguire Jr.

Each year there are thousands of degree projects done by students at KTH Royal Institute of Technology. These are one of the core elements of undergraduate education and producing high quality degree projects is of interest to the faculty, students, administration, UKÄ, and the public. The reasons for the work described in this document are (1) I have been an examiner for a lot (>560) of degree projects and I want to help students by providing them a template to start with (avoiding the “Blank paper” barrier) and to help facilitate the degree project process, (2) the previous LaTeX template that used to be widely available to students at KTH was no longer being maintained, and (3) the references from the KTH web to this template was removed by the GVS Communications unit.

Since ~2010, all KTH theses should have English **and** Swedish abstracts. Ideally the thesis would also include keywords and titles in both English and Swedish. Moreover, this information is useful not only in the thesis itself, but also in:

1. The announcements of degree project presentations: As, in a number of programs, students doing degree projects need to be an active listener for some number of other degree project presentations; hence, it is desirable to announce these presentation and have students easily about to find this information and select which they would like to attend. Moreover, degree project presentations are supposed to be public, so it would be useful and foster their visibility if it was easy to make such announcements (in both English and Swedish).
2. The final approved thesis needs to be entered into DiVA (both the metadata and the thesis itself) [There is a separate issue regarding whether the full-text is publicly available for not, but this is outside the scope of this document.]
3. Reporting both English and Swedish titles in LADOK.

Therefore, in my roles as an examiner, member of the “språkkommittén (referensgruppen för språkfrågor)”, and to foster the production of theses in both English and Swedish, I have written a DOCX and LaTeX template for 1st and 2nd cycle degree projects (as well as a LaTeX template for third cycle degree projects). Additionally, I have written some software to: (1) make it easier to announce these events to make it easier for people to know about them, so that they can attend, (2) make it easy to make the cover and apply it (or in the case of LaTeX integrate it into the template), and (3) facilitate entering the meta data and thesis into DiVA. There is also an experimental effort to be able to insert the two titles into LADOK (see Section 22).

The focus of this document will be on the LaTeX template for 1st and 2nd cycle degree projects. A separate document examines the DOCX version of the template and a future document will examine the 3rd cycle LaTeX template. There is also a separate document that addresses the quality of the data in DiVA and LADOK as the titles of theses as recorded in LADOK do not necessarily match those in either DiVA or the document itself (as a sample data point for the year 2020 and the EECS school, 9.4% of the entries for titles in LADOK are in error). Additionally, since only a small number of student’s KTHIDs are recorded in the DiVA entry for their thesis it is not easy to mechanically match the LADOK and DiVA records.

Most readers will only be interested in the high-level overview given in Section 1. Some readers may be interested iin accessibility, see Section 18. The rest of the document should contain enough information to let someone else deal with the problems that remain. A easy to read user manual for the LaTeX template is now integrated as an appendix in the template.

.

Table of Contents

[1 Accuracy and Economics 7](#_Toc99212483)

[2 Make it simple from the start 7](#_Toc99212484)

[2.1 Using Information from Canvas and putting information into Canvas 8](#_Toc99212485)

[2.2 How can we associate students with examiners and supervisors? 10](#_Toc99212486)

[2.3 What happens if I do not have such a spreadsheet? 11](#_Toc99212487)

[3 Automating later steps in the degree project 12](#_Toc99212488)

[3.1 Making an announcement 12](#_Toc99212489)

[3.2 Making a cover and applying –new model (the new cover as of December 2021) 13](#_Toc99212490)

[3.3 To facilitate entering a thesis into DiVA – make a MODS file 13](#_Toc99212491)

[3.4 Requirements 13](#_Toc99212492)

[4 Can someone else use these programs? 14](#_Toc99212493)

[5 LaTeX template in Overleaf 14](#_Toc99212494)

[6 Put information into LaTeX template to generate a draft or final thesis 15](#_Toc99212495)

[7 Now that data is in the template, what happens? 21](#_Toc99212496)

[8 How did the abstracts and the keywords appear? 24](#_Toc99212497)

[9 Extracting the information from the PDF file 25](#_Toc99212498)

[10 Optimized extraction from an Overleaf LaTeX project 26](#_Toc99212499)

[11 What can we do with the JSON file? 28](#_Toc99212500)

[12 Actual example 34](#_Toc99212501)

[13 Change in how to enter the abstracts in LaTeX 37](#_Toc99212502)

[14 Adding keywords and PDF meta data 38](#_Toc99212503)

[15 Other variants of the JSON\_to\_calendar.py 43](#_Toc99212504)

[16 JSON to MODS file 43](#_Toc99212505)

[16.1 Example 43](#_Toc99212506)

[16.2 Limitations 56](#_Toc99212507)

[16.3 Limitations overcome 57](#_Toc99212508)

[16.3.1 Entering number in series 57](#_Toc99212509)

[16.3.2 Including KTH affiliations 57](#_Toc99212510)

[17 Some enhancements 63](#_Toc99212511)

[17.1 Enhancements to template and supporting programs 63](#_Toc99212512)

[17.2 Better support for mathematical expressions in abstracts 63](#_Toc99212513)

[17.2.1 Better support for mathematics in Canvas course announcement and course calendar 63](#_Toc99212514)

[17.2.2 Better support for mathematics in Cortina 67](#_Toc99212515)

[17.2.3 Support for mathematics in DiVA 70](#_Toc99212516)

[17.3 Acronyms in abstracts 70](#_Toc99212517)

[17.4 URLs in abstracts 71](#_Toc99212518)

[17.5 Getting the PDF from a Canvas assignment and optionally extracting JSON 71](#_Toc99212519)

[17.6 Support for new cover 71](#_Toc99212520)

[18 Accessibility 72](#_Toc99212521)

[18.1 Calendar entries 72](#_Toc99212522)

[18.2 Cover and PDF file 72](#_Toc99212523)

[18.3 Template itself 73](#_Toc99212524)

[18.4 Improving accessibility 73](#_Toc99212525)

[19 The structure of the template and the report 73](#_Toc99212526)

[20 Generating LaTeX commands for in the student’s thesis source file 74](#_Toc99212527)

[21 Alternative way of inserting the covers 77](#_Toc99212528)

[22 JSON\_to\_ladok.py 78](#_Toc99212529)

[23 The big picture 80](#_Toc99212530)

[24 Alternative way of generating JSON for a LaTeX template 84](#_Toc99212531)

List of Figures

[Figure 1: A student in multiple sections – one for the course, one for the supervisor, and one for the examiner 8](#_Toc99212532)

[Figure 2: Examiner's view of the gradebook 8](#_Toc99212533)

[Figure 3: List of examiners 9](#_Toc99212534)

[Figure 4: Spreadsheet of CS students 10](#_Toc99212535)

[Figure 5: Some of the contents of an alias file for a course 10](#_Toc99212536)

[Figure 6: Section enrollments 11](#_Toc99212537)

[Figure 7: Search people box on People page 12](#_Toc99212538)

[Figure 8: Three vertical dots and Edit Sections button 12](#_Toc99212539)

[Figure 9: Title, subtitle, etc. 13](#_Toc99212540)

[Figure 10 Example of title page of a thesis 16](#_Toc99212541)

[Figure 11 Information about the title and the student authors of the thesis 17](#_Toc99212542)

[Figure 12 Supervisors and examiner information 18](#_Toc99212543)

[Figure 13: Information about where the thesis is taking place 19](#_Toc99212544)

[Figure 14: Information relevant to the oral presentation (both the location and the opponent or opponents) 19](#_Toc99212545)

[Figure 15: Information relevant to the oral presentation (both the location and the opponent or opponents) 19](#_Toc99212546)

[Figure 16: Storing the language in a scontents buffer named "lang" 20](#_Toc99212547)

[Figure 17: LaTeX to produce the English abstract (edited for formatting) 20](#_Toc99212548)

[Figure 18: LaTeX input to produce the Swedish abstract 21](#_Toc99212549)

[Figure 19: First part of the “For DiVA” information 22](#_Toc99212550)

[Figure 20: Second part of the “For DiVA” information (incomplete) 23](#_Toc99212551)

[Figure 21: Code in \divainfo command to output the abstracts and keywords 24](#_Toc99212552)

[Figure 22: Code in the document file to set up the command \typestoredx and to configure the listing environment to put the abstract into. 25](#_Toc99212553)

[Figure 23: Extract the JSON like information 25](#_Toc99212554)

[Figure 24: Example event.json output (note that this is not the same as currently extracted) 25](#_Toc99212555)

[Figure 25: Other logs and files button 26](#_Toc99212556)

[Figure 26: List of log and other files 27](#_Toc99212557)

[Figure 27: The several outputs of JSON\_to\_calendar.py 28](#_Toc99212558)

[Figure 28: Running JSON\_to\_calendar.py to produce all three outputs 28](#_Toc99212559)

[Figure 29: Canvas course announcement 29](#_Toc99212560)

[Figure 30: A course event in the Canvas calendar (the figure is zoomed in on 15 March 2021 30](#_Toc99212561)

[Figure 31: Zoomed view of the opened Canvas calendar event 30](#_Toc99212562)

[Figure 32: KTH’s Cortina calendar showing two degree project events 31](#_Toc99212563)

[Figure 33: English version of the calendar 31](#_Toc99212564)

[Figure 34: Swedish version of the calendar event 32](#_Toc99212565)

[Figure 35: Response from the KTH Cortina API 33](#_Toc99212566)

[Figure 36: Actual example of announcement in Canvas (top) 34](#_Toc99212567)

[Figure 37: Bottom of the announcement 35](#_Toc99212568)

[Figure 38: Opened version in course calendar 35](#_Toc99212569)

[Figure 39: The English (left) and Swedis (right) in the Cortina calendar 36](#_Toc99212570)

[Figure 40: Commands to extract the JSON and to make the calendar entries and announcements 36](#_Toc99212571)

[Figure 41: Extracted JSON file oscar.json - edited for appearance here 36](#_Toc99212572)

[Figure 42: Example of the revised format for entering an abstract 37](#_Toc99212573)

[Figure 43: Second example of the revised format for entering an abstract 38](#_Toc99212574)

[Figure 44: The two macros used to help enter the language specification 38](#_Toc99212575)

[Figure 45: New commands in kththesis.cls 39](#_Toc99212576)

[Figure 46: New additions to examplethesis.text 39](#_Toc99212577)

[Figure 47: lib/pdf\_related\_includes.tex (edited for readability) 40](#_Toc99212578)

[Figure 48: Including the English language keywords below the English abstract 41](#_Toc99212579)

[Figure 49: Including the Swedeish language keywords below the Swedish abstract 41](#_Toc99212580)

[Figure 50: The title page of the thesis and the PDF meta data 42](#_Toc99212581)

[Figure 51: The keywords appear as expected in the for DiVA data at the end of the PDF file 42](#_Toc99212582)

[Figure 52: Several possible inputs to JSON\_to\_calendar.py and its outputs 43](#_Toc99212583)

[Figure 53: Examples of using JSON\_to\_MODS.py 43](#_Toc99212584)

[Figure 54: test12.json 44](#_Toc99212585)

[Figure 55: test12.mods 46](#_Toc99212586)

[Figure 56: The import process – step 1 – click on the “Import from external database” buttom on the upper right corner 46](#_Toc99212587)

[Figure 57: The import process – step 2 – choose the MODV3 format and select a file to import 47](#_Toc99212588)

[Figure 58: The import process – step 3 – after clicking “Import” – it says that the file was successfully uploaded 48](#_Toc99212589)

[Figure 59: The import process – step 4 – after clicking “Import” once more – you can see the file has been added to the list of imported files 48](#_Toc99212590)

[Figure 60: Imported document (part 1) 49](#_Toc99212591)

[Figure 61: Imported document (part 2) – Note that both the English and Swedish titles are shown as well as the information about which level, the number of university credits, the education al program, and the subject.- Additionally, the year and number of pages are shown. 50](#_Toc99212592)

[Figure 62: Imported document (part 3) – Note that the number in the series is not being shown. The nation subject catergies specified in the JSON file are shown. Additionally, a number of the keyoards in different languages are shown. 51](#_Toc99212593)

[Figure 63: Imported document (part 4) – more keywords are shown as well as the abstract in English 52](#_Toc99212594)

[Figure 64: Imported document (part 5) – the bottom of the Engish abstract is shown .- showing some of the formatted information that can be shown. Additionally, the Swedish and other abstracts are shown. 53](#_Toc99212595)

[Figure 65: Imported document (part 6) – yet more abstracts 54](#_Toc99212596)

[Figure 66: Imported document (part 7) – the supervisors are shown 55](#_Toc99212597)

[Figure 67: Imported document (part 8) – the examiner is shown along with the information about the oral presentation 56](#_Toc99212598)

[Figure 68: Number in series correctly entered 57](#_Toc99212599)

[Figure 69: MODS data 57](#_Toc99212600)

[Figure 70: Two KTH supervisors in different schools 58](#_Toc99212601)

[Figure 71: Third supervisor 58](#_Toc99212602)

[Figure 72: Examiner 59](#_Toc99212603)

[Figure 73: MODS data for the supervisors and examiners 60](#_Toc99212604)

[Figure 74: **Corporate name** data in MODS data for the affiliations 61](#_Toc99212605)

[Figure 75: Reduced MODS for corporate names 61](#_Toc99212606)

[Figure 76: The two supervisors with the reduced MODS for corporate names 62](#_Toc99212607)

[Figure 77: The examiner with the reduced MODS for corporate names 62](#_Toc99212608)

[Figure 78: Commands to produce the JSON and make the calendar entries and announcement 63](#_Toc99212609)

[Figure 79: Examples of equations in an announcement 64](#_Toc99212610)

[Figure 80: Announcement summary 65](#_Toc99212611)

[Figure 81: Announcement with equation 65](#_Toc99212612)

[Figure 82: HTML for the announcement 66](#_Toc99212613)

[Figure 83: Response from Instructure support regarding the block equaiotns showing up in the summary of the announcements – Note that the URLs have been added in parenthesis. The screenshot is shown in Figure 84. 66](#_Toc99212614)

[Figure 84: The screenshot mentioned above 67](#_Toc99212615)

[Figure 85: Equations in Cortina claendar entry 68](#_Toc99212616)

[Figure 86: Zoom in on part of the Cortinal calendar entry 69](#_Toc99212617)

[Figure 87: Zoom in on lower part of the Cortinal calendar entry 70](#_Toc99212618)

[Figure 88: Specifying acronyms file 70](#_Toc99212619)

[Figure 89: get\_PDF\_submission program 71](#_Toc99212620)

[Figure 90: New fill\_in\_template.py program 71](#_Toc99212621)

[Figure 91: Files and folders in Overleaf template 74](#_Toc99212622)

[Figure 92: Generating LaTeX commands for a student author 75](#_Toc99212623)

[Figure 93: Generating the LaTeX data for the second author (in the case of a 1st cycle thesis) 76](#_Toc99212624)

[Figure 94: Attempting to generate a 2nd author for a 2nd cycle thesis 76](#_Toc99212625)

[Figure 95: Generating LaTeX commands for a supervisor 77](#_Toc99212626)

[Figure 96: Generating LaTeX commands for the examiner 77](#_Toc99212627)

[Figure 97: Inseert this include file either in the main text document or the includes.tex file 77](#_Toc99212628)

[Figure 98: Insert the front cover before the title page 77](#_Toc99212629)

[Figure 99: Insert the back cover page before the For DIVA section 77](#_Toc99212630)

[Figure 100: Using the extracted JSON to produce a LADOK entry 80](#_Toc99212631)

[Figure 101: New version of \Insert\Keywords tother with making the expl3 \strcompare function available. 85](#_Toc99212632)

[Figure 102: Original \schoolAcronym 85](#_Toc99212633)

[Figure 103: new \schoolAcronym 86](#_Toc99212634)

[Figure 104: New version of \programcode 87](#_Toc99212635)

[Figure 105: Code to write the first author's information to a "JSON" file 88](#_Toc99212636)

[Figure 106: New code to ouput the "Other information" including page numbers 89](#_Toc99212637)

[Figure 107: Function to get the scontents 89](#_Toc99212638)

[Figure 108: Writing out the language, abstract, and keywords information 90](#_Toc99212639)

[Figure 109: Resulting announcement 96](#_Toc99212640)

List of Tables

[Table 1: Information and how it is used 81](#_Toc99212641)

[Table 2: Resulting fordiva.json file 91](#_Toc99212642)

[Table 3: Resut after cleanup for the contents of the earlier fordiva.json file, called fordiva-cleaned 94](#_Toc99212643)

[Table 4: Source LaTeX (from inside the scontents of the abstract= 97](#_Toc99212644)

# Accuracy and Economics

An earlier 1st cycle thesis[[1]](#footnote-1) looked at the automation of entering a thesis and its metadata into DiVA and found that a large fraction of existing entries with manually entered metadata in DiVA had erors in them and also estimated that the total number of full-time equivalent (FTE) hours spent entering this data was several FTEs equivalent. Note that the change to having only administrative staff entering the meta data has *not* eliminated the problem of errors in data entry.

**Hypothesis**: Incorporating the relevant data into the thesis document will facilitate both announcements and DiVA data entry.

**Approach**: Make this data readily available via the template and collect & output it in a   
“For DIVA” set of information at the end of the PDF document.

# Make it simple from the start

If a Canvas course room is set up for degree projects one can greatly simplify the many processes involved for all the stakeholders (students, faculty, and staff). The student is assumed to be enrolled in the Canvas course room and that one knows the student’s e-mail address (or in the case of a 1st cycle thesis possible two students’ addresses) and the e-mail address of the examiner and supervisor[[2]](#footnote-2) and possibly an external supervisor exists.

For two 1st cycle students (s1 and s2), myself as the examiner, and Anders Västberg as the academic supervisor and an unknown supervisor at a company – together with the program code and course code, generating the pre-configuring of the LaTeX template is as simple as two (single line) commands:

./create\_customized\_JSON\_file.py --canvas\_course\_id 31167 --author s1@kth.se --author2 s2@kth.se --language eng --programCode TIDAB  
--Examiner maguire --Supervisor vastberg --Supervisor2 xxX  
--courseCode II142X --exam högskoleingenjör

This gets relevant data from Canvas and makes a JSON file with the relevant information in it. The xxx in the above will just generate dummy entries for an external supervisor.

./customize\_LaTeX\_project.py --json customize.json --file z23.zip --initialize

This takes a local zip file of the LaTeX project (in z23.zip) and adds the customization information to it. It produces a z23-modified.zip file - that I rename for the student and send to the student. The student only has to upload this into Overleaf and off they go. If they then invite the examiner and supervisor(s) to the Overleaf project they can directly see what the student is doing, comment, etc. This greatly simplifies things and collects some of the needed data already at the start of the degree project.

For a 2nd cycle thesis, an example of the first command would be:

./create\_customized\_JSON\_file.py --canvas\_course\_id 33514 --author s1 --language eng --programCode TCSCM --Examiner maguire --Supervisor vastberg --Supervisor2 xxx --courseCode DA231X --exam master

The same second command is used to make the customized LaTeX project for the student.

## Using Information from Canvas and putting information into Canvas

What kind of information can be gottern from Canvas and what can you do with it?

Given the course course\_id (31167 and 33514 in the examples above), you can look up information about the student via the Canvas course room. For example, if we assume the course\_id is 33514 (the EECS 2nd cycle degree project Canvas course room for lots of different courses) and the student is taking the course DA231X. When the student is enrolled in this Canvas course room they are added to a section that has their course code in it. For example:

Section for the course DA231X VT22-2 Degree Project in Computer Science and Engineering, Second Cycle

Sections are a very powerful mechanism in Canvas for organizing students and a student can be in up 18 different sections in a given Canvas course room. For example, it is possible to create sections for each of the examiners and even the academic supervisor(s). Now one can add the student to the sections for the examiner and supervisor(s). The result is that the student now appears in these different sections, see Figure 1.

|  |
| --- |
|  |

Figure 1: A student in multiple sections – one for the course, one for the supervisor, and one for the examiner

This greatly simplifies the work for the examiner and supervisor as now they can choose to view the gradebook and only need to see their students, see Figure 2.

|  |
| --- |
|  |

Figure 2: Examiner's view of the gradebook

Note that in Figure 2, there are two columns “Course code” and “Examiner”. These are what I call *administrative assignments* and the students do not need to see them (hence the eye symbol with a slash through it – indicating these are hidden from the students). In the “Course code” assignment, I have given the student a “grade” that is the course code of their degree project. This way it is very easy to see what course code a student is enrolled in – as an examiner may have students enrolled in several different course codes. The “grading scale” was created from the list of all the EECS 2nd cycle course codes that share this same Canvas course room. The “Examiner”.column contains a “grade” that is the name of the examiner assigned for this student. The examiner “grading scale” is created from the list of all of the examiners in the Canvas course room. Just the same as the way that a teacher can assign a grade, if you click on the gradebook cell for a student, you get a pulldown list of all of the possible “grades” in this case the list of examiners, see Figure 3.

|  |
| --- |
|  |

Figure 3: List of examiners

Did I have to manually enter all of the course codes for the 687 students in the course room? No, I used a script with the following command:

./add\_course\_codes\_for\_students\_in\_course.py 33514

Similarly, the grading scale with course codes was created with:

./insert\_course\_code\_grading\_standard.py 33514

The grading scale for examiners (and even one for teachers as supervisor) is created with:

./insert\_teachers\_grading\_standard.py –-examiners 33514

The above line creates a gradting scale with the suffix “\_Examiners”. Note that all of these grading scales are course specific grading scales, so they are not visible outside of this course. One caution about the use of grading scales is that if the list of course codes or list of teachers/examiners changes – there can be problems with the mappings as the grading scale creates a mapping between a numeric grade and a string (the course code or name); therefore, they should only be generated once at the starto f the course when the full list of values is known.

So clearly we can have information about the student, supervisors (who are teachers in the course), and examiner in the Canvas course room. Then since these people are in the Canvas course room we can know their sortable name, KTHID, e-mail address, etc. Clearly the students were added to the course when they were enrolle din the course. How did the teachers and examiners get into the course? These people were added to the course by a script that IT runs that takes this data from KOPPS!

## How can we associate students with examiners and supervisors?

In the case of the computer science (CS) students the exjobb coordinator Mats Nordahl has a spreadsheet with the student’s name, working title, examiner, and supervisors in it. It has a sheet named “Closed” that contains the information about the students who have been assigned an examiner and supervisor. Figure 4 shows the column headings of the spreadsheet “Masters\_thesis\_proposals-CS-P3-2022a.xlx” (this is my local file name for the spreadsheet).

|  |
| --- |
|  |

Figure 4: Spreadsheet of CS students

The following shows the use of a script that takes the data from this spreadsheet and adds it to the Canvas gradebook with the command:

./insert-examiners-and-supervisors-from-spreadsheet.py 33514 "Masters\_thesis\_proposals-CS-P3-2022a.xlsx"

The script uses the student’s e-mail address to look up the student in the Canvas course room, then adds the student to the examiner’s section and updates the “Examiner” grade for this student to show this assigned examiner. It also add the contents of “Proposal” as text in a custom column in the gradebook called “Tentative\_title” and adds the list of supervisors to the custom column called “Supervisors” – see Figure 2. Note that this program uses a file, spreadsheet\_aliases-33514.json, containing aliases for each of the examiners and supervisors. Figure 5 shows part of the alias file.

|  |
| --- |
| {  "examiners": {  "Anders Västberg": "Västberg, Anders",  …  "Erik Fransen": "Fransén, Erik",  "Erik Fransén": "Fransén, Erik",  "Gerald Maguire": "Maguire Jr, Gerald Quentin",  "Gerald Q. Maguire": "Maguire Jr, Gerald Quentin",  …  "joakim gustafson": "Gustafsson, Joakim",  "Joakim Gustafsson": "Gustafsson, Joakim",  "Joakim Gustafson": "Gustafsson, Joakim",  …  },  "teachers": {  …  "Bob Sturm": "Sturm JR, Bobby Lee Townsend",  "Christopher Peters": "Peters, Christopher",  …  "Mats Nordahl": "Nordahl, Mats",  …  "Wafaa A. H.": "Mushtaq, Wafaa",  "Wafaa A.H.": "Mushtaq, Wafaa",  "Wafaa A.H": "Mushtaq, Wafaa",  …  }  } |

Figure 5: Some of the contents of an alias file for a course

The alias file takes the names used in the spreadsheet and maps them to the sortable names used in Canvas. While this adds a little more complexity, it addressed the problem that the names used in the spreadsheet are not always consistent or even complete.

The availability of information in Canvas also suggests that some of the arguments that were shown earlier for the script create\_customized\_JSON\_file.py are actually *unnecessary* and the program can figure them out given the course\_id and student’s e-mail address. However, somethings are not accessible directly from Canvas, for example I have not been able to figure out what exam a student intends to apply for following their degree project course[[3]](#footnote-3). Another example of missing information is the program code; however, this information is in LADOK and Canvas has the LADOK identifier for each student; however, my script to access LADOK is currently not working.

## What happens if I do not have such a spreadsheet?

A Canvas administrator can add students to the examiner’s and supervisor’s sections, as shown in Figure 6. Note that the input field can do partial matches on the text that you enter and will give you a list of the partially matching entries and then you can click on on and then click on Update.

|  |
| --- |
|  |

Figure 6: Section enrollments

To get the Section Enrollments editing box popup (shown above) – you go to the “People” page – find the student (either scrolling down the page or starting to type the student’s name in “Search people” box (shown in Figure 7) then click on the vertical three dots on the right side of the row for this student and click on Edit section (as shown in Figure 8).

|  |
| --- |
|  |

Figure 7: Search people box on People page

|  |
| --- |
|  |

Figure 8: Three vertical dots and Edit Sections button

Now that we have tied together the student, examiner, supervisors we can make use of that in many ways. One of these ways was to create the customized LaTeX project described earlier. The next section will explain some of the additional steps in the degree project that can be partially automated.

# Automating later steps in the degree project

There are many different steps in the degree project that can be easied by using the information from the Canvas course room and from the student’s thesis. An example of one of these is making an announcement for the student’s oral presentation (described in Section 3.1).

## Making an announcement

Assuming that a student[[4]](#footnote-4) has submitted their thesis with the information in the “For DIVA” pages and that this information includes the information about the opponent(s) and presenation. The steps to make an announcement are:

1. Save the PDF file, for example: oscar.pdf
2. Extract the “For DIVA” information as JSON

./extract\_pseudo\_JSON-from\_PDF.py --pdf oscar.pdf --json oscar.json

If acronyms are used in the abstracts, you can add the “--acronyms acronyms.tex” aregument to the extract command line and the program will process the acronyms[[5]](#footnote-5). You will get a JSON file (oscar.json) that can be used to make an announcement:

1. Make the announcement for a Canvas course room (with course\_id 11):  
   ./JSON\_to\_calendar.py -c 11 --json oscar.json

The above will publish the announcement in the Canvas course 11, in the Canvas course’s calendar, and (eventually) in the KTH Calendar. The software supports the development version of the KTH Cortina API. However, this is API not yet in production and requires a KTH Cortina Access Key.

Replace 11 with the course\_id of the Canvas course room, for example: 33514 for the EECS 2nd cycle degree projects. For more information see Sections 11 and 15.

## Making a cover and applying –new model (the new cover as of December 2021)

As of December 2021, there is a new cover for 1st and 2nd cycle theses at KTH. Information about this is at <https://intra.kth.se/en/administration/kommunikation/mallar/avhandlingarochexamensarbeten/skapa-ett-omslag-till-ditt-exjobb-1.479838>

In the case of the LaTeX project, the cover is **automatically** made and updated as the student writes. The only thing that is missing is the TRITA number. The TRITA number needs to be (1) added to the project before the final printing or (2) alternatively it can be entered by editing the place holder text on backcover when the TRITA number is assigned (note that in this case, this data has to be (a) manually added to the JSON file used to make the MODS file or (b) manually edited in DiVA entry).

To set up the title and subtitle in English and Swedish the student simply edits the relevant lines in the LaTeX file, see Figure 9. This is futher described in detail in Section 6.

|  |
| --- |
| \title{This is the title in the language of the thesis}  \subtitle{A subtitle in the language of the thesis}  % give the alternative title - i.e., if the thesis is in English, then give a Swedish title  \alttitle{Detta är den svenska översättningen av titeln}  \altsubtitle{Detta är den svenska översättningen av undertiteln}  % alternative, if the thesis is in Swedish, then give an English title  %\alttitle{This is the English translation of the title}  %\altsubtitle{This is the English translation of the subtitle} |

Figure 9: Title, subtitle, etc.

So what we have seen is that we can use the information from Canvas and the thesis itself for simplifying many steps. This is true because the assumption is that the relevant data will be in the Canvas course room and in the thesis itself!

## To facilitate entering a thesis into DiVA – make a MODS file

Assuming that a student has submitted a thesis with the information in the “For DIVA pages” that includes the information for the DIVA entry **and** the examiner has approved the thesis.

The same steps as above are made to get the JSON file, then make a MODS file:

./JSON\_to\_MODS.py -c 11 --json oscar.json --trita "TRITA-EECS-EX-2021:xxxx"

For more information see Section 16.

## Requirements

To run the program to make announcements and calendar entries for a course rooms requires the program and appropriate permissions (for example as a Teacher) in the Canvas course room. To make the entry in the KTH Calendar requires a KTH Cortina access key.

# Can someone else use these programs?

The source code is at <https://github.com/gqmaguirejr/E-learning>

Any users with the Canvas permissions to post announcements in a course and insert course calendar events can use the basic Canvas-related functionality. However, to change sections (adding/removing/… a user to/from a section) requires the permission of a Canvas administrator. Additionally, the KTH Cortina Calendar API requires an access key (which you have to get from the IT unit). Get the programs from the github URL above. Create a config.json file to provide your Canvas access token and the URL of the Canvas instance you want to use, for details see <https://canvas.kth.se/courses/11/pages/making-a-config-dot-json-file-to-make-life-simpler> ..

# LaTeX template in Overleaf

A LaTeX template in Overleaf is at <https://www.overleaf.com/read/qxvttmmqbgdt> - this is the version used by many EECS students[[6]](#footnote-6).

Note that the rest of this document goes into deep detail about how the template works and how the programs work. It is primarily intended to give the information necessary so that someone else can take over the template and programs. Also note that Section 17 addressess issues about accessibility.

# Put information into LaTeX template to generate a draft or final thesis

Figure 10 shows the title page of a ficticous thesis created using a LaTeX.template that I have developed. The idea is to capture the information needed for the announcement of the oral presentation and for reporting the final approved thesis in DiVA via the thesis itself.

For the current documentation of the template for these details see the file “README\_notes.tex” in the template itself.

For the purpose of this document, the following information is present in the order of this exposition, not necessarily the order it is in the LaTeX source file. However, all of this information is prior to \begin{document}. Moreover, the actual template has been reorganized since the following text was written. In particular, information that should be knowable at the start of the degree project is placed into a file named custom\_configuration.tex – to make it easier to pre-configure a project for a student or students (as was described in Section 2).

Information about the title and the student author(s) of the thesis is entered via the set of macros as shown in Figure 11. Note that you can have either one or two authors (the latter in the case of a 1st cycle degree project). There are several other elements of metadata collected about the student (primarily driven by the current author metadata fields in DiVA). A brief description of them and why they are there are given below:

* One might question why have an e-mail address (when under the current policy the student will lose their e-mail address upon graduation)? One of the main reasons is so that the library (i.e., KTHB) can notify the student that the thesis has be stored in DiVA. A second reason is if the library needs to contact the student.
* One might also wonder, why include the student’s kthid. The main reason is that this identifier uniquely identifies this author within KTH, so that all of their publications can be found in a DiVA search. Note that a number of KTH students go on to write papers related to their thesis and a number of these are registered in DiVA.
* A very questionable field is authorsSchool. In a discussion with DiVA administrators at KTH on 2021-04-29, the consensus was that this should be the school of the thesis examiner, since 1st and 2nd cycle students are in *programs of study* and **not** schools, department, etc.
* Next there is programcode – this is used to generate the degree information just above the data on the title page and is also used to compute the “lead” for the calendar entry (which differs depending upon whether it is a 1st or 2nd cycle degree project presentation). Finally, the \degreeName{Bachelors degree} \subjectArea{Technology}- this information is also used later when generating the color KTH cover.

**NB**: A limitation of the current template is that I do not handle the case of two students who are in *different* programs.

|  |
| --- |
|  |

Figure 10 Example of title page of a thesis

|  |
| --- |
| %% Information for inside title page  \title{This is the title in the language of the thesis}  \subtitle{An subtitle in the language of the thesis}  % give the alternative title - i.e., if the thesis is in English, then give a Swedish title  \alttitle{Detta är den svenska översättningen av titeln}  \altsubtitle{Detta är den svenska översättningen av undertiteln}  \authorsLastname{Student}  \authorsFirstname{Fake A.}  \email{a@kth.se}  \kthid{u100001}  \authorsSchool{\schoolAcronym{EECS}}  \programcode{TCOMK}  \degreeName{Bachelors degree}  % Note that the subject area for a Bachelor's thesis (Kandidatexamen) should be either Technology or Architecture  % If the thesis is in Swedish these would be: teknik | arkitektur -- Note the use of lower case  \subjectArea{Technology}  % If there is a second author - add them here:  \secondAuthorsLastname{Student}  \secondAuthorsFirstname{Fake B.}  \secondemail{b@kth.se}  \secondkthid{u100002}  \secondAuthorsSchool{\schoolAcronym{ABE}} |

Figure 11 Information about the title and the student authors of the thesis

Figure 12 shows the information entered about the supervisor or supervisors and the examiner.

**NB**: A limitation of the current template is that I do not handle (a) more than three supervisors or (b) the case of multiple examiners.

|  |
| --- |
| \supervisorAsLastname{Supervisor}  \supervisorAsFirstname{A. Busy}  \supervisorAsEmail{sa@kth.se}  % If the supervisor is from within KTH add their KTHID, School and Department info  \supervisorAsKTHID{u100003}  \supervisorAsSchool{\schoolAcronym{EECS}}  \supervisorAsDepartment{Computer Science}  % other for a supervisor outside of KTH add their organization info  %\supervisorAsOrganization{Timbuktu University, Department of Pseudoscience}  %If there is a second supervisor add them here:  \supervisorBsLastname{Supervisor}  \supervisorBsFirstname{Another Busy}  \supervisorBsEmail{sb@kth.se}  % If the supervisor is from within KTH add their KTHID, School and Department info  \supervisorBsKTHID{u100003}  \supervisorBsSchool{\schoolAcronym{ABE}}  \supervisorBsDepartment{Public Buildings}  % other for a supervisor outside of KTH add their organization info  %\supervisorBsOrganization{Timbuktu University, Department of Pseudoscience}  \examinersLastname{Maguire Jr.}  \examinersFirstname{Gerald Q.}  \examinersEmail{maguire@kth.se}  % If the examiner is from within KTH add their KTHID, School and Department info  \examinersKTHID{u100004}  \examinersSchool{\schoolAcronym{EECS}}  \examinersDepartment{Computer Science}  % other for a examiner outside of KTH add their organization info  %\examinersOrganization{Timbuktu University, Department of Pseudoscience} |

Figure 12 Supervisors and examiner information

Figure 13 shows how to enter data about where the thesis is being done if outside of KTH.

**NB**: A limitation of the current template is that I do not handle multiple companies, as I assume that there is a single host company. However, you can have a list of names within the two text fields (but only a hostcompany or a hostorganization).

|  |
| --- |
| \hostcompany{Företaget AB} % Remove this line if the project was not done at a host company  %\hostorganization{CERN} % if there was a host organization  \date{\today} |

Figure 13: Information about where the thesis is taking place

Figure 14 collects the information regarding the time, place, and language of the presentation. Note that the opponents names are simply separated by ‘\&’ – so it is easy to have one more opponents.

|  |
| --- |
| %%%%% for the oral presentation  \presentationDateAndTimeISO{2021-03-15 13:00}  \presentationLanguage{eng}  \presentationRoom{via Zoom https://kth-se.zoom.us/j/ddddddddddd}  %\presentationAddress{}  \presentationCity{Stockholm}  % Opponent's information  \opponentsNames{A. B. Normal \& A. X. E. Normalè}  \nationalsubjectcategories{10201, 10206} |

Figure 14: Information relevant to the oral presentation (both the location and the opponent or opponents)

Finally, Figure 15 collects the information regarding the National Subject Categories – this is simply a list of 3 or 5 digit numbers separated by commas. The numbers come from <https://www.scb.se/contentassets/10054f2ef27c437884e8cde0d38b9cc4/oversattningsnyckel-forskningsamnen.pdf> while the Swedish and English versions are given in <https://www.scb.se/contentassets/3a12f556522d4bdc887c4838a37c7ec7/standard-for-svensk-indelning--av-forskningsamnen-2011-uppdaterad-aug-2016.pdf>. This information is for a required field in DiVA. Note that 5 digit codes are preferred over 3 digit codes.

|  |
| --- |
| \nationalsubjectcategories{10201, 10206} |

Figure 15: Information relevant to the oral presentation (both the location and the opponent or opponents)

The program create\_customized\_JSON\_file.py tries to guess the National Subject Categories based upon the particular course code that the student is enrolled in. Currently only the categories that I thought were useful for the EECS 2nd cycle courses codes are well addressed.

Figure 17 and Figure 18 show examples of abstracts that in a real thesis would be in English and Swedish with the first to appear being the abstract in the language of the thesis. Note that the actual content of these two abstracts is primarily for testing and is not meant to suggest real abstracts.

The template also supports a number of other languages (based upon the languages used for abstracts in undergraduate theses in 2020). It is straight forward to add an additional language as necessary. One of the reason for having abstracts in additional languages so that dual degree students do not have to write another document for their home/other university. While the template includes a number of place holders for these other abstracts, if they are unused they can simply be deleted.

The three character code used for the language is the ISO 639-2 Code – specifically the "B" (bibliographic) variant of these codes as these seem to be the codes used in DiVA when one access the MODS formatted metadata for publications. In the example below we see “eng” being stored into a scontents buffer called “lang”.

|  |
| --- |
| \begin{scontents}[store-env=lang]  eng  \end{scontents} |

Figure 16: Storing the language in a scontents buffer named "lang"

The abstract itself is stored into an scontents buffer called “abstracts” and the keywords are stored in an scontents buffer called “keywords”. These buffers are part of the LaTeX scontents package and allow contents to be stored and later retrieved.

|  |
| --- |
| \begin{abstract}  \markboth{\abstractname}{}  \begin{scontents}[store-env=lang]  eng  \end{scontents}  \begin{scontents}[store-env=abstracts,print-env=true]  All theses at KTH are \textbf{required} to have an abstract in both \textit{English} and \textit{Swedish}.  Exchange students many want to include one or more abstracts in the language(s) used in their home institutions to avoid the need to write another thesis when returning to their home institution.  Keep in mind that most of your potential readers are only going to read your \texttt{title} and \texttt{abstract}. This is why it is important that the abstract give them enough information that they can decide is this document relevant to them or not. Otherwise the likely default choice is to ignore the rest of your document.  A abstract should stand on its own, i.e., no citations, cross references to the body of the document, acronyms must be spelled out, … .  Write this early and revise as necessary. This will help keep you focused on what you are trying to do.  Example of a formula in an abstract: $c=2 \cdot \pi \cdot r$ or \[ \int\_{a}^{b} x^2 \,dx \]  two chemical formulas: H\textsubscript{2}O or $(C\_5O\_2H\_8)\_n$, copyright symbol: \textcopyright Maguire 2021, and some superscripts: \textsuperscript{99m}Tc, A\textsuperscript{\*},  A\textsuperscript{\textregistered}, and A\texttrademark.  Write an abstract with the following components:% key parts of the abstract  \begin{itemize}  \item …  \end{itemize}  % comment at end  \end{scontents}  \subsection\*{Keywords}  \begin{scontents}[store-env=keywords,print-env=true]  Canvas Learning Management System, Docker containers, performance tuning  \end{scontents}  \end{abstract} |

Figure 17: LaTeX to produce the English abstract (edited for formatting)

|  |
| --- |
| \selectlanguage{swedish}  \begin{abstract}  \markboth{\abstractname}{}  \begin{scontents}[store-env=lang]  swe  \end{scontents}  \begin{scontents}[store-env=abstracts,print-env=true]  Alla avhandlingar vid KTH måste ha ett abstrakt på både engelska och svenska.  If you are writing your thesis in English, you can leave this until the final version. If you are writing your thesis in Swedish then this should be done first – and you should revise as necessary along the way.  If you are writing your thesis in English, then this section can be a summary targeted at a more general reader. However, if you are writing your thesis in Swedish, then the reverse is true – your abstract should be for your target audience, while an English summary can be written targeted at a more general audience.  This means that the English abstract and Swedish sammnfattning  or Swedish abstract and English summary need not be literal translations of each other.  The abstract in the language used for the thesis should be the first abstract, while the Summary/Sammanfattning in the other language can follow.  \end{scontents}  \subsection\*{Nyckelord}  \begin{scontents}[store-env=keywords,print-env=true]  Canvas Lärplattform,Dockerbehållare, prestandajustering  \end{scontents}  \end{abstract} |

Figure 18: LaTeX input to produce the Swedish abstract

# Now that data is in the template, what happens?

The first thing that happens in the LaTeX code automatically generates one or more pages at the end of the PDF document that contain the data – primarily to be used to report the final approved thesis in DiVA. However, a secondary use is that if this information is added to the draft copy that is going to the opponent – then one can potentially automate many of the steps in announcing the oral presentation.

Figure 19 and Figure 20 show this “For DIVA” information. The format is supposed to look like JSON.

The ”organisation”: {”L1”: ”School of Electrical Engineering and Computer Science ”,”L2”: ”Computer Science” }} that for the examiner is used to determine which local part of the KTH calendar a calendar announcement should appear in. The Cortina calendar is divided by school and then by department. Note that the department name must be in Swedish.

|  |
| --- |
|  |

Figure 19: First part of the “For DiVA” information

|  |
| --- |
| …    … |

Figure 20: Second part of the “For DiVA” information (incomplete)

# How did the abstracts and the keywords appear?

The \divainfo command that generates the For DiVA information pages has the following little bit of code that walks the set of scontents buffers using the lang scontents buffer to put the language and the corresponding abstract and keywords (see Figure 21). Originally, I used \getstored[\i]{abstracts} to get an abstract, but this turns out to process the LaTeX into something to render in the PDF. However, I realized that it would be far better to get the actual LaTeX source code and then process it myself into HTML for the announcement and calendars.

The \typestored command that the scontents package provides will not take a variable argument, i.e., it only takes a constant, such as \ typestored[2]{abstracts}. Unfortunately, I need to have a loop to handle the variable number of abstracts that could be used. To do this required a new command \typestoredx that evaluates the variable and the calls the internal function that gets the contents of the correct scontents buffer! This new command is shown in Figure 22 – it uses ExplSyntax – see the LaTeX package expl3. The quad euro symbols are used as markers to avoid problems with quotation marks in the abstract itself. I have assumed that such a combination of characters will never occur in an abstract.

|  |
| --- |
| "Number of lang instances": \textquotedbl\relax\countsc{lang}\textquotedbl\relax,\\  \foreach \i in {1,...,\countsc{lang}} {  "Abstract[\getstored[\i]{lang}]": €€€€\\  \typestoredx{\i}{abstracts}  €€€€,\\  "Keywords[\getstored[\i]{lang}]": €€€€\\  \getstored[\i]{keywords}  €€€€,\\  } |

Figure 21: Code in \divainfo command to output the abstracts and keywords

Figure 22 also shows how to configure the listings environment to put the abstract into. One of the tricks here is that it was important to reduce the hyphenation penalty to enable the abstract text to nicely wrap text on the page. As an added benefit, the LaTeX syntax highlighting is on – so one can easily see the LaTeX commands that are used – as this might need manual editing before the event is announced.

|  |
| --- |
| \ExplSyntaxOn  \newcommand\typestoredx[2]{\expandafter\\_\_scontents\_typestored\_internal:nn\expandafter{#1} {#2}}  \ExplSyntaxOff  \makeatletter  \let\verbatimsc\@undefined  \let\endverbatimsc\@undefined  \lst@AddToHook{Init}{\hyphenpenalty=50\relax}  \makeatother  \lstnewenvironment{verbatimsc}  {  \lstset{%  basicstyle=\ttfamily\tiny,  %basicstyle=\tiny,  %columns=fullflexible,  columns=[l]fixed,  language=[LaTeX]TeX,  %numbers=left,  %numberstyle=\tiny\color{gray},  keywordstyle=\color{red},  breaklines=true, % sets automatic line breaking  breakatwhitespace=true, % sets if automatic breaks should only happen at whitespace  %keepspaces=false,  breakindent=0em,  %fancyvrb=true  }  }{} |

Figure 22: Code in the document file to set up the command \typestoredx and to configure the listing environment to put the abstract into.

# Extracting the information from the PDF file

Now that the JSON-like information is in the PDF file, the next step is to extract it. We use a command line, as shown in Figure 23, to extract the information. Figure 24 shows an example of the resulting file.

|  |
| --- |
| ./extract\_pseudo\_JSON-from\_PDF.py --pdf test5.pdf --json event.json |

Figure 23: Extract the JSON like information

|  |
| --- |
| {"Author1": {"Last name": "Student", "First name": "Fake A.", "Local User Id": "u100001", "E-mail": "a@kth.se", "ORCiD": "0000-0002-00001-1234", "organisation": {"L1": "School of Electrical Engineering and Computer Science "}}, "Author2": {"Last name": "Student", "First name": "Fake B.", "Local User Id": "u100002", "E-mail": "b@kth.se", "ORCiD": "0000-0002-00001-5678", "organisation": {"L1": "School of Architecture and the Built Environment "}}, "Degree": {"Educational program": "Bachelor’s Programme in Information and Communication Technology"}, "Title": {"Main title": "This is the title in the language of the thesis", "Subtitle": "An subtitle in the language of the thesis", "Language": "eng"}, "Alternative title": {"Main title": "Detta är den svenska översättningen av titeln", "Subtitle": "Detta är den svenska översättningen av undertiteln", "Language": "swe"}, "Supervisor1": {"Last name": "Supervisor", "First name": "A. Busy", "Local User Id": "u100003", "E-mail": "sa@kth.se", "organisation": {"L1": "School of Electrical Engineering and Computer Science ", "L2": "Computer Science"}}, "Supervisor2": {"Last name": "Supervisor", "First name": "Another Busy", "Local User Id": "u100003", "E-mail": "sb@kth.se", "organisation": {"L1": "School of Architecture and the Built Environment ", "L2": "Public Buildings"}}, "Examiner1": {"Last name": "Maguire Jr.", "First name": "Gerald Q.", "Local User Id": "u100004", "E-mail": "maguire@kth.se", "organisation": {"L1": "School of Electrical Engineering and Computer Science ", "L2": "Computer Science"}}, "Cooperation": {"Partner\_name": "Företaget AB"}, "Other information": {"Year": "2021", "Number of pages": "xxxiii,35"}, "Opponents": {"Name": "A. B. Normal & A. X. E. Normalè"}, "Presentation": {"Date": "2021-03-16 13:00", "Language": "eng", "Room": "via Zoom", "City": "Stockholm"}, "Number of lang instances": "10", "abstracts": {"eng": "<p>All theses at KTH are <bold>required</bold> to have an abstract in both <i>English</i> and <i>Swedish</i>.</p><p>Exchange students many want to include one or more abstracts in the language(s) used in their home institutions to avoid the need to write another thesis when returning to their home institution.</p><p>Keep in mind that most of your potential readers are only going to read your <tt>title</tt> and <tt>abstract</tt>. This is why it is important that the abstract give them enough information that they can decide is this document relevant to them or not. Otherwise the likely default choice is to ignore the rest of your document.</p><p>A abstract should stand on its own, i.e., no citations, cross references to the body of the document, acronyms must be spelled out, … .</p><p>Write this early and revise as necessary. This will help keep you focused on what you are trying to do.</p><p>Example of a formula in an abstract: $c=2 \\cdot \\pi \\cdot r$ or \\[ \\int\_{a}^{b} x^2 \\,dx \\] two chemical formulas: H<sub>2</sub>O or $(C\_5O\_2H\_8)\_n$, copyright symbol: &copy; Maguire 2021, and some superscripts: <sup>99m</sup>Tc, A<sup>\*</sup>, A<sup>&reg;</sup>, and A&trade;.</p><p>Write an abstract with the following components: </p><ul><li> What is the topic area? (optional) Introduces the subject area for the project. </li><li> Short problem statement </li><li> Why was this problem worth a ’Masters thesis project? (i.e., why is the problem both significant and of a suitable degree of difficulty for a ’Masters thesis project? Why has no one else solved it yet?) </li><li> How did you solve the problem? What was your method/insight? </li><li> Results/Conclusions/Consequences/Impact: What are your key results/conclusions? What will others do based upon your results? What can be done now that you have finished - that could not be done before your thesis project was completed?</li></ul>", "swe": "<p>Alla avhandlingar vid KTH måste ha ett abstrakt på både engelska och svenska.</p><p>If you are writing your thesis in English, you can leave this until the final version. If you are writing your thesis in Swedish then this should be done first – and you should revise as necessary along the way.</p><p>If you are writing your thesis in English, then this section can be a summary targeted at a more general reader. However, if you are writing your thesis in Swedish, then the reverse is true – your abstract should be for your target audience, while an English summary can be written targeted at a more general audience.</p><p>This means that the English abstract and Swedish sammnfattning or Swedish abstract and English summary need not be literal translations of each other.</p><p>The abstract in the language used for the thesis should be the first abstract, while the Summary/Sammanfattning in the other language can follow.</p>", "fre": "<p>Résumé en français.</p>", "spa": "<p>Résumé en espagnol.</p>", "ita": "<p>Sommario in italiano.</p>", "nor": "<p>Sammendrag på norsk.</p>", "ger": "", "dan": "<p>Abstrakt på dansk.</p><p>Zusammenfassung in Deutsch.</p>", "dut": "<p>Samenvatting in het Nederlands.</p><p>Eesti keeles kokkuvõte.</p>", "est": ""}, "keywords": {"eng": "Canvas Learning Management System, Docker containers, performance tuning ", "swe": "Canvas Lärplattform,Dockerbehållare, prestandajustering ", "fre": "5-6 mots-clés ", "spa": "5-6 Palabras claves ", "ita": "5-6 parole chiave ", "nor": "5-6 nøkkelord ", "ger": "5-6 Schlüsselwörter ", "dan": "5-6 Søgeord ", "dut": "5-6 trefwoorden ", "est": "5-6 Märksõnad "}} |

Figure 24: Example event.json output (note that this is not the same as currently extracted)

# Optimized extraction from an Overleaf LaTeX project

In the case of an Overleaf project, when compiling the LaTeX as a side-effect a fordiva.json file is generated, if you save this file locally then you can clean it up and make a MODS file with the following commands:

./cleanup\_pseudo\_JSON-from\_LaTeX.py --json xxx\_fordiva.json --acronyms acronyms.tex

./JSON\_to\_MODS.py --json xxx\_fordiva.json

Now all you have to do is rename the XML file that was produced to xxx.mods and you are all set to upload it into DiVA!

To find the fordiva.json file, look for the “Other logs and files” button shown at the bottom of the log output window. This button is shown in Figure 25. After clicking this button you will see a list of log and other files, such as shown in Figure 26.

|  |
| --- |
|  |

Figure 25: Other logs and files button

The details for how the fordiva.json file is generated are given in Section 24.

|  |
| --- |
|  |

Figure 26: List of log and other files

# What can we do with the JSON file?

Now that you have a JSON file, you can edit the HTML in the abstracts to deal with equations and things that were not automatically processed by the extraction program.

Once you are happy with the JSON file’s contents, the next step is to generate something interesting with this data – for this we use the program JSON\_to\_calendar.py.

|  |
| --- |
| JSON\_to\_calendar.py  KTH calendar event  Canvas calendar event  Canvas announcement |

Figure 27: The several outputs of JSON\_to\_calendar.py

We can produce all three outputs with the command shown in Figure 28. Note that the program was run with the event.json file to produce a Canvas course announcement, as shown in Figure 29. Note that this is being run in the Canvas test instance (hence the pink bar across the bottom of the figure).

|  |
| --- |
| JSON\_to\_calendar.py -c 11 --config config-test.json |

Figure 28: Running JSON\_to\_calendar.py to produce all three outputs

This same basic mechanism was extened to make and apply the cover for a thesis. There is still a need for a tool that can automate the insertion of the final thesis and metradata. While a prototype was shown earlier for this, the IT unit wants to wait for a DiVA API from the DiVA organization.

|  |
| --- |
|  |

Figure 29: Canvas course announcement

Figure 29 and Figure 30 show the event in the Canvas calendar (when I have selected to display the events for course 11 in green).

|  |
| --- |
|  |

Figure 30: A course event in the Canvas calendar (the figure is zoomed in on 15 March 2021

|  |
| --- |
|  |

Figure 31: Zoomed view of the opened Canvas calendar event

I also edited the event.json to create an event on the following date. The result is two Calendar events as shown in KTH’s Cortina calendar (in this case, it is in the development version of the Polopoly web – as this is the only place where I can use the as of yet unreleased Cortina API which is being developed by KTH’s IT unit. Figure 33 and Figure 34 show the English and Swedish versions of the event in the calendar.s

|  |
| --- |
|  |

Figure 32: KTH’s Cortina calendar showing two degree project events

|  |
| --- |
|  |

Figure 33: English version of the calendar

|  |
| --- |
|  |

Figure 34: Swedish version of the calendar event

Figure 35 shows the response from doing a POST to the KTH Cortina API. Note that this is a prototype and as of the date of my experiments did not yet support having an examiner in a calendar event (hence I had to save and remove this element of the dict before passing the data to the API, then I restored this element for use by the subsequent routines).

|  |
| --- |
| {  "advisor": "A. Busy Supervisor & Another Busy Supervisor",  "contentId": "1.1010375",  "contentName": {  "en\_GB": "This is the title in the language of the thesis: An subtitle in the language of the thesis",  "sv\_SE": "Detta är den svenska översättningen av titeln: Detta är den svenska översättningen av undertiteln"  },  "dates\_endtime": "2021-03-16T13:00:00.000Z",  "dates\_starttime": "2021-03-16T12:00:00.000Z",  "lead": {  "en\_GB": "Bachelor's thesis presentation",  "sv\_SE": "Kandidate Examensarbete presentation"  },  "lecturer": "Fake A. Student & Fake B. Student",  "location": "via Zoom",  "opponent": "A. B. Normal & A. X. E. Normalè",  "organisation": {  "school": "EECS",  "department": "Datavetenskap"  },  "respondent": "",  "respondentDepartment": "",  "subjectarea": {  "en\_GB": "Canvas Learning Management System, Docker containers, performance tuning ",  "sv\_SE": "Canvas Lärplattform,Dockerbehållare, prestandajustering "  },  "seminartype": "thesis",  "paragraphs\_text": {  "en\_GB": "<p>All theses at KTH are required to have an abstract in both <i>English</i> and <i>Swedish</i>.</p> \n<p>Exchange students many want to include one or more abstracts in the language(s) used in their home institutions to avoid the need to write another thesis when returning to their home institution.</p> \n<p>Keep in mind that most of your potential readers are only going to read your title and abstract. This is why it is important that the abstract give them enough information that they can decide is this document relevant to them or not. Otherwise the likely default choice is to ignore the rest of your document.</p> \n<p>A abstract should stand on its own, i.e., no citations, cross references to the body of the document, acronyms must be spelled out, … .</p> \n<p>Write this early and revise as necessary. This will help keep you focused on what you are trying to do.</p> \n<p>Example of a formula in an abstract: $c=2 \\cdot \\pi \\cdot r$ or \\[ \\int\_{a}^{b} x^2 \\,dx \\] two chemical formulas: H<sub>2</sub>O or $(C\_5O\_2H\_8)\_n$, copyright symbol: © Maguire 2021, and some superscripts: <sup>99m</sup>Tc, A<sup>\*</sup>, A<sup>®</sup>, and A™.</p> \n<p>Write an abstract with the following components: </p> \n<ul> \n <li> What is the topic area? (optional) Introduces the subject area for the project. </li> \n <li> Short problem statement </li> \n <li> Why was this problem worth a ’Masters thesis project? (i.e., why is the problem both significant and of a suitable degree of difficulty for a ’Masters thesis project? Why has no one else solved it yet?) </li> \n <li> How did you solve the problem? What was your method/insight? </li> \n <li> Results/Conclusions/Consequences/Impact: What are your key results/conclusions? What will others do based upon your results? What can be done now that you have finished - that could not be done before your thesis project was completed?</li> \n</ul>\n",  "sv\_SE": "<p>Alla avhandlingar vid KTH måste ha ett abstrakt på både engelska och svenska.</p> \n<p>If you are writing your thesis in English, you can leave this until the final version. If you are writing your thesis in Swedish then this should be done first – and you should revise as necessary along the way.</p> \n<p>If you are writing your thesis in English, then this section can be a summary targeted at a more general reader. However, if you are writing your thesis in Swedish, then the reverse is true – your abstract should be for your target audience, while an English summary can be written targeted at a more general audience.</p> \n<p>This means that the English abstract and Swedish sammnfattning or Swedish abstract and English summary need not be literal translations of each other.</p> \n<p>The abstract in the language used for the thesis should be the first abstract, while the Summary/Sammanfattning in the other language can follow.</p>\n"  },  "uri": "https://www.kth.se"  } |

Figure 35: Response from the KTH Cortina API

# Actual example

This example appear with the student’s permission. Figure 36shows the announcement for a 2nd cycle thesis presentation in a Canvas course while Figure 37 shows the bottom part of the announcement. Figure 39 shows the Cortina calendar entry. Note that the Cortina calendar entry is in the development system and not the production calendar. Figure 40shows the command to extract the JSON information from the student’s PDF file and then the command to make the announcement and calendar entries. Figure 41 shows the extracted JSON (edited for appearance here).

Note that the entry was made on 2021-06-17, but the event was earlier. This entry was made with the new (as of this date) KTH Cortina Calendar API that supports the examiner and language of presentation fields. Additionally, it returns a canonicalUrl (the URL to this calendar entry). The program was extened to add this URL to the course announcement and course calendar entry.

|  |
| --- |
|  |

Figure 36: Actual example of announcement in Canvas (top)

|  |
| --- |
|  |

Figure 37: Bottom of the announcement

|  |
| --- |
|  |

Figure 38: Opened version in course calendar

|  |  |
| --- | --- |
|  |  |

Figure 39: The English (left) and Swedis (right) in the Cortina calendar

|  |
| --- |
| ./extract\_pseudo\_JSON-from\_PDF.py --pdf oscar.pdf --json oscar.json  ./JSON\_to\_calendar.py -c 11 --config config-test.json --json oscar.json |

Figure 40: Commands to extract the JSON and to make the calendar entries and announcements

|  |
| --- |
| {"Author1": {"Last name": "Rosquist", "First name": "Oscar", "Local User Id": "u1tmg8l6", "E-mail": "oscarros@kth.se", "organisation": {"L1": "School of Electrical Engineering and Computer Science "}}, "Degree": {"Educational program": "Degree Programme in Computer Science and Engineering"}, "Title": {"Main title": "Adapting to the new remote work era", "Subtitle": "Improving social well-being among IT remote workers", "Language": "eng"}, "Alternative title": {"Main title": "Anpassningar för den digitala arbetsplatsen", "Subtitle": "Förbättringar av det sociala välmåendet hos distansarbete inom IT", "Language": "swe"}, "Supervisor1": {"Last name": "Västberg", "First name": "Anders", "Local User Id": "u1ft3a12", "E-mail": "vastberg@kth.se", "organisation": {"L1": "School of Electrical Engineering and Computer Science ", "L2": "Computer Science"}}, "Supervisor2": {"Last name": "XXXXX", "First name": "XXXXX", "E-mail": "XXXXXX", "Other organisation": "XXXXX"}, "Examiner1": {"Last name": "Maguire Jr.", "First name": "Gerald Q.", "Local User Id": "u1d13i2c", "E-mail": "maguire@kth.se", "organisation": {"L1": "School of Electrical Engineering and Computer Science ", "L2": "Computer Science"}}, "Cooperation": {"Partner\_name": "XXXXX"}, "Other information": {"Year": "2021", "Number of pages": "xvii,115"}, "Opponents": {"Name": "XXXXXX"}, "Presentation": {"Date": "2021-05-31 13:00", "Language": "eng", "Room": "via Zoom https://kth-se.zoom.us/j/61957949263", "Address": "Isafjordsgatan 22 (Kistagången 16)", "City": "Stockholm", "National Subject Categories": "10201, 10206"}, "Number of lang instances": "2", "abstracts": {"eng": "<p>In 2020, the world was struck by the Covid-19 pandemic. …negative effects.</p>", "swe": "<p>Under början av året 2020 drabbades världen av Covid-19 pandemin. … ett liknande experiment.</p>"}, "keywords": {"eng": "\n\fAppendix B: Surveys results | 117\n\nRemote work, Work from home, Social well-being, Covid-19 pandemic, Digital social interactions, Information and Communication\nTechnologies ", "swe": "Distansarbete, Hemarbete, Socialt välmående, Covid-19 pandemin, Digitala sociala interaktioner, Informations- och kommunikationsteknik\n"}} |

Figure 41: Extracted JSON file oscar.json - edited for appearance here

# Change in how to enter the abstracts in LaTeX

In order to deal with both babl and Polyglossia and both bibtex and biblatex, I have changed how the abstracts should be entered. Basically the idea is to insert a \babelpolyLangStart{language\_name} before the start of the abstract and \babelpolyLangStop after the end of the abstract. These macros hide the difference between using Babel or Polyglossia. Additionally, they avoid the problem of the Overleaf GUI being confused about matching beginning and ending statements. Figure 42 and Figure 43 show examples of how to enter an abstract and keywords while Figure 44 shows the defition of the two macros.

Note that both Babel and Polyglossia expand the \abstractname into the correct version of the name for an abstract in the current language. Unfortunately, neither package has an equivalent to provide the language specific version of “keywords”, so these have to be provided by the person entering the keywords.

|  |
| --- |
| \babelpolyLangStart{swedish}  \begin{abstract}  \markboth{\abstractname}{}  \begin{scontents}[store-env=lang]  swe  \end{scontents}  \begin{scontents}[store-env=abstracts,print-env=true]  Alla avhandlingar vid KTH måste ha ett abstrakt på både engelska och svenska.  Om du skriver din avhandling på svenska ska detta göras först (och placera det som det första abstraktet) - och du bör revidera det vid behov.  If you are writing your thesis in English, you can leave this until the draft version that goes to your opponent for the written opposition. In this way you can provide the English and Swedish abstract/summary information that can be used in the announcement for your oral presentation.  If you are writing your thesis in English, then this section can be a summary targeted at a more general reader. However, if you are writing your thesis in Swedish, then the reverse is true – your abstract should be for your target audience, while an English summary can be written targeted at a more general audience.  This means that the English abstract and Swedish sammnfattning  or Swedish abstract and English summary need not be literal translations of each other.  The abstract in the language used for the thesis should be the first abstract, while the Summary/Sammanfattning in the other language can follow.  \end{scontents}  \subsection\*{Nyckelord}  \begin{scontents}[store-env=keywords,print-env=true]  Canvas Lärplattform,Dockerbehållare, prestandajustering  \end{scontents}  \end{abstract}  \babelpolyLangStop{swedish} |

Figure 42: Example of the revised format for entering an abstract

|  |
| --- |
| \todo[inline]{Use the relevant language for abstracts for your home university.\\  Note that you may need to augment the set of language used in polyglossia or  babel (see the file kththesis.cls). The following languages include those languages that were used in theses at KTH in 2018-2019, except for one in Chinese.\\  Remove those versions that you do not need.\\  If adding a new language, when specifying the language for the abstract use the three letter ISO 639-2 Code – specifically the "B" (bibliographic) variant of these codes (note that this is the same language code used in DiVA).  \babelpolyLangStart{french}  \begin{abstract}  \markboth{\abstractname}{}  \begin{scontents}[store-env=lang]  fre  \end{scontents}  \begin{scontents}[store-env=abstracts,print-env=true]  Résumé en français.  \end{scontents}  \subsection\*{Mots clés}  \begin{scontents}[store-env=keywords,print-env=true]  5-6 mots-clés  \end{scontents}  \end{abstract}  \babelpolyLangStop{french}  \cleardoublepage |

Figure 43: Second example of the revised format for entering an abstract

|  |
| --- |
| \ifxeorlua  \newcommand{\babelpolyLangStop}[1]{\end{#1}}  \else  \newcommand{\babelpolyLangStop}[1]{\end{otherlanguage}}  \fi  \ifxeorlua  \newcommand{\babelpolyLangStart}[1]{\begin{#1}}  \else  \newcommand{\babelpolyLangStart}[1]{\begin{otherlanguage}{#1}}  \fi |

Figure 44: The two macros used to help enter the language specification

# Adding keywords and PDF meta data

In an effort to add the PDF meta via the hyperref package, I also decided to add the keywords part of the PDF meta. However, in order to do this I had to have the keywords before the \begin{document} command in the LaTeX file. To do so, I added three new commands to the kththesis.cls file, as shown in Figure 45. The commands are used in the examplethesis.tex file to set up the keywords in both English and Swedeish as well as include a new set of LaTeX commands to store the PDF meta data (as shown in Figure 46) using a file called lib/pdf\_related\_includes.tex (shown in Figure 47). Later the keywords that have been stored are inserted into the LaTeX after their respective language abstracts as shown in Figure 48 and Figure 49. The title page of the thesis and the PDF meta data are shown in Figure 50 Finally, the keywords appear (as expected) in the for DiVA data at the end of the PDF file as shown in Figure 51.

Note that \makeatletter and \makeatother are use to include the character “@” as a letter and then return “@” to being a punction code. This use of “@” protects the internal names from being accessed outside of these two commands. More explicitly, \EnglishKeywords is a function that takes one argument, the text of the English keywords, and then stores them into “@EnglishKeywords”. Later the text can be retried with the command \InsertKeywords{english} or \InsertKeywords{swedish}.

|  |
| --- |
| % Keywords  \let\@EnglishKeywords\@empty  \newcommand{\EnglishKeywords}[1]{\def\@EnglishKeywords{#1}}  \let\@SwedishKeywords\@empty  \newcommand{\SwedishKeywords}[1]{\def\@SwedishKeywords{#1}}  \makeatletter  \newcommand{\InsertKeywords}[1]{  \IfEqCase{#1}{%  {english}{\@EnglishKeywords}  {swedish}{\@SwedishKeywords}  }[\typeout{argument must be english or swedish}]  } |

Figure 45: New commands in kththesis.cls

Figure 46 shows the storing of the keywords using the above commands and the include of the library to set up the PDF meta data.

|  |
| --- |
| % Enter the English and Swedish keywords here for use in the PDF meta data \_and\_ for later use  % following the respective abstract.  % Try to put the words in the same order in both to facilitate matching.  \EnglishKeywords{Canvas Learning Management System, Docker containers, performance tuning}  \SwedishKeywords{Canvas Lärplattform, Dockerbehållare, prestandajustering}  % Put the title, author, and keyword information into the PDF meta information  \include{lib/pdf\_related\_includes} |

Figure 46: New additions to examplethesis.text

The lib/pdf\_related\_includes.tex file contains the LaTeX to add information to the PDF file (specifically, author(s), title(s), and keywords. It uses the hyperref package and should be be included before the \begin{document}.

I want to acknowledge the inspiration of Karl Voit's template for TU Graz that inspired me to add the PDF document information. For more information about his template see <https://github.com/novoid/LaTeX-KOMA-template>

Note that this template does not use anything from his template other than the names of the information for the PDF meta fields, i.e., mytitle, myauthor, and mykeywords together with the idea of defining the corresponding newcommand to set the relevant hyperref parameters. A result is that these command names are visible to the rest of the LaTeX file.

|  |
| --- |
| \makeatletter  \ifx\@subtitle\@empty  \newcommand{\mytitle}{\@title}  \else  \newcommand{\mytitle}{\@title: \@subtitle}  \fi  \hypersetup{  pdftitle={\mytitle} % Title field  }  \ifx\@secondAuthorsLastname\@empty  \newcommand{\myauthor}{\@authorsFirstname \@authorsLastname}  \else  \ifinswedish  \newcommand{\myauthor}{\@authorsFirstname\space\@authorsLastname\space\relax och\space\@secondAuthorsFirstname \@secondAuthorsLastname}  \else  \newcommand{\myauthor}{\@authorsFirstname\space\@authorsLastname\space\relax and\space\@secondAuthorsFirstname \@secondAuthorsLastname}  \fi  \fi  \hypersetup{  pdfauthor={\myauthor} % Author field  }  % Put the alternative title (and subtitle) into the PDF Subject meta  \ifx\@altsubtitle\@empty\relax  \newcommand{\myalttitle}{\@alttitle}  \else  \newcommand{\myalttitle}{\@alttitle: \@altsubtitle}  \fi  \hypersetup{  pdfsubject={\myalttitle} % Subject field  }  \ifx\@EnglishKeywords\@empty  \ifx\@SwedishKeywords\@empty  \newcommand{\mykeywords}{}  \else  \newcommand{\mykeywords}{\@SwedishKeywords}  \fi  \else  \ifx\@SwedishKeywords\@empty  \newcommand{\mykeywords}{\@EnglishKeywords}  \else  \ifinswedish  \newcommand{\mykeywords}{\@SwedishKeywords, \@EnglishKeywords}  \else  \newcommand{\mykeywords}{\@EnglishKeywords, \@SwedishKeywords}  \fi  \fi  \fi  \makeatother  \hypersetup{  pdfkeywords={\mykeywords} % Keywords field  } |

Figure 47: lib/pdf\_related\_includes.tex (edited for readability)

|  |
| --- |
| \subsection\*{Keywords}  \begin{scontents}[store-env=keywords,print-env=true]  % If you set the EnglishKeywords earlier, you can retrieve them with:  % **Alternative 1:**  %\makeatletter  %\@EnglishKeywords  %\makeatother  %  % **Alternative 2:**  \InsertKeywords{english}  % If you did not set the EnglishKeywords earlier then simply enter the keywords here:  % **Alternative 3:**  % comma separate keywords, such as: Canvas Learning Management System, Docker containers, performance tuning  \end{scontents} |

Figure 48: Including the English language keywords below the English abstract

|  |
| --- |
| \subsection\*{Nyckelord}  \begin{scontents}[store-env=keywords,print-env=true]  % SwedishKeywords were set earlier, hence we can use alternative 2  \InsertKeywords{swedish}  \end{scontents}  \end{abstract} |

Figure 49: Including the Swedeish language keywords below the Swedish abstract

|  |
| --- |
|  |

Figure 50: The title page of the thesis and the PDF meta data

|  |
| --- |
|  |

Figure 51: The keywords appear as expected in the for DiVA data at the end of the PDF file

# Other variants of the JSON\_to\_calendar.py

For testing purposes, I also created functionality in JSON\_to\_calendar.py to insert a fixed event (this was my first test) and take in a MODS file. The MODS file was created from a DiVA feed of theses presented in 2020 through to the 25th of April. However, one limitation that I found is that other than myself, few people have been entering the date and time for the oral presentation. Since I wanted to test making calendar announcements, I needed data and time!

**NB**: I have assumed that each degree project presentation lasts one hour – since the KTH Cortina calendar needs both a starting and ending time.

These other variants probably should not be kept, but rather the architecture should be similar to that shown in Figure 52. Additionally, when taking data from other types of sources, one can take advantage of the data that is in a Canvas degree project course to help out the processing of the source data.

|  |
| --- |
| JSON  Extractor  JSON\_to\_calendar.py  Canvas announcement  Canvas calendar event  KTH calendar event  LaTeX  DOCX  PDF |

Figure 52: Several possible inputs to JSON\_to\_calendar.py and its outputs

# JSON to MODS file

In keeping with the idea of using the JSON file to drive other applications, I have added a program JSON\_to\_MODS.py to create a MODS file using the information from the arguments and a JSON file.

The program is patterned after the JSON\_to\_calendar and JSON\_to\_cover program. The program outputs a MODS file called: MODS.pdf.

|  |
| --- |
| ./JSON\_to\_MODS.py -c 11 --json jussi.json --trita "TRITA-EECS-EX-2021:219" --testing  or  ./JSON\_to\_MODS.py -c 11 --json test12.json --trita "TRITA-EECS-EX-2021:219" --testing |

Figure 53: Examples of using JSON\_to\_MODS.py

Note that currently the Canvas course information is not used. Note also that the “—testing” flag foces the report series to be "TRITA-ICT-EX" since the actual series in in the earlier version of DiVA.

If acronyms are used in the abstracts, you can also add the “- -acronyms acronyms.tex” aregument to the command line and the program will process the acronyms. You will get a file named modsXML.xml that you can inport into DiVA. I seem to be able to import these files into the test instance of DiVA: <https://kth.test.diva-portal.org/dream/import/importList.jsf>

This program has been extened to be able to take the TRITA number information from the MODS file.

## Example

The process of importing the MODS data into DiVA begins with the extracted JSON file (note that I have changed some of the formatting of the pseudo JSON information and added the export of the program code). An example of the JSON information is shown in Figure 54. The resulting MODS file is shown in Figure 55.

|  |
| --- |
| {"Author1": {"Last name": "Student", "First name": "Fake A.", "Local User Id": "u100001", "E-mail": "a@kth.se", "ORCiD": "0000-0002-00001-1234", "organisation": {"L1": "School of Electrical Engineering and Computer Science "}}, "Author2": {"Last name": "Student", "First name": "Fake B.", "Local User Id": "u100002", "E-mail": "b@kth.se", "ORCiD": "0000-0002-00001-5678", "organisation": {"L1": "School of Architecture and the Built Environment "}}, "Degree": {"Educational program": "Bachelor’s Programme in Information and Communication Technology", "programcode": "TCOMK", "Level": "1", "Course code": "IA150X", "Credits": "15.0", "Exam": "Bachelors degree", "subjectArea": "Information and Communication Technology"}, "Title": {"Main title": "This is the title in the language of the thesis", "Subtitle": "An subtitle in the language of the thesis", "Language": "eng"}, "Alternative title": {"Main title": "Detta är den svenska översättningen av titeln", "Subtitle": "Detta är den svenska översättningen av undertiteln", "Language": "swe"}, "Supervisor1": {"Last name": "Supervisor", "First name": "A. Busy", "Local User Id": "u100003", "E-mail": "sa@kth.se", "organisation": {"L1": "School of Electrical Engineering and Computer Science ", "L2": "Computer Science"}}, "Supervisor2": {"Last name": "Supervisor", "First name": "Another Busy", "Local User Id": "u100003", "E-mail": "sb@kth.se", "organisation": {"L1": "School of Architecture and the Built Environment ", "L2": "Public Buildings"}}, "Supervisor3": {"Last name": "Supervisor", "First name": "Third Busy", "E-mail": "sc@tu.va", "Other organisation": "Timbuktu University, Department of Pseudoscience"}, "Examiner1": {"Last name": "Maguire Jr.", "First name": "Gerald Q.", "Local User Id": "u1d13i2c", "E-mail": "maguire@kth.se", "organisation": {"L1": "School of Electrical Engineering and Computer Science ", "L2": "Computer Science"}}, "Cooperation": {"Partner\_name": "Företaget AB"}, "National Subject Categories": "10201, 10206", "Other information": {"Year": "2021", "Number of pages": "xxxiii,35"}, "Opponents": {"Name": "A. B. Normal & A. X. E. Normalè"}, "Presentation": {"Date": "2021-03-15 13:00", "Language": "eng", "Room": "via Zoom https://kth-se.zoom.us/j/ddddddddddd", "Address": "Isafjordsgatan 22 (Kistagången 16)", "City": "Stockholm"}, "Number of lang instances": "10", "abstracts": {"eng": "<p>All theses at KTH are <bold>required</bold> to have an abstract in both <i>English</i> and <i>Swedish</i>.</p><p>Exchange students many want to include one or more abstracts in the language(s) used in their home institutions to avoid the need to write another thesis when returning to their home institution.</p><p>Keep in mind that most of your potential readers are only going to read your <tt>title</tt> and <tt>abstract</tt>. This is why it is important that the abstract give them enough information that they can decide is this document relevant to them or not. Otherwise the likely default choice is to ignore the rest of your document.</p><p>A abstract should stand on its own, i.e., no citations, cross references to the body of the document, acronyms must be spelled out, … .</p><p>Write this early and revise as necessary. This will help keep you focused on what you are trying to do.</p><p>Write an abstract with the following components: </p><ul><li> What is the topic area? (optional) Introduces the subject area for the project. </li><li> Short problem statement </li><li> Why was this problem worth a Bachelor’s/’Masters thesis project? (i.e., why is the problem both significant and of a suitable degree of difficulty for a Bachelor’s/’Masters thesis project? Why has no one else solved it yet?) </li><li> How did you solve the problem? What was your method/insight? </li><li> Results/Conclusions/Consequences/Impact: What are your key results/conclusions? What will others do based upon your results? What can be done now that you have finished - that could not be done before your thesis project was completed?</li></ul><p>announcement of the oral presentation and for entering data into DiVA</p><p>\\pi \\cdot r$ or \\[ \\int\_{a}^{b} x^2 \\,dx \\]</p><p></p><p> A<sup>\*</sup>, A<sup>&reg;</sup>, and A&trade;.</p><p> first chemical formula can be handled, while the second will require hand editing</p>", "swe": "<p>Alla avhandlingar vid KTH <bold>måste ha</bold> ett abstrakt på både <i>engelska</i> och <i>svenska</i>.</p><p>Om du skriver din avhandling på svenska ska detta göras först (och placera det som det första abstraktet) - och du bör revidera det vid behov.</p><p>If you are writing your thesis in English, you can leave this until the draft version that goes to your opponent for the written opposition. In this way you can provide the English and Swedish abstract/summary information that can be used in the announcement for your oral presentation.</p><p>If you are writing your thesis in English, then this section can be a summary targeted at a more general reader. However, if you are writing your thesis in Swedish, then the reverse is true – your abstract should be for your target audience, while an English summary can be written targeted at a more general audience.</p><p>This means that the English abstract and Swedish sammnfattning or Swedish abstract and English summary need not be literal translations of each other.</p><p>The abstract in the language used for the thesis should be the first abstract, while the Summary/Sammanfattning in the other language can follow.</p>", "fre": "<p>Résumé en français.</p>", "spa": "<p>Résumé en espagnol.</p>", "ita": "<p>Sommario in italiano.</p>", "nor": "<p>Sammendrag på norsk.</p>", "ger": "", "dan": "<p>Abstrakt på dansk.</p>", "dut": "<p>Zusammenfassung in Deutsch.</p><p>Samenvatting in het Nederlands.</p><p>Eesti keeles kokkuvõte.</p>", "est": ""}, "keywords": {"eng": "Canvas Learning Management System, Docker containers, Performance tuning ", "swe": "Canvas Lärplattform, Dockerbehållare, Prestandajustering\nNyckelord som beskriver innehållet i uppsatsrapporten\n", "fre": "5-6 mots-clés ", "spa": "5-6 Palabras claves ", "ita": "5-6 parole chiave ", "nor": "5-6 nøkkelord ", "ger": "5-6 Schlüsselwörter ", "dan": "5-6 Søgeord ", "dut": "5-6 trefwoorden ", "est": "5-6 Märksõnad "}} |

Figure 54: test12.json

|  |
| --- |
| {"Author1": {"Last name": "Student", "First name": "Fake A.", "Local User Id": "u100001", "E-mail": "a@kth.se", "ORCiD": "0000-0002-00001-1234", "organisation": {"L1": "School of Electrical Engineering and Computer Science "}}, "Author2": {"Last name": "Student", "First name": "Fake B.", "Local User Id": "u100002", "E-mail": "b@kth.se", "ORCiD": "0000-0002-00001-5678", "organisation": {"L1": "School of Architecture and the Built Environment "}}, "Degree": {"Educational program": "Bachelor’s Programme in Information and Communication Technology", "programcode": "TCOMK", "Level": "1", "Course code": "IA150X", "Credits": "15.0", "Exam": "Bachelors degree", "subjectArea": "Information and Communication Technology"}, "Title": {"Main title": "This is the title in the language of the thesis", "Subtitle": "An subtitle in the language of the thesis", "Language": "eng"}, "Alternative title": {"Main title": "Detta är den svenska översättningen av titeln", "Subtitle": "Detta är den svenska översättningen av undertiteln", "Language": "swe"}, "Supervisor1": {"Last name": "Supervisor", "First name": "A. Busy", "Local User Id": "u100003", "E-mail": "sa@kth.se", "organisation": {"L1": "School of Electrical Engineering and Computer Science ", "L2": "Computer Science"}}, "Supervisor2": {"Last name": "Supervisor", "First name": "Another Busy", "Local User Id": "u100003", "E-mail": "sb@kth.se", "organisation": {"L1": "School of Architecture and the Built Environment ", "L2": "Public Buildings"}}, "Supervisor3": {"Last name": "Supervisor", "First name": "Third Busy", "E-mail": "sc@tu.va", "Other organisation": "Timbuktu University, Department of Pseudoscience"}, "Examiner1": {"Last name": "Maguire Jr.", "First name": "Gerald Q.", "Local User Id": "u1d13i2c", "E-mail": "maguire@kth.se", "organisation": {"L1": "School of Electrical Engineering and Computer Science ", "L2": "Computer Science"}}, "Cooperation": {"Partner\_name": "Företaget AB"}, "National Subject Categories": "10201, 10206", "Other information": {"Year": "2021", "Number of pages": "xxxiii,35"}, "Opponents": {"Name": "A. B. Normal & A. X. E. Normalè"}, "Presentation": {"Date": "2021-03-15 13:00", "Language": "eng", "Room": "via Zoom https://kth-se.zoom.us/j/ddddddddddd", "Address": "Isafjordsgatan 22 (Kistagången 16)", "City": "Stockholm"}, "Number of lang instances": "10", "abstracts": {"eng": "<p>All theses at KTH are <bold>required</bold> to have an abstract in both <i>English</i> and <i>Swedish</i>.</p><p>Exchange students many want to include one or more abstracts in the language(s) used in their home institutions to avoid the need to write another thesis when returning to their home institution.</p><p>Keep in mind that most of your potential readers are only going to read your <tt>title</tt> and <tt>abstract</tt>. This is why it is important that the abstract give them enough information that they can decide is this document relevant to them or not. Otherwise the likely default choice is to ignore the rest of your document.</p><p>A abstract should stand on its own, i.e., no citations, cross references to the body of the document, acronyms must be spelled out, … .</p><p>Write this early and revise as necessary. This will help keep you focused on what you are trying to do.</p><p>Write an abstract with the following components: </p><ul><li> What is the topic area? (optional) Introduces the subject area for the project. </li><li> Short problem statement </li><li> Why was this problem worth a Bachelor’s/’Masters thesis project? (i.e., why is the problem both significant and of a suitable degree of difficulty for a Bachelor’s/’Masters thesis project? Why has no one else solved it yet?) </li><li> How did you solve the problem? What was your method/insight? </li><li> Results/Conclusions/Consequences/Impact: What are your key results/conclusions? What will others do based upon your results? What can be done now that you have finished - that could not be done before your thesis project was completed?</li></ul><p>announcement of the oral presentation and for entering data into DiVA</p><p>\\pi \\cdot r$ or \\[ \\int\_{a}^{b} x^2 \\,dx \\]</p><p></p><p> A<sup>\*</sup>, A<sup>&reg;</sup>, and A&trade;.</p><p> first chemical formula can be handled, while the second will require hand editing</p>", "swe": "<p>Alla avhandlingar vid KTH <bold>måste ha</bold> ett abstrakt på både <i>engelska</i> och <i>svenska</i>.</p><p>Om du skriver din avhandling på svenska ska detta göras först (och placera det som det första abstraktet) - och du bör revidera det vid behov.</p><p>If you are writing your thesis in English, you can leave this until the draft version that goes to your opponent for the written opposition. In this way you can provide the English and Swedish abstract/summary information that can be used in the announcement for your oral presentation.</p><p>If you are writing your thesis in English, then this section can be a summary targeted at a more general reader. However, if you are writing your thesis in Swedish, then the reverse is true – your abstract should be for your target audience, while an English summary can be written targeted at a more general audience.</p><p>This means that the English abstract and Swedish sammnfattning or Swedish abstract and English summary need not be literal translations of each other.</p><p>The abstract in the language used for the thesis should be the first abstract, while the Summary/Sammanfattning in the other language can follow.</p>", "fre": "<p>Résumé en français.</p>", "spa": "<p>Résumé en espagnol.</p>", "ita": "<p>Sommario in italiano.</p>", "nor": "<p>Sammendrag på norsk.</p>", "ger": "", "dan": "<p>Abstrakt på dansk.</p>", "dut": "<p>Zusammenfassung in Deutsch.</p><p>Samenvatting in het Nederlands.</p><p>Eesti keeles kokkuvõte.</p>", "est": ""}, "keywords": {"eng": "Canvas Learning Management System, Docker containers, Performance tuning ", "swe": "Canvas Lärplattform, Dockerbehållare, Prestandajustering\nNyckelord som beskriver innehållet i uppsatsrapporten\n", "fre": "5-6 mots-clés ", "spa": "5-6 Palabras claves ", "ita": "5-6 parole chiave ", "nor": "5-6 nøkkelord ", "ger": "5-6 Schlüsselwörter ", "dan": "5-6 Søgeord ", "dut": "5-6 trefwoorden ", "est": "5-6 Märksõnad "}} |

Figure 55: test12.mods

Figure 56 to Figure 59 show the process of importing the MODS file into DiVA. Note that it skips the first user interface form (as this differs between users).

|  |
| --- |
|  |

Figure 56: The import process – step 1 – click on the “Import from external database” buttom on the upper right corner

|  |
| --- |
|  |

Figure 57: The import process – step 2 – choose the MODV3 format and select a file to import

|  |
| --- |
|  |

Figure 58: The import process – step 3 – after clicking “Import” – it says that the file was successfully uploaded

Click on “Import” again to load the file. The MODS files is loaded and shown in Figure 59.

|  |
| --- |
|  |

Figure 59: The import process – step 4 – after clicking “Import” once more – you can see the file has been added to the list of imported files

The imported document is shown in Figure 60 to Figure 67.

|  |
| --- |
|  |

Figure 60: Imported document (part 1)

|  |
| --- |
|  |

Figure 61: Imported document (part 2) – Note that both the English and Swedish titles are shown as well as the information about which level, the number of university credits, the education al program, and the subject.- Additionally, the year and number of pages are shown.

|  |
| --- |
|  |

Figure 62: Imported document (part 3) – Note that the number in the series is not being shown. The nation subject catergies specified in the JSON file are shown. Additionally, a number of the keyoards in different languages are shown.

|  |
| --- |
|  |

Figure 63: Imported document (part 4) – more keywords are shown as well as the abstract in English

|  |
| --- |
|  |

Figure 64: Imported document (part 5) – the bottom of the Engish abstract is shown .- showing some of the formatted information that can be shown. Additionally, the Swedish and other abstracts are shown.

|  |
| --- |
|  |

Figure 65: Imported document (part 6) – yet more abstracts

|  |
| --- |
|  |

Figure 66: Imported document (part 7) – the supervisors are shown

|  |
| --- |
|  |

Figure 67: Imported document (part 8) – the examiner is shown along with the information about the oral presentation

## Limitations

I was initially unable to get the KTH affiliations correctly entered. I think that it is because I did not understand the description of how to do this at: <https://wiki.epc.ub.uu.se/pages/viewpage.action?pageId=27466001> where it says:

För att göra en sådan koppling skall dels ”affiliation” finnas i personelementet dels ett name-element med ID i xlink:href samt ett namePart-element med samma namn som affiliation. Om man har bägge nedanstående element i exemplet kommer Urban Ericsson automatiskt att kopplas till Universitetsbiblioteket i Uppsala vid importen. Om enbart personelementet finns men inte det andra kommer Urban Ericsson att få Uppsala universitet, Universitetsbiblioteket som ”Annat lärosäte” vid importen.

I did not find any examples of how to do what they describe. The results is that all of my affiliations information ended up in the "Other university" field.

I was also unable to get the issue number to appear, even those it seems to be set - since if you click on the box after the MODS import, it knows the value. Note that I force the TRITA series to be "TRITA-ICT-EX" since the actual series in in the earlier version of DiVA.

Finally, it seems that DiVA does not import the cooperation data at all: according to <https://wiki.epc.ub.uu.se/display/divainfo/Externt+samarbete>

**The first two limitations have been overcome as described in Section 16.3.**

## Limitations overcome

This section describes how two of the limitations were overcome.

### Entering number in series

The first limitation to be over come was entering the particular number in the series for a TRITA number. Figure 68 shows the value being entered. The error was in the structure of the object, see Figure 69. My earlier error was due to having the <identifier type="issue number">2021:00</identifier> ***within*** the <titleInfo>.

|  |
| --- |
|  |

Figure 68: Number in series correctly entered

|  |
| --- |
| <relatedItem type="series">  <titleInfo><title>TRITA-EECS-EX</title>  <identifier type="local">16855</identifier>  </titleInfo>  <identifier type="issue number">2021:00</identifier>  </relatedItem > |

Figure 69: MODS data

### Including KTH affiliations

Earlier those with KTH affilations were appearing in the “Other organization” filed. The error in my understanding was how the controlled fields for the KTH authors was being used. We will start by looking at the two KTH affiliated supervisors (see Figure 70), the third supervisor (see Figure 71), and the examiner (see Figure 72). The MODS data for these four people is shown in Figure 73 and the “magic” corporate name MODS data is shown in Figure 74. The trick is that there needs to be the **corporate name MODS data for each of the possible affiliations** – as this is the way that one passes the DiVA code via the xlink information. Note that in the case of the examiner, two different affilations were specified and there was a corporate name entry for each of them. Moreover, the key is just to have test extra corporate name entries, only one needs to be for the publishers, as shown in Figure 75 – with the results in Figure 76 and shown Figure 77.

**Note**: The ICT affilations (and DiVA codes) were used and the pictures are all from the test DiVA environment: <https://kth.test.diva-portal.org/dream/import/importList.jsf>

|  |
| --- |
|  |

Figure 70: Two KTH supervisors in different schools

|  |
| --- |
|  |

Figure 71: Third supervisor

|  |
| --- |
|  |

Figure 72: Examiner

|  |
| --- |
| <name type="personal" authority="kth" xlink:href="u1d13i2c">  <namePart type="family">Maguire Jr.</namePart>  <namePart type="given">Gerald Q.</namePart>  <description>email=maguire@kth.se</description>  <affiliation>KTH, Skolan för informations- och kommunikationsteknik (ICT), Kommunikationssystem, CoS</affiliation>  <affiliation>KTH, Skolan för informations- och kommunikationsteknik (ICT)</affiliation>  <role><roleTerm type="code" authority="marcrelator">mon</roleTerm></role>  </name>  <name type="personal" authority="kth" xlink:href="u100003">  <namePart type="family">Supervisor</namePart>  <namePart type="given">A. Busy</namePart>  <affiliation>KTH, Skolan för informations- och kommunikationsteknik (ICT)</affiliation>  <description>email=sa@kth.se</description>  <role><roleTerm type="code" authority="marcrelator">ths</roleTerm></role>  </name>  <name type="personal" authority="kth" xlink:href="u100003">  <namePart type="family">Supervisor</namePart>  <namePart type="given">Another Busy</namePart>  <affiliation>KTH, Skolan för arkitektur och samhällsbyggnad (ABE)</affiliation>  <description>email=sb@kth.se</description>  <role><roleTerm type="code" authority="marcrelator">ths</roleTerm></role>  </name>  <name type="personal">  <namePart type="family">Supervisor</namePart>  <namePart type="given">Third Busy</namePart>  <affiliation>Timbuktu University, Department of Pseudoscience</affiliation>  <description>email=sc@tu.va</description>  <role><roleTerm type="code" authority="marcrelator">ths</roleTerm></role>  </name> |

Figure 73: MODS data for the supervisors and examiners

|  |
| --- |
| <name type="corporate" authority="kth" xlink:href="5994">  <namePart>KTH</namePart>  <namePart>Skolan för informations- och kommunikationsteknik (ICT)</namePart>  <role><roleTerm type="code" authority="marcrelator">pbl</roleTerm></role>  </name>  <name type="corporate" authority="kth" xlink:href="5998">  <namePart>KTH</namePart>  <namePart>Skolan för informations- och kommunikationsteknik (ICT)</namePart>  <namePart>Kommunikationssystem, CoS</namePart>  <role><roleTerm type="code" authority="marcrelator">pbl</roleTerm></role>  </name>  <name type="corporate" authority="kth" xlink:href="5850">  <namePart>KTH</namePart>  <namePart>Skolan för arkitektur och samhällsbyggnad (ABE)</namePart>  <role><roleTerm type="code" authority="marcrelator">pbl</roleTerm></role>  </name> |

Figure 74: **Corporate name** data in MODS data for the affiliations

|  |
| --- |
| <name type="corporate" authority="kth" xlink:href="5994">  <namePart>KTH</namePart>  <namePart>Skolan för informations- och kommunikationsteknik (ICT)</namePart>  <role><roleTerm type="code" authority="marcrelator">pbl</roleTerm></role>  </name>  <name type="corporate" authority="kth" xlink:href="5998">  <namePart>KTH</namePart>  <namePart>Skolan för informations- och kommunikationsteknik (ICT)</namePart>  <namePart>Kommunikationssystem, CoS</namePart>  </name>  <name type="corporate" authority="kth" xlink:href="5850">  <namePart>KTH</namePart>  <namePart>Skolan för arkitektur och samhällsbyggnad (ABE)</namePart>  </name> |

Figure 75: Reduced MODS for corporate names

|  |
| --- |
|  |

Figure 76: The two supervisors with the reduced MODS for corporate names

|  |
| --- |
|  |

Figure 77: The examiner with the reduced MODS for corporate names

# Some enhancements

The enhancements are (§17.1) to the template and supporting programs to avoid the user having to enter some information on the command line when making a cover and to make the information better available for a future DiVA entry and in Section 17.2 enhancements to better support mathematical expressions in abstracts. Section 17.3 describes support for acronyms in abstracts. Section 17.4 describes the state of support for URLs in abstracts. Section 17.5 describes a program that can extract an assignment from a Canvas course and optionally feed it to the program to extract the pseudo-JSON information.

## Enhancements to template and supporting programs

Add better support for the subject area (or areas in the case of a student doing both a Civ. Ing. Degree and a Master’s degree).

## Better support for mathematical expressions in abstracts

LaTeX in the abstracts is passed through to the pseudo JSON in the “For DIVA” text at the end of the thesis. When preparing this text for Cortina calendar entries or for the announcement in the Canvas course room (and for the calendar event in the course room) some simple LaTeX commands are converted to HTML. As few abstracts (in the many abstracts that I looked at in DiVA have equations I have only done these few transformations. However, if there were to be more use of equations, then there is probably a need to support them for the different platforms (Cortina, DiVA, and Canvas).

### Better support for mathematics in Canvas course announcement and course calendar

As of 2021-06-18, MathJAX and URLs are now supported in the Canvas course room announcement and calendar.

Using the Overleaf project: <https://www.overleaf.com/read/qsyddnhhvkgr> to provide a test source document. The results can be seen in Figure 79. These results and the results shown in Section 17.2.2 were generated using the commands in Figure 78.

|  |
| --- |
| ./extract\_pseudo\_JSON-from\_PDF.py --pdf abstracts\_with\_equations\_in\_them.pdf --json eqtest.json  ./JSON\_to\_calendar.py -c 11 --config config-test.json --json eqtest.json |

Figure 78: Commands to produce the JSON and make the calendar entries and announcement

|  |
| --- |
|  |

Figure 79: Examples of equations in an announcement

Note that block/display math are displayed in the Canvas summary for the announcements and cause Canvas to stop summarizing the abstract. The cause for this is not yet know, but has been reported to e-learning ([#ID:KTH-INC-3677258#]) and I have blogged about it in the Canvas Community. An example of the equation being displayed in the summary of the announcement is shown in Figure 80. The announcement is shown in Figure 81 while Figure 82 shows the HTML for this announcement.

|  |
| --- |
|  |

Figure 80: Announcement summary

|  |
| --- |
|  |

Figure 81: Announcement with equation

|  |
| --- |
| <h2 lang="en">Abstract</h2>  <p>All theses at KTH are required to have an abstract in both <i>English</i> and <i>Swedish</i>.</p>  <p>Exchange students many want to include one or more abstracts in the language(s) used in their home institutions to avoid the need to write another thesis when returning to their home institution.</p>  <p><span class="math-tex">\(\pi \cdot r\)</span> or <span class="math-tex">\[ \int\_{a}^{b} x^2 \,dx \]</span></p>  <p>Some more text: A<sup>\*</sup>, A<sup>&reg;</sup>, and A&trade;.</p>  <p><strong>Keywords:</strong> <em>Canvas Learning Management System, Docker containers, Performance tuning </em></p> |

Figure 82: HTML for the announcement

Via Martin Löfgren at IT-support, I raised the question about why the block equations were appearing in the summary of the announcement. Figure 83 shows the response from someone at Instructure. Indeed Instructure got the same behavior from the block equations in the snippets of the announcements. However, their solution for this is to suggest that we submit a feature request.

|  |
| --- |
| Thank you for reaching back out. I did some testing on this and I have a few things to point out and some follow up questions.  First off, the syntax of the block equation isn't the officially supported format. The inline equation is correct with the \(...\) delimiters, but the block equation should be formatted with dollar signs, as $$...$$. I see the square bracket delimiters \[...\] are working currently, but this is not the officially supported format and may not always work. Information on this can be found in these release notes (<https://community.canvaslms.com/t5/Canvas-Releases/Canvas-Release-Notes-2021-02-20/ta-p/434781#toc-hId-698876024> ).  Regardless of that though, I wanted to get clarification on the issue at hand. When I open the announcement, I see both the inline and block equations, but I see only the block equation when viewing it on the main Announcements page. Is this specifically what you're reporting? (sandbox screenshot for reference (<https://share.getcloudapp.com/jkuPOKRD>)) If so, this is because the Announcements page only shows a snippet of the text, so an inline equation beyond the point the text is truncated isn't treated any differently than the rest of the text not shown. It is interesting that Canvas decides to show the block equation in this view though. Having that removed from the preview would best be submitted as a feature request (<https://community.canvaslms.com/t5/forums/postpage/board-id/ideas/search-before-post-mode/true> ).  If the issues you're reporting are when the full announcement is open, I'm not seeing the same behavior on this end. If that's the case, can you provide a screenshot of what you're seeing? And does the behavior persist in a different browser?  Let me know if I've misinterpreted anything or if there are any other questions as well. We'll be happy to help. Thanks and have a great day!" |

Figure 83: Response from Instructure support regarding the block equaiotns showing up in the summary of the announcements – Note that the URLs have been added in parenthesis. The screenshot is shown in Figure 84.

|  |
| --- |
|  |

Figure 84: The screenshot mentioned above

### Better support for mathematics in Cortina

Won-Kyung Chung [wkchung@kth.se](mailto:wkchung@kth.se) indicated that CORTINA supports mathematical expressions using a class name of “math-tex” and gave an example:

<span class=\"math-tex\">\\(x = {-b \\pm \\sqrt{b^2-4ac} \\over 2a}\\)</span>

Won-Kyung also noted that Cortina does **not** allow images in paragraphs, so the solutions for Cortina and DiVA have to be different.

An important note about the above example is that \over is deprecated and one should use \frac{}{} or one of its variance instead, hence I used this version of the equation in my source LaTeX file.

Figure 85 shows the entry in the Cortina calendar, while Figure 86 and Figure 87 shows zoomed in views of the equations.

|  |
| --- |
|  |

Figure 85: Equations in Cortina claendar entry

|  |
| --- |
|  |

Figure 86: Zoom in on part of the Cortinal calendar entry

|  |
| --- |
|  |

Figure 87: Zoom in on lower part of the Cortinal calendar entry

### Support for mathematics in DiVA

In contrast to Canvas and Cortina, DiVA uses pictures (although these can have a LaTeX expression as an “alt” description of the image contents).

The code does **not** (yet) support the addition of equations to the DiVA entry of the abstracts.

The current DiVA user interface seems to strip out the math-tex classes, i.e., <span class='math-tex'>\[ \int\_{a}^{b} x^2 \,dx \]</span> becomes simply: \[ \int\_{a}^{b} x^2 \,dx \]. However, I do not know what happens when this is later displayed by DiVA, i.e., whether it will appear as an equation or not.

## Acronyms in abstracts

As students may use acronyms in abstracts, there is now support for the commands:\gls{}, \glspl{}, \Gls{}, \Glspl{}, \acrlong{}, \acrshort{}, and \acrfull{}. There is now an optional argument to the extraction program to specify the name of the file with defintions of acronyms using the \newacronyms{label}{acronym}{phrase} for of definitions. This optional argument is shown in Figure 88. Note that expanding acronyms are handled independently for the abstracts, so that they are spell out on first use in each abstract. Note also that there is **no** support for multiple languages for the phrase that is used, i.e., the expansion simply uses the phrase defined in the \newacronyms definition.

|  |
| --- |
| ./extract\_pseudo\_JSON-from\_PDF.py --pdf xxxx.pdf --json xxxx.json --acronyms acronyms.tex |

Figure 88: Specifying acronyms file

Additionally, the macros (from defines.tex) are supported in abstracts: \ie, \eg, \etc, \etal, \first, \second, \third, … . Now that all the acronyms are spelled out, there is no problem with them when making a cover, a calendar entry, or MODS file.

## URLs in abstracts

Currently URLs in abstracts are supported in Canvas course room announcements and calendar but not in the Cortina calendar – where the URL is simply shown as text.

## Getting the PDF from a Canvas assignment and optionally extracting JSON

To simply getting a PDF file from an Canvas course assignment submission there is a new program get\_PDF\_submission.py to help with this. The program checks that the submission has been graded and has the grade 'complete' and then gets the PDF file submitted for a specified assignment. The generic form of the command and an example are shown in Figure 89. Note that the students name is used to for the name of the output file (as a prefix) while the submitted filename is used as the rest of the file name. With the optional [-e] argument, it runs the extraction program. Note that this program is my Canvas-tools github: <https://github.com/gqmaguirejr/Canvas-tools>.

|  |
| --- |
| ./get\_PDF\_submission.py -c course\_id -a assignment\_id -u user\_id  Example:  ./get\_PDF\_submission.py -c 25434 -a 150953 -u 746 -e |

Figure 89: get\_PDF\_submission program

The program could be made more user friendly by being able to specifiy the name of the assignment and the user’s e-mail address or other information to avoid the need to enter the assignment\_id and user\_id.

## Support for new cover

Given that the cover for 3rd cycle dissertations changed during this academic year I have been expecting that there will be a new format for the cover for 1st and 2nd cycle thesis. I received a note about this on 2021-06-23. It seems that it will be based upon a PDF form and will come into use this Fall. So I started to do some experiments about filling in their form. An initial program is fill\_in\_template.py and the command line interface and an example are shown in Figure 90.

|  |
| --- |
| ./fill\_in\_template.py --pdf template.pdf --json data.json  Output: outputs a pdf file named "output.pdf" (currently a fixed name)  Example:  ./fill\_in\_template.py --pdf "KTH\_Omslag\_Exjobb\_Formulär\_Final\_dummy\_EN-20210623.pdf" –json xxxx.json --trita "TRITA-EECS-EX-2021:330" |

Figure 90: New fill\_in\_template.py program

Note that the new template is not yet ready for prime time and this program is a simple hack to see if I can mechanically generate the new format of cover. I had to manually adjust the position of the boxes for the title, substitle, and author (via the code) to fit the actual title of a recent thesis. Ideally the positioning of this text should be done via:

1. Directly use LaTeX or similar program to position the title, subtitle, and author on a page (with the template either as background or as input)
2. Do the computations within the program to compute the size of each box, based on filling the text into a box and computing the sum of the width of the characters and spacing and the heights of the lines to dynamically compute the size of the box. This should also incorporate computation of a fixed leading between the boxes.

Note that the new cover template uses the following fonts: "TheSans B4 SemiLight" and "TheSans B6 SemiBold". I have addressed the problem of making these fonts available from within LaTeX (for the 1st alternative) and made a small demonstration of this[[7]](#footnote-7). Unfortunately, I do not yet understand the font metrics enough to write the code for the 2nd alternative nor do I think it is worth doing this since this is just what a text formatting system does.

Once both the new cover template and the program are more mature the code should get integrated into JSON\_to\_cover.py - with a new option to specify whether you want to "new" or "old" cover.

# Accessibility

Accessibility has been divided into accessibility of the calendar entries, the cover and PDF files, and the template itself. There is also a subsection regarding improving accessibility.

## Calendar entries

Note that the calendar entries that are generated in the Canvas course room are as accessible as all content in Canvas (as Instructure tries to follow the W3C Web Content Accessibility Guidelines (WCAG)). The European standard EN 301 549 V2.1.2 (based upon WACG 2.1) are the accessibility guidelines for web content that are recommended by DIGG (*Myndigheten för digital förvaltning*), based upon the presentation by Tommy Olsson of DIGG to the SUNSET SALSA group on 2021-06-03. Morover, the contents are HTML language tagged, so that a text to speech program that has access *within* Canvas (such as ReaderSpaker[[8]](#footnote-8) is capable of doing as an LTI app) could read the content with the correct pronounciation for each of the two lanaguages. Note that KTH’s current screen reader solution does **not** access the HTML of the page and hence it does **not** use the language tags, thus the user must *manually* choose the language for output.

The calendar entries in the KTH Calendar are in the same format as the current calendar entries. The structure of these entries have been developed in consultation with Robert Lawesson ([lawesson@kth.se](mailto:lawesson@kth.se)) and the KTH Calendar API developer Won-Kyung Chung ([wkchung@kth.se](mailto:wkchung@kth.se)) with the additional help of Niklas Olsson ([hoyce@kth.se](mailto:hoyce@kth.se)). Note that Niklas is the original developer of the KTH cover generator.

## Cover and PDF file

The PDF meta data (author(s), title(s), keywords, etc.) is accessible to any programs that uses the PDF meta data (this is a standard feature of PDF files). Unfortunately, there are no provisions for lanaguge markup for this data.

No investigations has been made of the accessibility of the KTH cover nor of the contents of the PDF file (i.e., the thesis itself). The PDF output by Overleaf appears to be PDF version 1.5 (i.e., accessible via Acrobat version 6 or later).

The template makes use of color with with regard to the **hyperref colors** and **todonotes**. The **hyperref** colors are defined as:

\hypersetup{

colorlinks = true,

breaklinks = true,

linkcolor = \linkscolor,

urlcolor = \urlscolor,

citecolor = \refscolor,

anchorcolor = black

}

Where the colors are defined as (the ForestGreen make lack sufficient contrast for readability):

\definecolor{ForestGreen} {RGB}{34, 139, 34}

\definecolor{HeraldRed2} {rgb}{0.81, 0.12, 0.15}

\newcommand{\refscolor} {blue}

\newcommand{\linkscolor}{HeraldRed2}

\newcommand{\urlscolor} {ForestGreen}

Note that the colors do not encode any special meaning (i.e., they could all be turned to black), since the citations are recognizable by their format, the URL (and URI) by linking to an external document, and the links (by linking within the document.

Some additional colors are defined and used for **todonotes** – they should of of course be removed before the thesis is finalized. The default background color for todonotes is orange (which is predefined as #FF7F00\definecolor{orange(colorwheel)}{rgb}{1.0, 0.5, 0.0}). The color for notes in Swedish is aqua (defined as:\definecolor{aqua}{rgb}{0.0, 1.0, 1.0}). For notes by the authors to themselves using \todoinline are red (defined as \definecolor{red}{rgb}{0.7,0.0,0.0}).

## Template itself

The template itself is written in LaTeX using UTF8 characters. The template uses packages from TeX Live version 2020. It can be compiled with XeLaTeX and LuaLaTeX (note that both natively support UTF8 input). With pdfLaTeX it needs an initial declaration that it should use UTF8 input encoding.´

The template is designed with a set of options to generate a thesis in English or Swedish and to use either bibtex or biblatex, as explained at the top of the main file:

%% set the default language to english or swedish by passing an option to the documentclass - this handles the inside tile page

% to use bibtex or biblatex - edited the following line:

\documentclass[english, bibtex]{kththesis}

%\documentclass[swedish, biblatex]{kththesis}

The template is available from Overleaf (both as a template and via a share URL). The template is also available from a github repository at KTH.

The text in the template in in Swedish and English. The notes regarding the LaTeX class file, the various lib files, and the examplethesis.tex file are in English.

The default bibliographic styles is my own adaptation of the IEEE Tranactions format (i.e., numbered citations, numbered references, references in order of use) with the extention of adding DOIs, URLs, and ISBNs (when relevant).

## Improving accessibility

One method for improving accessibility would be to include the accessibility package, i.e., \usepackage{axessibility} as this would include a comment in the PDF file for each equation with the LaTeX that generated the equation. However, this package is no longer maintained and is incompatible with many other packages. For an introduction to what the LaTeX project has been working on see “LATEX Tagged PDF|A blueprint for a large Project” [1]. Based upon this article and the status of the tagpdf package, I conclude that it is too early to worry about properly tagging PDF files, doing so will have to await the release of packaged designed for produciotn use.

# The structure of the template and the report

The report in itself (i.e., the thesis) has a classical IMRAD structure. In some areas, such as mathimatics there is a tradition for another structure.

The files and folders in Overleaf (or from the github) have the form shown in Figure 91. Some students add folders per chapter and reduce the main document to a skeleton that includes the other parts of the document.

There is currently **not** a user’s manual with instructions concerning the template. Such a user’s manual should be available in both English and Swedish. Most student will only need to add acronyms to the acronyms.tex file and perhaps additional packages to includes.text (in some cases these may need to be added to the includes-after-hyperref.tex file is there is a conflict with hyperref).

The schools\_and\_programs.ins and old\_schools\_and\_programs.ins are generated from the data in KOPPS. Students will not have to change these.

|  |
| --- |
|  |

Figure 91: Files and folders in Overleaf template

# Generating LaTeX commands for in the student’s thesis source file

The thesis makes use of a set of LaTeX commands to collect information about the student, the supervisor(s), and examiner. However, how can a student know what their KTHID and other necessary information is? They can make use of the program: whoami\_for\_latex.py. This program makes use of information from the Canvas course room where the student is enrolled for their degree project. In the examples below we will assume that this is the Canvas course\_id 25434.

Figure 92 shows the case for a student authors who has the e-mail address “[oscarros@kth.se](mailto:oscarros@kth.se)”. Note that the user is prompted to enter their KTH account name and password (in order to access LADOK). Note that this student has two programs of study listed in LADOK (i.e. CDATE and TCSCM). The student will have to manually select one of these and comment out or delete the other, The course code is derived from the section of the course that the user is enrolled in. This section name (in the case of this student: “DA231XVT212”) is reduced to a course code by truncation. The course cycle information is taken from the first digit of the course code. This should work for most course code, except for one program in architecture which is still using a very old course code.

|  |
| --- |
| ./whoami\_for\_latex.py oscarros@kth.se 25434  login\_id=oscarros@kth.se  \authorsLastname{Rosquis}  \authorsFirstname{Oscar}  \email{oscarros@kth.se}  \kthid{u1tmg8l6}  % If the student has an ORCiD - add it here  %\orcid{0000-0002-xxxx-zzzz  %\authorsSchool{\schoolAcronym{EECS}}  \programcode{CDATE}  % Note that the following line should be commented out, as the programme is derived from the school\_andPorgrams.ins information  %\programname{Degree Programme in Computer Science and Engineering}  \programcode{TCSCM}  % Note that the following line should be commented out, as the programme is derived from the school\_andPorgrams.ins information  %\programname{Master's Programme, Computer Science}  \courseCycle{2}  \courseCode{DA231X} |

Figure 92: Generating LaTeX commands for a student author

Figure 93 show the case of generating the data for a second author (NB this is only possible for a 1st cycle thesis). The actual user data has been anonymized in this figure.

|  |
| --- |
| ./whoami\_for\_latex.py -2 xxxx@kth.se 22156  \secondAuthorsLastname{xxxxx}  \secondAuthorsFirstname{Oscar}  \secondemail{xxxx@kth.se}  \secondkthid{u1xxxxx}  \secondAuthorsSchool{\schoolAcronym{ABE}}  %Note that the LaTeX template does not support having students with different programs or course codes  \programcode{CINTE}  % Note that the following line should be commented out, as the programme is derived from the school\_andPorgrams.ins information  %\programname{Degree Programme in Information and Communication Technology}  \courseCycle{1}  \courseCode{IA150X} |

Figure 93: Generating the LaTeX data for the second author (in the case of a 1st cycle thesis)

Figure 94 shows the case where the program has been run to add a second author to a 2nd cycle thesis, note the warning message (**highlighted**) in the last line of this figure. Unfortunately, the highlighting is only shown in the figure and not in the actual output from the program.

|  |
| --- |
| ./whoami\_for\_latex.py -2 oscarros@kth.se 25434  \secondAuthorsLastname{Rosquis}  \secondAuthorsFirstname{Oscar}  \secondemail{oscarros@kth.se}  \secondkthid{u1tmg8l6}  \secondAuthorsSchool{\schoolAcronym{ABE}}  %Note that the LaTeX template does not support having students with different programs or course codes  \programcode{CDATE}  % Note that the following line should be commented out, as the programme is derived from the school\_andPorgrams.ins information  %\programname{Degree Programme in Computer Science and Engineering}  \courseCycle{2}  \courseCode{DA231X}  latex\_courseCycle=\courseCycle{2}, type=<class 'str'>  %%%% Note that a 2nd cycle thesis should not have two authors!!!  \programcode{TCSCM}  % Note that the following line should be commented out, as the programme is derived from the school\_andPorgrams.ins information  %\programname{Master's Programme, Computer Science}  \courseCycle{2}  \courseCode{DA231X}  latex\_courseCycle=\courseCycle{2}, type=<class 'str'>  **%%%% Note that a 2nd cycle thesis should not have two authors!!!** |

Figure 94: Attempting to generate a 2nd author for a 2nd cycle thesis

Figure 95 shows the case for a supervisor who has the e-mail address “vastberg@kth.se”. Figure 96 shows the case for a supervisor who has the e-mail address “maguire@kth.se”. Note that the information about the school and department (in the case of a supervisor and examiner) are currently fixed and commented out – so that the user can edit them as necessary. However, since EECS and Computer Science are the largest school and department (respectively) – it is a good guess ☺.

|  |
| --- |
| ./whoami\_for\_latex.py vastberg@kth.se 25434  %If not the first supervisor, then replace supervisorAs with supervisorBs or supervisorDCAs as appropriate  \supervisorAsLastname{Västber}  \supervisorAsFirstname{Anders}  \supervisorAsEmail{vastberg@kth.se}  % If the supervisor is from within KTH add their KTHID, School and Department info  %\supervisorAsSchool{\schoolAcronym{EECS}}  %\supervisorAsDepartment{Computer Science} |

Figure 95: Generating LaTeX commands for a supervisor

|  |
| --- |
| ./whoami\_for\_latex.py -e maguire@kth.se 25434  \examinersLastname{Maguire J}  \examinersFirstname{Gerald Quentin}  \examinersEmail{maguire@kth.se}  % If the examiner is from within KTH add their KTHID, School and Department info  \examinersSchool{\schoolAcronym{EECS}}  \examinersDepartment{Computer Science} |

Figure 96: Generating LaTeX commands for the examiner

The program: whoami\_for\_latex.py is far more powerful that shown in the examples above, it can take other identifiers for the user in instead of an e-mail address, such a a Canvas user\_id or a KTHID. Additionally, rather than specifying the Canvas course room’s course\_id, it is also possible to use the name of the course room and even a nickname (if you have defined a nickname in your Canvas dashboard). See the README file in <https://gits-15.sys.kth.se/maguire/ladok3>.

# Alternative way of inserting the covers

Another way that the covers can be inserted is to use the pdfpages package (as shown in Figure 97), then insert the two cover pages at the appropriate place (as shown in Figure 98 and Figure 99). Unfortunately, I do not yet know how I can programmatically insert these two files into the Overleaf project (presumably into a folder “covers”). In each figure the existing text is shown in black and the new text to be inserted is shown in red.

|  |
| --- |
| % To use KTH pdf covers  \usepackage{pdfpages}  \include{lib/includes-after-hyperref} |

Figure 97: Inseert this include file either in the main text document or the includes.tex file

|  |
| --- |
| % Add front cover  \includepdf[pages=-]{covers/front.pdf}  %%% Set the numbering for the title page to a numbering series not in the preface or body  \pagenumbering{alph} |

Figure 98: Insert the front cover before the title page

|  |
| --- |
| % Add back cover, unsure if this is supposed to be before or after the "For DIVA" pages  \includepdf[pages=-]{covers/back.pdf}  \section\*{For DIVA} |

Figure 99: Insert the back cover page before the For DIVA section

# JSON\_to\_ladok.py

As part of the effort to be able to minimize the effort of cutting and pasting. I have made a program JSON\_to\_ladok.py that takes the extracted JSON information and uses the information about the **author**(s) and the **title** and **alternative title** to try to insert this information into LADOK for the moment that requires a project title, i.e., 'KravPaProjekttitel' is True. Figure 100 shows an example of using this program to try to put the title and alternative title into LADOK. Basically the program logic should work, but I do not seem to have permission to make entries of this data.

Note that this program uses the ladok3 python library but extends it with some features that are not (yet) in the library. It should be regarded as very much a work in progress.

|  |
| --- |
| ./JSON\_to\_ladokK.py -c 11 --json xxx.json --code DA213X  testing=False  d={'Author1': {'Last name': ‘xxx’, 'First name': 'xxx', 'Local User Id': 'yyyyy', 'E-mail': 'xxx@kth.se', 'organisation': {'L1': 'School of Electrical Engineering and Computer Science '}}, 'Degree': {'Educational program': 'Master’s Programme, Computer Science, 120 credits', 'programcode': 'TCSCM', 'Level': '2', 'Course code': 'DA231X', 'Credits': '30.0', 'Exam': 'Master’s Programme, Computer Science, 120 credits', 'subjectArea': 'Computer Science'}, 'Title': {'Main title': 'xxx', 'Subtitle': yyyy', 'Language': 'eng'}, 'Alternative title': {'Main title': 'zzzz', 'Subtitle': wwwww', 'Language': 'swe'}, 'Supervisor1': {'Last name': 'x', 'First name': 'y', 'Local User Id': 'xx', 'E-mail': 'xxxxx@kth.se', 'organisation': {'L1': 'School of Electrical Engineering and Computer Science ', 'L2': 'Computer Science'}}, 'Supervisor2': {'Last name': 'yyy', 'First name': 'xxx', 'E-mail': 'xxx@zzzz.com', 'Other organisation': ‘zzzz'}, 'Examiner1': {'Last name': 'Maguire Jr.', 'First name': 'Gerald Q.', 'Local User Id': 'u1d13i2c', 'E-mail': 'maguire@kth.se', 'organisation': {'L1': 'School of Electrical Engineering and Computer Science ', 'L2': 'Computer Science'}}, 'Cooperation': {'Partner\_name': 'xxx'}, 'National Subject Categories': '10201, 10206', 'Other information': {'Year': '2021', 'Number of pages': 'xix,99'}, 'Opponents': {'Name': 'A. B. Normal & A. X. E. Normalè'}, 'Presentation': {'Date': '2021-03-15 13:00', 'Language': 'eng', 'Room': 'via Zoom', 'Address': 'Isafjordsgatan 22 (Kistagången 16)', 'City': 'Stockholm'}, 'Number of lang instances': '2', 'abstracts': {'eng': xxxx", 'swe': 'yyyy'}, 'keywords': {'eng': 'x, y z', 'swe': x, y, z'}}  course\_code=DA231X  author={'Last name': 'xxx', 'First name': 'yyy', 'Local User Id': 'u1xxxx', 'E-mail': 'oxxx@kth.se', 'organisation': {'L1': 'School of Electrical Engineering and Computer Science '}}  sortable name=X,Y  Canvas user\_id=dddd  integration\_id=gggggg-gggg-gggg-gggg-ggggggg  ladoK\_course\_info={'id': '6683207e-5a5d-11eb-9b32-eeb44fb14647', 'round\_id': '8e15ae14-1d86-11ea-a622-3565135944de', 'education\_id': '374ea085-73d8-11e8-afa7-8e408e694e54', 'instance\_id': '8eee8da9-dd0a-11e8-bb7a-19f8cd1a470e', 'swe\_name': 'Examensarbete i datalogi och datateknik, avancerad nivå', 'eng\_name': 'Degree Project in Computer Science and Engineering, Second Cycle'}  moment code=PRO1, requires title=False  moment code=PRO2, requires title=False  moment code=PRO3, requires title=True  trying to store a passing grade for moment=PRO3  Traceback (most recent call last):  File "./JSON\_to\_ladok.py", line 533, in <module>  sys.exit(main(sys.argv[1:]))  File "./JSON\_to\_ ladok.py", line 519, in main  status=save\_result\_degree\_project3(ladok, integration\_id, course\_code, mom['Utbildningskod'], '2021-07-14', 'P', "PF", main\_title, alternative\_main\_title)  File "./JSON\_to\_ ladok.py", line 374, in save\_result\_degree\_project3  raise Exception("Couldn't register " + course\_moment + "=" + grade\_raw + " " + result\_date\_raw + ": " + r.json()["Meddelande"])  Exception: Couldn't register PRO3=P 2021-07-14: Hinder mot skapa resultat påträffat: Rapporteringsrättighet saknas |

Figure 100: Using the extracted JSON to produce a LADOK entry

# The big picture

Table 1 shows the relationship between the information and the programs that use it. The sections highlighted in yellow are optional. For example, only 1st cycle theses will have a second author. While all theses will have one supervisor, some may have more (the programs support up to 10). The assumption is that there is a single examiner, but there is the ability to add more (this also requires modifying the programs that use this information). Not every project involves external cooperation – note also that while the MODS program writes out this information the DiVA import does not import this information.

Individuals can be associated with KTH, in this case they have a Local User Id and an organisation element (with the implicit top level (L0) being KTH)). The organization levels of L1 == school and L2 == Department. It is possible to extend this farther, but the programs currently only support L1 and L2. The Title and Alternative title fields are used to support the English and Swedish title (or visa versa) – similarly the abstracts & keywords must include English and Swedish – versions with support for abstracts & keywords in many languages. Note that Credits **must** be given to one decimal place.

Note that would be very rare that the examiner is from outside of KTH but this should (probably) be supported. Where a school should be specified one can use \schoolAcronym{xxxx}.

The Canvas course room’s course\_id is often an argument to the programs. One reason for this is to publish the announcement in the correct course room and calendar. Another reason for this is to be able to use the information in course room to translate a local user ID (i.e. a kthid) to an integration ID (i.e., a LADOK user ID).

Table 1: Information and how it is used

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Top level | elements | subelements | LaTeX Macro | DOCX DocProperty or fields | calendar | cover | LADOK | MODS |
| Author1 |  | | | | | | | |
| Last name |  | \authorsLastname{} | Author1\_Last\_name | x | x | x | x |
| First name |  | \authorsFirstname{} | Author1\_First\_name | x | x | x | x |
| Local User Id |  | \kthid{} | Author1\_Local User Id |  |  | x | x |
| E-mail |  | \email{} | Author1\_E-mail |  |  |  | x |
| organisation |  | | | | | | |
| L1 | \authorsSchool{} | Author1\_organization\_L1 |  |  |  | x |
|  |  | Author1\_organization\_L2 |  |  |  |  |
| Other organisation |  |  | Author1\_Other\_organisation |  |  |  |  |
| Author2 |  | | | | | | | |
|  | Last name |  | \secondAuthorsLastname{} | Author2\_Last\_name | x | x | x | x |
| First name |  | \secondAuthorsFirstname{} | Author2\_First\_name | x | x | x | x |
| Local User Id |  | \secondkthid{} | Author2\_Local User Id |  |  | x | x |
| E-mail |  | \secondemail{} | Author2\_E-mail |  |  |  | x |
| organisation |  | | | | | | |
| L1 | \secondAuthorsSchool{ } | Author2\_organization\_L1 |  |  |  | x |
|  |  | Author2\_organization\_L2 |  |  |  |  |
| Other organisation |  |  | Author2\_Other\_organisation |  |  |  |  |
| Cycle |  |  | \courseCycle{} | Cycle | x | x |  |  |
| Course code |  |  | \courseCode{} | Course\_code | i | i |  |  |
| Credits |  |  | \courseCredits{} | Credits |  |  |  |  |
| Degree1 |  | | | | | | | |
| Educational program |  | Derived from programcode | Educational program |  | x |  | x |
| programcode |  | \programcode{} | programcode |  | i |  | i |
| Degree |  | \degreeName{} | Degree |  | x |  | x |
| subjectArea |  | \subjectArea{ } | subjectArea |  | x |  | x |
| Degree2 |  | | | | | | | |
| Educational program |  | Derived from secondProgramcode | second\_Educational program |  | x |  | x |
| programcode |  | \secondProgramcode{} | second\_Programcode |  | i |  | i |
| Degree |  | \secondDegreeName{} | second\_Degree |  | x |  | x |
| subjectArea |  | \secondSubjectArea{} | second\_subjectArea |  | x |  | x |
| Title |  | | | | | | | |
| Main title |  | \title{} | (exiting document property) | x | x | x | x |
| Subtitle |  | \subtitle{} | Subtitle | x | x | o | x |
| Language |  | Derived from option to document class |  | x |  |  | x |
| Alternative title |  | | | | | | | |
| Main title |  | \alttitle{} | Alternative\_main\_title | x | x | x | x |
| Subtitle |  | \altsubtitle{} | Alternative\_subtitle | x | x | o | x |
| Language |  | Derived from option to documentclass |  | x |  |  | x |
| Supervisor1 |  | | | | | | | |
| Last name |  | \supervisorAsLastname{} | Supervisor1\_Last\_name | x |  |  | x |
| First name |  | \supervisorAsFirstname{} | Supervisor1\_First\_name | x |  |  | x |
| Local User Id |  | \supervisorAsKTHID{} | Supervisor1\_Local User Id |  |  |  | x |
| E-mail |  | \supervisorAsEmail{} | Supervisor1\_E-mail |  |  |  | x |
| organisation |  | | | | | | |
|  | L1 | \supervisorAsSchool{} | Supervisor1\_organization\_L1 |  |  |  | x |
|  | L2 | supervisorAsDepartment{} | Supervisor1\_organization\_L2 |  |  |  | o |
| Other organisation |  | supervisorAsOrganization{} | Supervisor1\_Other\_organisation |  |  |  |  |
| Supervisor2 |  | | | | | | | |
| Last name |  | \supervisorBsLastname{ } | Supervisor2\_Last\_name | x |  |  | x |
| First name |  | \supervisorBsFirstname{} | Supervisor2\_First\_name | x |  |  | x |
| Local User Id |  | \supervisorBsKTHID{} | Supervisor2\_Local User Id |  |  |  | x |
| E-mail |  | \supervisorBsEmail{} | Supervisor2\_E-mail |  |  |  | x |
| organisation |  | | | | | | |
|  | L1 | \supervisorBsSchool{} | Supervisor2\_organization\_L1 |  |  |  | x |
|  | L2 | \supervisorBsDepartment{} | Supervisor2\_organization\_L2 |  |  |  | o |
| Other organisation |  | \supervisorBsOrganization{} | Supervisor2\_Other\_organisation |  |  |  |  |
| Supervisor3 |  | | | | | | | |
| Last name |  | \supervisorCsLastname{} | Supervisor3\_Last\_name | x |  |  | x |
| First name |  | \supervisorCsFirstname{} | Supervisor3\_First\_name | x |  |  | x |
| Local User Id |  | \supervisorCsKTHID{} |  |  |  |  |  |
| E-mail |  | \supervisorCsEmail{} | Supervisor3\_E-mail |  |  |  | o |
| organisation |  | | | | | | |
|  | L1 | \supervisorCsSchool{} | Supervisor3\_organization\_L1 |  |  |  | x |
|  | L2 | \supervisorCsDepartment{} | Supervisor3\_organization\_L2 |  |  |  | o |
| Other organisation |  | \supervisorCsOrganization{} | Supervisor3\_Other\_organisation | x |  |  | x |
| Examiner1 |  | | | | | | | |
| Last name |  | \examinersLastname{} | Examiner1\_Last\_name | x |  | i | x |
| First name |  | \examinersFirstname{} | Examiner1\_First\_name | x |  | i | x |
| Local User Id |  | \examinersKTHID{} | Examiner1\_Local User Id |  |  | i | x |
| E-mail |  | \examinersEmail{} | Examiner1\_E-mail |  |  |  | x |
| organisation |  | | | | | | |
|  | L1 | \examinersSchool{} | Examiner1\_organization\_L1 |  |  |  | x |
|  | L2 | \examinersDepartment{} | Examiner1\_organization\_L2 |  |  |  |  |
| Other organisation |  | \examinersOrganization{} |  |  |  |  | o |
| Cooperation |  | | | | | | | |
| Partner\_name |  | \hostcompany{} or \hostorganization{} | Cooperation\_Partner\_name |  |  |  | o |
| National Subject Categories |  |  | \nationalsubjectcategories{} | National Subject Categories |  |  |  | x |
| Other information |  | | | | | | | |
| Year |  | \the\year | (derived from document property Date) |  |  |  | x |
| Number of pages |  | (derived from labels: pg:lastPageofPreface and pg:lastPageofMainmatter) | (derived from bookmarks: lastPageofPreface and lastPageofMainmatter) |  |  |  | x |
| Series |  | | | | | | | |
| Title of series |  | \trita{TRITA-EECS-EX}{2021:00} | Series\_name |  | x |  | x |
| No. in series |  | Number\_in\_series |  | x |  | x |
| Opponents |  |  | \opponentsNames{} | Opponents\_Name | x |  |  |  |
| Presentation |  | | | | | | | |
| Date |  | \presentationDateAndTimeISO{} | Presentation\_Date | x |  |  | x |
| Language |  | \presentationLanguage{ } | Presentation\_Language | x |  |  | x |
| Room |  | \presentationRoom{} | Presentation\_Room | x |  |  | x |
| Address |  | \presentationAddress{ } | Presentation\_Address | x |  |  | x |
| City |  | \presentationCity{ } | Presentation\_City | x |  |  | x |
| Number of lang instances |  |  |  | [Manually entered as 2, 3, 4, …] |  |  |  |  |
| abstracts |  | | | | | | | |
| eng |  | (saved in lang array using scontents) | (derived from bookmark: EnglishAbstract and then text inserted) | x |  |  | x |
| swe |  | (derived from bookmark: SwedishAbstract) | x |  |  | x |
| fre |  |  |  |  |  | o |
| … |  |  |  |  |  | o |
| keywords |  | | | | | | | |
| eng |  | (saved in lang array using scontents) | (derived from bookmark: EnglishKeywords) | x |  |  | x |
| swe |  | (derived from bookmark: SwedishKeywords) | x |  |  | x |
| fre |  |  |  |  |  | o |
| … |  |  |  |  |  | o |

# Alternative way of generating JSON for a LaTeX template

After writing a program to extract the JSON information directly from the DOCX file, I was inspired to try to more directly get the JASON data from the LaTeX template; thus, in both cases avoiding the need to parse the PDF file.

While one might think about simply writing the information to a file rather than rendering it in the PDF document, there turned out to be several problems, as the LaTeX \write turns out **not** to expand macros and commands that are ***not expandable*** – while writing this same content to the page ends up with the correct rendering of the page (as this action causes the expansion and evaluation of the macros).

The first problem that I encountered was that the xstring function \IfEqCase that I had used in a number of places turned out to not be expandable in a \immediate\write call. Thus I needed to use an expandable version. The new code for kththesis.cls is shown in Figure 101. I changed to use expl3 strcompare rather than xstring's IfEqCase, since the later is not expandable[[9]](#footnote-9). For example, the code in Figure 102 needed to be replaced by the code in Figure 103. Similarly for the code of \programmecode (partially) shown in Figure 104. This means that there needs to be a new schools\_and\_programs.ins file. [Note that I have not yet written a new version of the program to generate this file.]

|  |
| --- |
| \ExplSyntaxOn  \cs\_new\_eq:NN \strcompare \str\_if\_eq:nnTF  \ExplSyntaxOff  \newcommand{\InsertKeywords}[1]{  \strcompare{#1}{english}{\@EnglishKeywords}{}%  \strcompare{#1}{swedish}{\@SwedishKeywords}{}% |

Figure 101: New version of [\Insert\Keywords](file:///\\Insert\Keywords) tother with making the expl3 \strcompare function available.

|  |
| --- |
| \newcommand{\schoolAcronym}[1]{%  \ifinswedish  \IfEqCase{#1}{%  {ABE}{Skolan för Arkitektur och samhällsbyggnad}%  {ITM}{Skolan för Industriell teknik och management}%  {SCI}{Skolan för Teknikvetenskap}%  {CBH}{Skolan för Kemi, bioteknologi och hälsa}%  {EECS}{Skolan för Elektroteknik och datavetenskap}%  }[\typeout{school's code not found}]  \else  \IfEqCase{#1}{%  {ABE}{School of Architecture and the Built Environment}%  {ITM}{School of Industrial Engineering and Management}%  {SCI}{School of Engineering Sciences}%  {CBH}{School of Engineering Sciences in Chemistry, Biotechnology and Health}%  {EECS}{School of Electrical Engineering and Computer Science}%  }[\typeout{school's code not found}]  \fi  } |

Figure 102: Original \schoolAcronym

|  |
| --- |
| % The strcompare is added to kththesis.cls  \newcommand{\schoolAcronym}[1]{%  \ifinswedish  \strcompare{#1}{ABE}{Skolan för Arkitektur och samhällsbyggnad}{}%  \strcompare{#1}{ITM}{Skolan för Industriell teknik och management}{}%  \strcompare{#1}{SCI}{Skolan för Teknikvetenskap}{}%  \strcompare{#1}{CBH}{Skolan för Kemi, bioteknologi och hälsa}{}%  \strcompare{#1}{EECS}{Skolan för Elektroteknik och datavetenskap}{}%  \else  \strcompare{#1}{ABE}{School of Architecture and the Built Environment}{}%  \strcompare{#1}{ITM}{School of Industrial Engineering and Management}{}%  \strcompare{#1}{SCI}{School of Engineering Sciences}{}%  \strcompare{#1}{CBH}{School of Engineering Sciences in Chemistry, Biotechnology and Health}{}%  \strcompare{#1}{EECS}{School of Electrical Engineering and Computer Science}{}%  \fi  } |

Figure 103: new \schoolAcronym

|  |
| --- |
| \newcommand{\programmecode}[1]{%  \ifinswedish  \strcompare{#1}{ARKIT}{Arkitektutbildning}{}%  \strcompare{#1}{CBIOT}{Civilingenjörsutbildning i bioteknik}{}%  ...  \strcompare{#1}{TURSM}{Magisterprogram, urbana studier}{}%  \else  \strcompare{#1}{ARKIT}{Degree Programme in Architecture}{}%  \strcompare{#1}{CBIOT}{Degree Programme in Biotechnology}{}%  ...  \strcompare{#1}{TURSM}{Master's Programme, Urbanism Studies, 60 credits}{}%  \fi  } |

Figure 104: New version of \programcode

Figure 105 shows the code to write the first author’s information to the JSON file.

|  |
| --- |
| \newwrite\jsonfile  \immediate\openout\jsonfile=fordiva.json  \immediate\write\jsonfile{ \@charlb }  \immediate\write\jsonfile{"Author1": \@charlb \ifx\@authorsLastname\@empty%\relax  \else  "Last name": "\@authorsLastname",  \fi  \ifx\@authorsFirstname\@empty%\relax  \else  "First name": "\@authorsFirstname",  \fi  \ifx\@kthid\@empty%\relax  \else  "Local User Id": "\@kthid",  \fi  \ifx\@email\@empty%\relax  \else  "E-mail": "\@email",  \fi  \ifx\@authorsSchool\@empty%\relax  \else  "organisation": \@charlb"L1": "\@authorsSchool"  \ifx\@authorsDepartment\@empty%\relax  \else  ,"L2": "\@authorsDepartment"  \fi  \@charrb  \fi  \@charrb,}  …  \immediate\write\jsonfile{\@charrb} % end the JSON dict  \closeout\jsonfile |

Figure 105: Code to write the first author's information to a "JSON" file

The second problem that I encountered was that the function \pageref used to get the page numbers of the end of the preface and the last page of the document were also not expandable. This was solved using the package refcount to be able to get an expandable \getpagerefnumber. This new code is shown in Figure 106

|  |
| --- |
| % "\pageref{pg:lastPageofPreface},\pageref{pg:lastPageofMainmatter}" \@charrb,  \immediate\write\jsonfile{  "Other information": \@charlb"Year": "\the\year", "Number of pages":  "\getpagerefnumber{pg:lastPageofPreface},\getpagerefnumber{pg:lastPageofMainmatter}" \@charrb,  } |

Figure 106: New code to ouput the "Other information" including page numbers

The third problem turned out to be **much** harder to solve. The scontents related functions that I used in the newcommand divainfo: \getstored[\i]{lang}, \getstored[\i]{keywords}, and \typestoredx{\i}{abstracts} also turned out to not be expandable. The approach taken here is somewhat of a cheat in that it required getting the *strings* **directly** from the scontents buffer (see the code in Figure 107) and then using xstring’s tokenize function (see thc code in Figure 108).

|  |
| --- |
| \ExplSyntaxOn  % Note that this function returns the \_string\_ stored in the scontents buffer  \cs\_new:Npn \qgetstored #1 #2  {  \\_\_scontents\_getfrom\_seq:nn {#1} {#2}  } |

Figure 107: Function to get the scontents

|  |
| --- |
| \immediate\write\jsonfile{"Number of lang instances": "\countsc{lang}",}  \immediate\write\jsonfile{"abstracts": \@charlb}  \foreach \i in {1,...,\countsc{lang}} {  %\getstored[\i]{lang}  \newcommand{\templangs}{\qgetstored{\i}{lang}}  \tokenize{\lang}{\templangs}  \immediate\write\jsonfile{"\lang": "\qgetstored{\i}{abstracts}"}  \immediate\write\jsonfile{,}  }  \immediate\write\jsonfile{\@charrb} % end the abstracts dict  \immediate\write\jsonfile{"keywords": \@charlb}  \foreach \i in {1,...,\countsc{lang}} {  \newcommand{\templangs}{\qgetstored{\i}{lang}}  \tokenize{\lang}{\templangs}  \newcommand{\fee}{\qgetstored{\i}{keywords}}  % Because \qgetstored{\i}{keywords} returned a string,  % we need to convert this string into tokens so it can be evaluate  % we do this using the function \tokenisze from the xstring package  \tokenize{\fumble}{\fee}  \immediate\write\jsonfile{"\lang": "\fumble"}  \immediate\write\jsonfile{,}  }  \immediate\write\jsonfile{\@charrb} % end the keywords dict |

Figure 108: Writing out the language, abstract, and keywords information

The result is that LaTeX directly generates a file named “fordiva.json” with the (almost) desired information in it. See Table 2 for an example of such a file.

Table 2: Resulting fordiva.json file

|  |
| --- |
| {  "Author1": {"Last name": "Student", "First name": "Fake A.", "Local User Id": "u100001", "E-mail": "a@kth.se", "organisation": {"L1": "School of Electrical Engineering and Computer Science" }},  "Author2": {"Last name": "Student", "First name": "Fake B.", "Local User Id": "u100002", "E-mail": "b@kth.se", "organisation": {"L1": "School of Architecture and the Built Environment" }},  "Cycle": "1", "Course code": "IA150X", "Credits": "15.0",  "Degree1": {"Educational program": "" ,"programcode": "TCOMK" ,"Degree": "Bachelors degree" ,"subjectArea": "Information and Communication Technology" },  "Title": {"Main title": "This is the title in the language of the thesis", "Subtitle": "An subtitle in the language of the thesis", "Language": "eng" }, "Alternative title": {"Main title": "Detta är den svenska översättningen av titeln", "Subtitle": "Detta är den svenska översättningen av undertiteln", "Language": "swe" },  "Supervisor1": {"Last name": "Supervisor", "First name": "A. Busy", "Local User Id": "u100003", "E-mail": "sa@kth.se", "organisation": {"L1": "School of Electrical Engineering and Computer Science" ,"L2": "Computer Science" }},  "Supervisor2": {"Last name": "Supervisor", "First name": "Another Busy", "Local User Id": "u100003", "E-mail": "sb@kth.se", "organisation": {"L1": "School of Architecture and the Built Environment" ,"L2": "Architecture" }},  "Supervisor3": {"Last name": "Supervisor", "First name": "Third Busy", "E-mail": "sc@tu.va", "Other organisation": "Timbuktu University, Department of Pseudoscience" },  "Examiner1": {"Last name": "Maguire Jr.", "First name": "Gerald Q.", "Local User Id": "u1d13i2c", "E-mail": "maguire@kth.se", "organisation": {"L1": "School of Electrical Engineering and Computer Science" ,"L2": "Computer Science" }},  "Cooperation": {"Partner\_name": "Företaget AB"},  "National Subject Categories": "10201, 10206",  "Other information": {"Year": "2021", "Number of pages": "xxxv,35" },  "Series": {"Title of series": "TRITA-EECS-EX" , "No. in series": "2021:00" },  "Opponents": {"Name": "A. B. Normal \& A. X. E. Normalè"},  "Presentation": {"Date": "2021-03-15 13:00" ,"Language": "eng" ,"Room": "via Zoom https://kth-se.zoom.us/j/ddddddddddd" ,"Address": "Isafjordsgatan 22 (Kistagången 16)" ,"City": "Stockholm" },  "Number of lang instances": "10",  "abstracts": {  "eng": €€€€  "Write an abstract that is about \textit{250 and 350 words} (1/2 A4-page) with the following components: % key parts of the abstract  \begin{itemize}  \item What is the topic area? (optional) Introduces the subject area for the project.  \item Short problem statement  \item Why was this problem worth a Bachelor's/Master’s thesis project? (\ie, why is the problem both significant and of a suitable degree of difficulty for a Bachelor's/Master’s thesis project? Why has no one else solved it yet?)  \item How did you solve the problem? What was your method/insight?  \item Results/Conclusions/Consequences/Impact: What are your key results/\linebreak[4]conclusions? What will others do based upon your results? What can be done now that you have finished - that could not be done before your thesis project was completed?  \end{itemize}  For testing the following lines were added.  Choice of typeface with \textbackslash textit, \textbackslash textbf, and \textbackslash texttt: \textit{x}, \textbf{x}, and \texttt{x}  Text superscripts and subscripts with \textbackslash textsubscript and \textbackslash textsuperscript: A\textsubscript{x} and A\textsuperscript{x}  Some useful symbols: \textbackslash textregistered, \textbackslash texttrademark, and \textbackslash textcopyright. For example, copyright symbol: \textbackslash textcopyright Maguire 2021, and some superscripts: \textbackslash textsuperscript\{99m\}Tc, A\textbackslash textsuperscript\{\*\}, A\textbackslash textsuperscript\{\textbackslash textregistered\}, and A\textbackslash texttrademark : \textcopyright Maguire 2021, and some superscripts: \textsuperscript{99m}Tc, A\textsuperscript{\*}, A\textsuperscript{\textregistered}, and A\texttrademark. Another example: H\textbackslash textsubscript\{2\}O: H\textsubscript{2}O  \begin{enumerate}  \item The first item is a simple \textbf{strong} statement  \item The second item is a simple statement  \item The third item is an \textit{italicized} statement  \end{enumerate}  Yet more text can be added.  The following macros can be used: \textbackslash eg, \textbackslash Eg, \textbackslash ie, \textbackslash Ie, \textbackslash etc, and \textbackslash etal: \eg, \Eg, \ie, \Ie, \etc, and \etal  The following macros for numbering with lower case roman numerals: \textbackslash first, \textbackslash second, \textbackslash third, \textbackslash fourth, \textbackslash fifth, \textbackslash sixth, \textbackslash seventh, and \textbackslash eighth: \first, \second, \third, \fourth, \fifth, \sixth, \seventh, and \eighth.  Equations using \textbackslash( xxxx \textbackslash) or \textbackslash[ xxxx \textbackslash] can be used in the abstract. For example: \( (C\_5O\_2H\_8)\_n \)  or \[ \int\_{a}^{b} x^2 \,dx \]  The above showed even more text and equations.  "  €€€€,  "swe": €€€€  "Denna avhandling undersöker problemet \ldots"  €€€€,  "fre": €€€€  "Résumé en français."  €€€€,  "spa": €€€€  "Résumé en espagnol."  €€€€,  "ita": €€€€  "Sommario in italiano."  €€€€,  "nor": €€€€  "Sammendrag på norsk."  €€€€,  "ger": €€€€  "Zusammenfassung in Deutsch."  €€€€,  "dan": €€€€  "Abstrakt på dansk."  €€€€,  "dut": €€€€  "Samenvatting in het Nederlands."  €€€€,  "est": €€€€  "Eesti keeles kokkuvõte."  €€€€,  },  "keywords": {  "eng": €€€€  " Canvas Learning Management System, Docker containers, Performance tuning"  €€€€,  "swe": €€€€  " Canvas Lärplattform, Dockerbehållare, Prestandajustering"  €€€€,  "fre": €€€€  "5-6 mots-clés"  €€€€,  "spa": €€€€  "5-6 Palabras claves"  €€€€,  "ita": €€€€  "5-6 parole chiave"  €€€€,  "nor": €€€€  "5-6 nøkkelord"  €€€€,  "ger": €€€€  "5-6 Schlüsselwörter"  €€€€,  "dan": €€€€  "5-6 Søgeord"  €€€€,  "dut": €€€€  "5-6 trefwoorden"  €€€€,  "est": €€€€  "5-6 Märksõnad"  €€€€,  }  } |

The Overeleaf project used to workout the string reading and tokenize is at <https://www.overleaf.com/read/mqtrxrdcssmg> while the “KTH thesis template for 1st and 2nd cycle degree projects with JSON output” can be found at <https://www.overleaf.com/read/gxhwkywbtgsn>.

The reason for adding the reservation “(almost)” in the earlier paragraph was that this file is not directly readable by my programs that could accept the pseudo JSON that I produced with earlier programs. Some of the problems are that the opponents’ name string contains “\&”, i.e. it is "A. B. Normal \& A. X. E. Normalè" and this is not acceptable to python’s json.loads{} function; thus, this has to be rewritten. Additionally, the abstracts contain the raw LaTeX and thus can include acronyms that need spelled out. This means that the same transformations used in extract\_pseudo\_JSON-from\_PDF.py need to be implemented in a new program to clean up the pseudo JSON. The new program is named: cleanup\_pseudo\_JSON-from\_LaTeX.py .

Table 3: Resut after cleanup for the contents of the earlier fordiva.json file, called fordiva-cleaned

|  |
| --- |
| {"Author1": {"Last name": "Student", "First name": "Fake A.", "Local User Id": "u100001", "E-mail": "a@kth.se", "organisation": {"L1": "School of Electrical Engineering and Computer Science"}}, "Author2": {"Last name": "Student", "First name": "Fake B.", "Local User Id": "u100002", "E-mail": "b@kth.se", "organisation": {"L1": "School of Architecture and the Built Environment"}}, "Cycle": "1", "Course code": "IA150X", "Credits": "15.0", "Degree1": {"Educational program": "", "programcode": "TCOMK", "Degree": "Bachelors degree", "subjectArea": "Information and Communication Technology"}, "Title": {"Main title": "This is the title in the language of the thesis", "Subtitle": "An subtitle in the language of the thesis", "Language": "eng"}, "Alternative title": {"Main title": "Detta är den svenska översättningen av titeln", "Subtitle": "Detta är den svenska översättningen av undertiteln", "Language": "swe"}, "Supervisor1": {"Last name": "Supervisor", "First name": "A. Busy", "Local User Id": "u100003", "E-mail": "sa@kth.se", "organisation": {"L1": "School of Electrical Engineering and Computer Science", "L2": "Computer Science"}}, "Supervisor2": {"Last name": "Supervisor", "First name": "Another Busy", "Local User Id": "u100003", "E-mail": "sb@kth.se", "organisation": {"L1": "School of Architecture and the Built Environment", "L2": "Architecture"}}, "Supervisor3": {"Last name": "Supervisor", "First name": "Third Busy", "E-mail": "sc@tu.va", "Other organisation": "Timbuktu University, Department of Pseudoscience"}, "Examiner1": {"Last name": "Maguire Jr.", "First name": "Gerald Q.", "Local User Id": "u1d13i2c", "E-mail": "maguire@kth.se", "organisation": {"L1": "School of Electrical Engineering and Computer Science", "L2": "Computer Science"}}, "Cooperation": {"Partner\_name": "Företaget AB"}, "National Subject Categories": "10201, 10206", "Other information": {"Year": "2021", "Number of pages": "xxxv,35"}, "Series": {"Title of series": "TRITA-EECS-EX", "No. in series": "2021:00"}, "Opponents": {"Name": "A. B. Normal &amp; A. X. E. Normalè"}, "Presentation": {"Date": "2021-03-15 13:00", "Language": "eng", "Room": "via Zoom https://kth-se.zoom.us/j/ddddddddddd", "Address": "Isafjordsgatan 22 (Kistagången 16)", "City": "Stockholm"}, "Number of lang instances": "10", "abstracts": {"eng": "<p>Write an abstract that is about <i>250 and 350 words</i> (1/2 A4-page) with the following components: </p><p><ul><li>What is the topic area? (optional) Introduces the subject area for the project.</li><li>Short problem statement</li><li>Why was this problem worth a Bachelor's/Master’s thesis project? (i.e., why is the problem both significant and of a suitable degree of difficulty for a Bachelor's/Master’s thesis project? Why has no one else solved it yet?)</li><li>How did you solve the problem? What was your method/insight?</li><li>Results/Conclusions/Consequences/Impact: What are your key results/ conclusions? What will others do based upon your results? What can be done now that you have finished - that could not be done before your thesis project was completed?</li></ul></p><p>For testing the following lines were added.<BR>Choice of typeface with \\textit, \\textbf, and \\texttt: <i>x</i>, <strong>x</strong>, and <tt>x</tt></p><p>Text superscripts and subscripts with \\textsubscript and \\textsuperscript: A<sub>x</sub> and A<sup>x</sup></p><p>Some useful symbols: \\textregistered, \\texttrademark, and \\textcopyright. For example, copyright symbol: \\textcopyright Maguire 2021, and some superscripts: \\textsuperscript\\{99m\\}Tc, A\\textsuperscript\\{\*\\}, A\\textsuperscript\\{\\textregistered\\}, and A\\texttrademark : &copy; Maguire 2021, and some superscripts: <sup>99m</sup>Tc, A<sup>\*</sup>, A<sup>&reg;</sup>, and A&trade;. Another example: H\\textsubscript\\{2\\}O: H<sub>2</sub>O</p><p><ol><li>The first item is a simple <strong>strong</strong> statement</li><li>The second item is a simple statement</li><li>The third item is an <i>italicized</i> statement</li></ol></p><p>Yet more text can be added.</p><p>The following macros can be used: \\eg, \\Eg, \\ie, \\Ie, \\etc, and \\etal: e.g., E.g., i.e., I.e., etc., and et al.</p><p>The following macros for numbering with lower case roman numerals: \\first, \\second, \\third, \\fourth, \\fifth, \\sixth, \\seventh, and \\eighth: (i) , (ii) , (iii) , (iv) , (v) , (vi) , (vii) , and (viii) .</p><p>Equations using \\textbackslash( xxxx \\textbackslash) or \\textbackslash[ xxxx \\textbackslash] can be used in the abstract. For example: \\( (C\_5O\_2H\_8)\_n \\)<BR>or \\[ \\int\_{a}^{b} x^2 \\,dx \\]</p><p>The above showed even more text and equations.</p>", "swe": "<p>Denna avhandling undersöker problemet ... </p>", "fre": "<p>Résumé en français.</p>", "spa": "<p>Résumé en espagnol.</p>", "ita": "<p>Sommario in italiano.</p>", "nor": "<p>Sammendrag på norsk.</p>", "ger": "<p>Zusammenfassung in Deutsch.</p>", "dan": "<p>Abstrakt på dansk.</p>", "dut": "<p>Samenvatting in het Nederlands.</p>", "est": "<p>Eesti keeles kokkuvõte.</p>"}, "keywords": {"eng": "Canvas Learning Management System, Docker containers, Performance tuning", "swe": "Canvas Lärplattform, Dockerbehållare, Prestandajustering", "fre": "5-6 mots-clés", "spa": "5-6 Palabras claves", "ita": "5-6 parole chiave", "nor": "5-6 nøkkelord", "ger": "5-6 Schlüsselwörter", "dan": "5-6 Søgeord", "dut": "5-6 trefwoorden", "est": "5-6 Märksõnad"}} |

Executing the command:

./JSON\_to\_calendar.py -c 11 --config config-test.json --json fordiva-cleaned.json --nocortina

will produce the cource announcement shown in Figure 109.

|  |
| --- |
|  |

Figure 109: Resulting announcement

This contrived example shows that a LaTeX abstract of the form shown in Table 4 can be handled. While more testing should be done, this shows that it is possible to directly extract the pseudo JSON from the LaTeX without having to parse and process the PDF file; however, you still have to compile the LaTeX and produce the PDF file in order to produce the fordiva.json file.

Note that the cleanup\_pseudo\_JSON-from\_LaTeX.py program does simple text substitution and **cannot** handle nested LaTeX commands.

Table 4: Source LaTeX (from inside the scontents of the abstract=

|  |
| --- |
| Write an abstract that is about \textit{250 and 350 words} (1/2 A4-page) with the following components: % key parts of the abstract  \begin{itemize}  \item What is the topic area? (optional) Introduces the subject area for the project.  \item Short problem statement  \item Why was this problem worth a Bachelor's/Master’s thesis project? (\ie, why is the problem both significant and of a suitable degree of difficulty for a Bachelor's/Master’s thesis project? Why has no one else solved it yet?)  \item How did you solve the problem? What was your method/insight?  \item Results/Conclusions/Consequences/Impact: What are your key results/\linebreak[4]conclusions? What will others do based upon your results? What can be done now that you have finished - that could not be done before your thesis project was completed?  \end{itemize}  For testing the following lines were added.  Choice of typeface with \textbackslash textit, \textbackslash textbf, and \textbackslash texttt: \textit{x}, \textbf{x}, and \texttt{x}  Text superscripts and subscripts with \textbackslash textsubscript and \textbackslash textsuperscript: A\textsubscript{x} and A\textsuperscript{x}  Some useful symbols: \textbackslash textregistered, \textbackslash texttrademark, and \textbackslash textcopyright. For example, copyright symbol: \textbackslash textcopyright Maguire 2021, and some superscripts: \textbackslash textsuperscript\{99m\}Tc, A\textbackslash textsuperscript\{\*\}, A\textbackslash textsuperscript\{\textbackslash textregistered\}, and A\textbackslash texttrademark : \textcopyright Maguire 2021, and some superscripts: \textsuperscript{99m}Tc, A\textsuperscript{\*}, A\textsuperscript{\textregistered}, and A\texttrademark. Another example: H\textbackslash textsubscript\{2\}O: H\textsubscript{2}O  \begin{enumerate}  \item The first item is a simple \textbf{strong} statement  \item The second item is a simple statement  \item The third item is an \textit{italicized} statement  \end{enumerate}  Yet more text can be added.  The following macros can be used: \textbackslash eg, \textbackslash Eg, \textbackslash ie, \textbackslash Ie, \textbackslash etc, and \textbackslash etal: \eg, \Eg, \ie, \Ie, \etc, and \etal  The following macros for numbering with lower case roman numerals: \textbackslash first, \textbackslash second, \textbackslash third, \textbackslash fourth, \textbackslash fifth, \textbackslash sixth, \textbackslash seventh, and \textbackslash eighth: \first, \second, \third, \fourth, \fifth, \sixth, \seventh, and \eighth.  Equations using \textbackslash( xxxx \textbackslash) or \textbackslash[ xxxx \textbackslash] can be used in the abstract. For example: \( (C\_5O\_2H\_8)\_n \)  or \[ \int\_{a}^{b} x^2 \,dx \]  The above showed even more text and equations. |

References

[1] Frank Mittelbach and Chris Rowley, ‘LATEX Tagged PDF|A blueprint for a large project’, *TUGboat,* vol. 41, no. 3, pp. 292–298, 2020 [Online]. Available: <https://www.latex-project.org/publications/2020-FMi-TUB-tb129mitt-tagpdf.pdf>

1. Shiva Besharat and Qi Li*, Connecting Silos: Automation system for thesis processing in Canvas and DiVA.* Bachelor's thesis, Stockholm, Sweden: KTH: Skolan för elektroteknik och datavetenskap (EECS): Kommunikationssystem (CoS), *2018, TRITA-EECS-EX-2018:164.* [*http://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Akth%3Adiva-230996*](http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-230996) [↑](#footnote-ref-1)
2. Sometimes also referred to as an advisor. [↑](#footnote-ref-2)
3. The particular exam that a student is going to apply for affects the thesis cover in a complex way. For a gallery of covers see <https://canvas.kth.se/courses/11/pages/templates-for-1st-and-2nd-cycle-theses-some-templates> and in Overleaf at <https://www.overleaf.com/read/fmpsjbjcfhpj>. [↑](#footnote-ref-3)
4. I would like to acknowledge the permission of Oscar Rosquist for the permission to use his thesis as an example in my documentation. [↑](#footnote-ref-4)
5. Note that you will need to download the student’s acronyms.tex file from their project to have their list of acronyms. [↑](#footnote-ref-5)
6. The template used to make the examples in this document is at: <https://www.overleaf.com/read/xmrfhcchgnvq> [↑](#footnote-ref-6)
7. However, the situation with the font licenses is unclear to me, hence I have not included a link to this project.. [↑](#footnote-ref-7)
8. <https://www.readspeaker.com/education/learning-management-systems/> [↑](#footnote-ref-8)
9. The insight for this is from Enrico Gregorio - egreg's posing of 26 April 2016 at <https://tex.stackexchange.com/questions/306484/how-do-i-perform-an-expandable-string-comparison> [↑](#footnote-ref-9)