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 publicaly 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, and (3) facilitate entering the meta data and thesis into DiVA.

The focus of this document will be on the LaTeX template for 1st and 2nd cycle degree projects. Future document will examine the DOCX version of the template and the 3rd cycle LaTeX template.

Most readers will only be interested in the high-level overview given in Section 1. Some readers may be interested iin accessibility, see Section 16. The rest of the document should contain enough information to let someone else deal with the problems that remain. Missing is a easy to read user manual.

.

Table of Contents

[1 Accuracy and Economics 6](#_Toc75776655)

[1.1 Making an announcement 6](#_Toc75776656)

[1.2 Making the cover and applying it 6](#_Toc75776657)

[1.3 To facilitate entering a thesis into DiVA – make a MODS file 6](#_Toc75776658)

[1.4 Requirements 7](#_Toc75776659)

[2 Can someone else use these programs? 7](#_Toc75776660)

[3 LaTeX template in Overleaf 7](#_Toc75776661)

[4 Put information into LaTeX template to generate a draft or final thesis 8](#_Toc75776662)

[5 Now that data is in the template, what happens? 14](#_Toc75776663)

[6 How did the abstracts and the keywords appear? 16](#_Toc75776664)

[7 Extracting the information from the PDF file 17](#_Toc75776665)

[8 What can we do with this event.json file? 19](#_Toc75776666)

[9 Actual example 25](#_Toc75776667)

[10 Change in how to enter the abstracts in LaTeX 28](#_Toc75776668)

[11 Adding keywords and PDF meta data 29](#_Toc75776669)

[12 Other variants of the JSON\_to\_calendar.py 34](#_Toc75776670)

[13 Making and applying covers 34](#_Toc75776671)

[14 JSON to MODS file 35](#_Toc75776672)

[14.1.1 Limitations 35](#_Toc75776673)

[14.1.2 Example 35](#_Toc75776674)

[15 Some enhancements 48](#_Toc75776675)

[15.1 Enhancements to template and supporting programs 48](#_Toc75776676)

[15.2 Better support for mathematical expressions in abstracts 48](#_Toc75776677)

[15.2.1 Better support for mathematics in Canvas course announcement and course calendar 48](#_Toc75776678)

[15.2.2 Better support for mathematics in Cortina 52](#_Toc75776679)

[15.2.3 Support for mathematics in DiVA 55](#_Toc75776680)

[15.3 Acronyms in abstracts 55](#_Toc75776681)

[15.4 URLs in abstracts 56](#_Toc75776682)

[15.5 Getting the PDF from a Canvas assignment and optionally extracting JSON 56](#_Toc75776683)

[15.6 Support for new cover 56](#_Toc75776684)

[16 Accessibility 57](#_Toc75776685)

[16.1 Calendar entries 57](#_Toc75776686)

[16.2 Cover and PDF file 57](#_Toc75776687)

[16.3 Template itself 58](#_Toc75776688)

[16.4 Improving accessibility 58](#_Toc75776689)

[17 The structure of the template and the report 58](#_Toc75776690)

[18 Generating LaTeX commands for in the student’s thesis source file 59](#_Toc75776691)

[19 Alternative way of inserting the covers 60](#_Toc75776692)

List of Figures

[Figure 1: Example of title page of a thesis 9](#_Toc75776693)

[Figure 2: Information about the title and the student authors of the thesis 10](#_Toc75776694)

[Figure 3: Supervisors and examiner information 11](#_Toc75776695)

[Figure 4: Information about where the thesis is taking place 12](#_Toc75776696)

[Figure 5: Information relevant to the oral presentation (both the location and the opponent or opponents) 12](#_Toc75776697)

[Figure 6: Information relevant to the oral presentation (both the location and the opponent or opponents) 12](#_Toc75776698)

[Figure 7: LaTeX to produce the English abstract (edited for formatting) 13](#_Toc75776699)

[Figure 8: LaTeX input to produce the Swedish abstract 14](#_Toc75776700)

[Figure 9: First part of the “For DiVA” information 15](#_Toc75776701)

[Figure 10: Second part of the “For DiVA” information 16](#_Toc75776702)

[Figure 11: Code in \divainfo command to output the abstracts and keywords 17](#_Toc75776703)

[Figure 12: Code in the document file to set up the command \typestoredx and to configure the listing environment to put the abstract into. 17](#_Toc75776704)

[Figure 13: Extract the JSON like information 17](#_Toc75776705)

[Figure 14: Example event.json output 18](#_Toc75776706)

[Figure 15: The several outputs of JSON\_to\_calendar.py 19](#_Toc75776707)

[Figure 16: Funning JSON\_to\_calendar.py to produce all three outputs 19](#_Toc75776708)

[Figure 17: Canvas course announcement 20](#_Toc75776709)

[Figure 18: A course event in the Canvas calendar (the figure is zoomed in on 15 March 2021 21](#_Toc75776710)

[Figure 19: Zoomed view of the opened Canvas calendar event 21](#_Toc75776711)

[Figure 20: KTH’s Cortina calendar showing two degree project events 22](#_Toc75776712)

[Figure 21: English version of the calendar 22](#_Toc75776713)

[Figure 22: Swedish version of the calendar event 23](#_Toc75776714)

[Figure 23: Response from the KTH Cortina API 24](#_Toc75776715)

[Figure 24: Actual example of announcement in Canvas (top) 25](#_Toc75776716)

[Figure 25: Bottom of the announcement 26](#_Toc75776717)

[Figure 26: opened version in course calendar 26](#_Toc75776718)

[Figure 27: The English (left) and Swedis (right) in the Cortina calendar 27](#_Toc75776719)

[Figure 28: Commands to extract the JSON and to make the calendar entries and announcements 27](#_Toc75776720)

[Figure 29: Extracted JSON file oscar.json - edited for appearance here 27](#_Toc75776721)

[Figure 30. Example of the revised format for entering an abstract 28](#_Toc75776722)

[Figure 31. Second example of the revised format for entering an abstract 29](#_Toc75776723)

[Figure 32: The two macros used to help enter the language specification 29](#_Toc75776724)

[Figure 33: New commands in kththesis.cls 30](#_Toc75776725)

[Figure 34: New additions to examplethesis.text 30](#_Toc75776726)

[Figure 35: lib/pdf\_related\_includes.tex (edited for readability) 31](#_Toc75776727)

[Figure 36: Including the English language keywords below the English abstract 32](#_Toc75776728)

[Figure 37: Including the Swedeish language keywords below the Swedish abstract 32](#_Toc75776729)

[Figure 38: The title page of the thesis and the PDF meta data 33](#_Toc75776730)

[Figure 39: The keywords appear as expected in the for DiVA data at the end of the PDF file 33](#_Toc75776731)

[Figure 40: Several possible inputs to JSON\_to\_calendar.py and its outputs 34](#_Toc75776732)

[Figure 41: Several possible used of the JSON file 34](#_Toc75776733)

[Figure 42: Examples of using JSON\_to\_MODS.py 35](#_Toc75776734)

[Figure 43: test12.json 36](#_Toc75776735)

[Figure 44: test12.mods 37](#_Toc75776736)

[Figure 45: The import process – step 1 – click on the “Import from external database” buttom on the upper right corner 38](#_Toc75776737)

[Figure 46: The import process – step 2 – choose the MODV3 format and select a file to import 39](#_Toc75776738)

[Figure 47: the import process – step 3 – after clicking “Import” – it says that the file was successfully uploaded 40](#_Toc75776739)

[Figure 48: The import process – step 4 – after clicking “Import” once more – you can see the file has been added to the list of imported files 40](#_Toc75776740)

[Figure 49: Imported document (part 1) 41](#_Toc75776741)

[Figure 50: 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. 42](#_Toc75776742)

[Figure 51: 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. 43](#_Toc75776743)

[Figure 52: Imported document (part 4) – more keywords are shown as well as the abstract in English 44](#_Toc75776744)

[Figure 53: 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. 45](#_Toc75776745)

[Figure 54: Imported document (part 6) – yet more abstracts 46](#_Toc75776746)

[Figure 55: Imported document (part 7) – the supervisors are shown 47](#_Toc75776747)

[Figure 56: Imported document (part 8) – the examiner is shown along with the information about the oral presentation 48](#_Toc75776748)

[Figure 57: Commands to produce the JSON and make the calendar entries and announcement 49](#_Toc75776749)

[Figure 58: Examples of equations in an announcement 50](#_Toc75776750)

[Figure 59: Announcement summary 51](#_Toc75776751)

[Figure 60: Announcement with equation 51](#_Toc75776752)

[Figure 61: HTML for the announcement 52](#_Toc75776753)

[Figure 62: Equations in Cortina claendar entry 53](#_Toc75776754)

[Figure 63: Zoom in on part of the Cortinal calendar entry 54](#_Toc75776755)

[Figure 64: Zoom in on lower part of the Cortinal calendar entry 55](#_Toc75776756)

[Figure 65: Specifying acronyms file 55](#_Toc75776757)

[Figure 66: get\_PDF\_submission program 56](#_Toc75776758)

[Figure 67: New fill\_in\_template.py program 56](#_Toc75776759)

[Figure 68: Files and folders in Overleaf template 59](#_Toc75776760)

[Figure 69: Generating LaTeX commands for a student author 59](#_Toc75776761)

[Figure 70: Generating LaTeX commands for a supervisor 60](#_Toc75776762)

[Figure 71: Generating LaTeX commands for the examiner 60](#_Toc75776763)

[Figure 72: Inseert this include file either in the main text document or the includes.tex file 60](#_Toc75776764)

[Figure 73: Insert the front cover before the title page 60](#_Toc75776765)

[Figure 74: Insert the back cover page before the For DIVA section 60](#_Toc75776766)

# 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.

## Making an announcement

Assuming that a student[[2]](#footnote-2) has submitted their thesis with the information in the For DIVA 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. You will get a file named modsXML.xml that you can inport into DiVA.

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, 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 not yet in production and requires a KTH Cortina Access Key.

Replace 11 with the course\_id of the Canvas course room, for example: 35434 for the EECS 2nd cycle degree projects. For more information see Section 12.

## Making the cover and applying it

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 you can make the covers and apply them. For example, for a file (oscar.pdf) with two (2) “For DIVA” pages:

./JSON\_to\_cover.py --json oscar.json --exam 4 --file oscar.pdf --diva 2

The above will generate the color cover using the KTH cover generator, remove the two “For DIVA” pages, and apply the cover, creating a file: oscar-with-cover.pdf . Note that this command needed the type of exam, in this case 4 indicates a 'Degree of Master of Science in Engineering' or 'Civilingenjörsexamen'. One could also use either of these strings to indicate the type of exam. You can include the TRITA number using the option “--trita "TRITA-EECS-EX-2021:xxxx" For more information see Section 13.

## 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 14.

## 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. To run the program to make the covers requires the program and a widely used utility called “qpdf” (<http://qpdf.sourceforge.net/>).

# 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 Canvas-related functionality. However, 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> .

See the examples of running the program in Sections 1.1 and 1.2.

# LaTeX template in Overleaf

The template used to make the examples in this document is at:

<https://www.overleaf.com/read/xmrfhcchgnvq>

**NB**: This is where I do experiments with the template and is not intended to be stable.

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 15 addressess issues about accessibility.

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

Figure 1 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.

Note that 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}.

Information about the title and the student author(s) of the thesis is entered via the set of macros as shown in Figure 2. 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.
* Note that ORCiD information is unnecessary, but if the student has an ORCiD identifier, it will be added to the record in DiVA. Student’s may find it useful to have such an identifier if they are going to go on to write conference, journal, and other publications in the future.
* 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.
* Finally, 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). It 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 1: 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}  % If the student has an ORCiD - add it here  \orcid{0000-0002-00001-1234}  \authorsSchool{\schoolAcronym{EECS}}  \programcode{TCOMK}  %% Alternatively, you can say \programme{Civilingenjör Datateknik} to directly set the programme string  % If there is a second author - add them here:  \secondAuthorsLastname{Student}  \secondAuthorsFirstname{Fake B.}  \secondemail{b@kth.se}  \secondkthid{u100002}  % If the student has an ORCiD - add it here  \secondorcid{0000-0002-00001-5678}  \secondAuthorsSchool{\schoolAcronym{ABE}} |

Figure 2: Information about the title and the student authors of the thesis

Figure 3 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 3: Supervisors and examiner information

Figure 4 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 4: Information about where the thesis is taking place

Figure 5 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 5: Information relevant to the oral presentation (both the location and the opponent or opponents)

Finally, Figure 6 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 6: Information relevant to the oral presentation (both the location and the opponent or opponents)

Figure 7 and Figure 8 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} |

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 7: 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 8: 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 9 and Figure 10 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 9: First part of the “For DiVA” information

|  |
| --- |
|  |

Figure 10: Second part of the “For DiVA” information

# 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 11). 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 12 – 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 11: Code in \divainfo command to output the abstracts and keywords

Figure 12 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 12: 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 13, to extract the information. Figure 14 shows an example of the resulting file.

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

Figure 13: 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 14: Example event.json output

# What can we do with this event.json file?

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

Once you are happy with the JSON file 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 15: The several outputs of JSON\_to\_calendar.py

We can produce all three outputs with the command shown in Figure 16. Note that the program was run with the event.json file to produce a Canvas course announcement, as shown in Figure 17. 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 16: Funning JSON\_to\_calendar.py to produce all three outputs

Note that as described in Section 1.2, 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 17: Canvas course announcement

Figure 17 and Figure 18 show the event in the Canvas calendar (when I have selected to display the events for course 11 in green).

|  |
| --- |
|  |

Figure 18: A course event in the Canvas calendar (the figure is zoomed in on 15 March 2021

|  |
| --- |
|  |

Figure 19: 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 21 and Figure 22 show the English and Swedish versions of the event in the calendar.s

|  |
| --- |
|  |

Figure 20: KTH’s Cortina calendar showing two degree project events

|  |
| --- |
|  |

Figure 21: English version of the calendar

|  |
| --- |
|  |

Figure 22: Swedish version of the calendar event

Figure 23 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 23: Response from the KTH Cortina API

# Actual example

This example appear with the student’s permission. Figure 24shows the announcement for a 2nd cycle thesis presentation in a Canvas course while Figure 25 shows the bottom part of the announcement. Figure 27 shows the Cortina calendar entry. Note that the Cortina calendar entry is in the development system and not the production calendar. Figure 28shows 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 29 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 24: Actual example of announcement in Canvas (top)

|  |
| --- |
|  |

Figure 25: Bottom of the announcement

|  |
| --- |
|  |

Figure 26: opened version in course calendar

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

Figure 27: 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 28: 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 29: 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 30 and Figure 31 show examples of how to enter an abstract and keywords while Figure 32 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 30. 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 31. 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 32: 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 33. 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 34) using a file called lib/pdf\_related\_includes.tex (shown in Figure 35). Later the keywords that have been stored are inserted into the LaTeX after their respective language abstracts as shown in Figure 36 and Figure 37. The title page of the thesis and the PDF meta data are shown in Figure 38 Finally, the keywords appear (as expected) in the for DiVA data at the end of the PDF file as shown in Figure 39.

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 33: New commands in kththesis.cls

Figure 34 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 34: 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 35: 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 36: 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 37: Including the Swedeish language keywords below the Swedish abstract

|  |
| --- |
|  |

Figure 38: The title page of the thesis and the PDF meta data

|  |
| --- |
|  |

Figure 39: 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 40. 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  ZIP of LaTeX  DOCX  PDF |

Figure 40: Several possible inputs to JSON\_to\_calendar.py and its outputs

# Making and applying covers

Once we have extracted the JSON from the “For DIVA” section at the end of the PDF file we can also make covers (as described in Section 1.2) and later make a MODS file for import into DiVA (along with the PDF file with covers)., see Figure 40.

|  |
| --- |
| TRITA number  Import into DiVA  PDF file with cover  JSON\_to\_cover.py  JSON\_to\_MODS.py  MODS file  JSON\_to\_calendar.py  Canvas announcement  Canvas calendar event  KTH calendar event  JSON file  PDF  Extractor |

Figure 41: Several possible used of the JSON file

Note that the PDF file can also be taken from Overleaf or from a submission to a Canvas course (see 15.5).

# 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 42: 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>

### Limitations

Currently, I do not seem to be able to get the KTH affiliations correctly entered. I think that it is because I do 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 have not found any example of how to do what they describe. All of my descriptions end up in the "Other university" field.

I also do not seem to be able 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>

So this remains a work in progress, but it certainly does a lot of the work of getting the abstracts, keywords, titles, etc.as meta data into DiVA.

### Example

The process 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 43. The resulting MODS file is shown in Figure 44.

|  |
| --- |
| {"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 43: 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 44: test12.mods

Figure 45 to Figure 48 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 45: The import process – step 1 – click on the “Import from external database” buttom on the upper right corner

|  |
| --- |
|  |

Figure 46: The import process – step 2 – choose the MODV3 format and select a file to import

|  |
| --- |
|  |

Figure 47: 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 48.

|  |
| --- |
|  |

Figure 48: 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 49 to Figure 56.

|  |
| --- |
|  |

Figure 49: Imported document (part 1)

|  |
| --- |
|  |

Figure 50: 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 51: 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 52: Imported document (part 4) – more keywords are shown as well as the abstract in English

|  |
| --- |
|  |

Figure 53: 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 54: Imported document (part 6) – yet more abstracts

|  |
| --- |
|  |

Figure 55: Imported document (part 7) – the supervisors are shown

|  |
| --- |
|  |

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

# Some enhancements

The enhancements are (§15.1) 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 15.2 enhancements to better support mathematical expressions in abstracts. Section 15.3 describes support for acronyms in abstracts. Section 15.4 describes the state of support for URLs in abstracts. Section 15.5 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 level of the degree, number of credits, the specific exam that the student is doing the degree project for, and the subject area (or areas in the case of a student doing both a Civ. Ing. Degree and a Master’s degree). Currently the program for making covers uses some guesses and some manually (and incomplete) configured data. Therefore, if these guesses are incorrect, then one has to explicitly provide the correct data in the command line for the program – this will be used by the program when making the cover.

## 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 58. These results and the results shown in Section 15.2.2 were generated using the commands in Figure 57.

|  |
| --- |
| ./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 57: Commands to produce the JSON and make the calendar entries and announcement

|  |
| --- |
|  |

Figure 58: 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 59. The announcement is shown in Figure 60 while Figure 61 shows the HTML for this announcement.

|  |
| --- |
|  |

Figure 59: Announcement summary

|  |
| --- |
|  |

Figure 60: 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 61: HTML for the announcement

### 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 62 shows the entry in the Cortina calendar, while Figure 63 and Figure 64 shows zoomed in views of the equations.

|  |
| --- |
|  |

Figure 62: Equations in Cortina claendar entry

|  |
| --- |
|  |

Figure 63: Zoom in on part of the Cortinal calendar entry

|  |
| --- |
|  |

Figure 64: 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 65. 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 65: 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 support 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 66. 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 66: 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 67.

|  |
| --- |
| ./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 67: New fill\_in\_template.py program

Note that the new template is net 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 computationof 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 not yet addressed the problem of making this available from within LaTeX (for the 1st alternative) or understand the font metrics enough to write the code for the 2nd alternative.

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[[3]](#footnote-3) 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 intself (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 68. 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 68: 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 69 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 : Generating LaTeX commands for a student author

Figure 70 shows the case for a supervisor who has the e-mail address “vastberg@kth.se”. Figure 71 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 : 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 : Generating LaTeX commands for the examiner

# Alternative way of inserting the covers

Another way that the covers can be inserted is to use the pdfpages package (as shown in Figure 72), then insert the two cover pages at the appropriate place (as shown in Figure 73 and Figure 74). 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 72: 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 73: 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 74: Insert the back cover page before the For DIVA section

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. I would loke to acknowledge the permission of Oscar Rosquist for the permission to use his thesis as an example in my documentation. [↑](#footnote-ref-2)
3. <https://www.readspeaker.com/education/learning-management-systems/> [↑](#footnote-ref-3)