

Titel der Arbeit

Optionaler Untertitel der Arbeit

BACHELORARBEIT

zur Erlangung des akademischen Grades

Bachelor of Science

im Rahmen des Studiums

Software und Information Engineering

eingereicht von

Kevin Haller

Matrikelnummer 1325694

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Pretitle Reka Marta Sabou, Posttitle

Mitwirkung: Pretitle Forename Surname, Posttitle

Pretitle Forename Surname, Posttitle

Pretitle Forename Surname, Posttitle

Wien, 1. August 2016

Kevin Haller

Reka Marta Sabou

Title

Optional Subtitle of the Thesis

BACHELOR'S THESIS

submitted in partial fulfillment of the requirements for the degree of

Bachelor of Science

in

Software and Information Engineering

by

Kevin Haller

Registration Number 1325694

to the Faculty of Informatics

at the TU Wien

Advisor: Pretitle Reka Marta Sabou, Posttitle

Assistance: Pretitle Forename Surname, Posttitle

Pretitle Forename Surname, Posttitle

Pretitle Forename Surname, Posttitle

Vienna, 1st August, 2016

Kevin Haller

Reka Marta Sabou

Erklärung zur Verfassung der Arbeit

Kevin Haller
Address

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 1. August 2016

Kevin Haller

Danksagung

TODO: Ihr Text hier.

Acknowledgements

TODO: Enter your text here.

Kurzfassung

TODO: Ihr Text hier.

Abstract

TODO: Enter your text here.

Contents

Kurzfassung	xi
Abstract	xiii
Contents	xv
1 Introduction	1
1.1 Motivation	1
1.2 Problem description	1
1.3 Structure of the thesis	2
2 Background	3
2.1 Semantic Web	3
2.2 RDF	3
2.3 Ontology	3
2.4 SPARQL	3
3 Related Work	5
3.1 Linked Open Data based map applications	5
3.2 Geospatial ontologies	5
3.3 Accessibility ontologies	6
3.4 Indoor modelling	6
4 Solution	7
4.1 Data sources	7
4.2 Prototype of ontology	7
4.3 Architectural prototype	7
4.4 Web application	8
5 Discussion	9
5.1 Evaluation	9
5.2 Further work	9
5.3 Conclusion	9

6	Introduction to L^AT_EX	11
6.1	Installation	11
6.2	Editors	11
6.3	Compilation	12
6.4	Basic Functionality	13
6.5	Bibliography	14
6.6	Table of Contents	15
6.7	Acronyms / Glossary / Index	15
6.8	Tips	15
6.9	Resources	16
	List of Figures	19
	List of Tables	19
	List of Algorithms	21
	Index	23
	Glossary	25
	Acronyms	27
	Bibliography	29

Introduction

1.1 Motivation

TODO: Intro Web of Documents, missed potential, isolated data sources

The same applies at the moment to the Vienna University of Technology. In order to find a special location like a certain lecture hall and how it can be accessed without obstacles like stairways, one has to search for information on web sites, scan floor plans and eventually construct a convenient route to the location based on the gathered knowledge. For humans this procedure does not constitute a problem, but think of machines. **TODO: Elaborate on problem for machines, Deep learning, NLP state of the art reference**

As a consequence the barrier for application developers is quite high due to the time consuming task of gathering and extracting the data from multiple sources and transforming it into a representation that is usable by the planned application.

TODO: Linked Data, goal of thesis

1.2 Problem description

TODO: more comprehensive description

- prototype of a map application based on the transformation of spatial data of the Vienna University of Technology into Linked Data to evaluate and discuss the potential benefits and caveats.
- prototype of an ontology to model spatial data as well as the indoor plans of facilities taking accessibility into consideration. In compliance to best practises terms of already existing ontologies shall be reused whenever possible.

- linking of the transformed spatial data with external sources like GeoNames or LinkedGeoData.

1.3 Structure of the thesis

This chapter outlined the motivation for writing this thesis and gave a description of the problems for which a solution will be suggested. The rest of this thesis is structured as follows: Chapter 2 discusses the basic concepts, principles and technologies that build the foundation of Linked Data and the Semantic Web. It is intended to be a brief introduction for readers that are not familiar with this topic. Chapter 3 provides a summary of past efforts in research and works related to the focus of this thesis. In the subsequent chapter 4 a solution for the given problems will be suggested and discussed. This includes the prototype of an ontology that models the problem domain as well as an architectural prototype of a system that manages a subset of spatial data about the Vienna University of Technology in form of Linked Data and provides a human- and machine-friendly interface to it. Finally, the implemented solution is evaluated and conclusions are drawn in the last chapter 5. It provides furthermore an outlook to future work and potential improvements.

CHAPTER 2

Background

2.1 Semantic Web

TODO: Enter your text here.

2.2 RDF

TODO: Enter your text here.

2.3 Ontology

TODO: Enter your text here.

2.3.1 RDFS

TODO: Enter your text here.

2.3.2 OWL

TODO: Enter your text here.

2.4 SPARQL

TODO: Enter your text here.

Related Work

3.1 Linked Open Data based map applications

TODO: Existing map application based on Linked Open Data, Southampton, Oxford, LODUM

3.2 Geospatial ontologies

TODO: Evaluation of existing spatial ontologies as related work for the designed ontology prototype

3.2.1 WGS84

...

3.2.2 ISA Programme Location Core Vocabulary

...

3.2.3 schema.org

...

3.2.4 NeoGeo Geometry Ontology

...

3.2.5 GeoSPARQL

...

3.3 Accessibility ontologies

TODO: Evaluation of accessibility ontologies as related work for the designed ontology prototype

3.4 Indoor modelling

TODO: Indoor modelling research

CHAPTER 4

Solution

4.1 Data sources

TODO: A description of the data sources and the transformation process

4.2 Prototype of ontology

TODO: Description of the developed ontology with newly created terms and reused ones, and why using GeoSPARQL and what are the disadvantages

4.3 Architectural prototype

TODO: A description of the system, the different layers, integration, linking and web framework, REST API, used technologies

4.3.1 System architecture

TODO: Description of the system architecture

4.3.2 RDF Framework

TODO: Reasoning why using Sesame, not Apache Jena

4.3.3 Linked Data management

TODO: Description of the data management and compliance to best practise

4.3.4 Linking

TODO: Description of strategies for linking with external sources and to enrich the data

4.3.5 Linked Data publishing

TODO: Description of the data publishing and compliance to best practise

4.4 Web application

TODO: A more comprehensive presentation of the user interface for humans to show the targeted use case.

CHAPTER 5



Discussion

5.1 Evaluation

TODO: Enter your text here.

5.2 Further work

TODO: Enter your text here.

5.3 Conclusion

TODO: Enter your text here.

Introduction to L^AT_EX

Since L^AT_EX is widely used in academia and industry, there exists a plethora of freely accessible introductions to the language. Reading through the guide at <https://en.wikibooks.org/wiki/LaTeX> serves as a comprehensive overview for most of the functionality and is highly recommended before starting with a thesis in L^AT_EX.

6.1 Installation

A full L^AT_EX distribution consists of not only of the binaries that convert the source files to the typeset documents, but also of a wide range of packages and their documentation. Depending on the operating system, different implementations are available as shown in Table 6.1. **Due to the large amount of packages that are in everyday use and due to their high interdependence, it is paramount to keep the installed distribution up to date.** Otherwise, obscure errors and tedious debugging ensue.

6.2 Editors

A multitude of T_EX editors are available differing in their editing models, their supported operating systems and their feature sets. A comprehensive overview of editors can be

Distribution	Unix	Windows	MacOS
TeX Live	yes	yes	(yes)
MacTeX	no	no	yes
MikTeX	no	yes	no

Table 6.1: T_EX/L^AT_EX distributions for different operating systems. Recommended choice in **bold**.

Description	
1	Scan for refs, toc/lof/lot/loa items and cites
2	Build the bibliography
3	Link refs and build the toc/lof/lot/loa
4	Link the bibliography
5	Build the glossary
6	Build the acronyms
7	Build the index
8	Link the glossary, acronyms, and the index
9	Link the bookmarks
Command	
1	<code>pdflatex.exe example</code>
2	<code>bibtex.exe example</code>
3	<code>pdflatex.exe example</code>
4	<code>pdflatex.exe example</code>
5	<code>makeindex.exe -t example.glg -s example.ist</code> <code>-o example.gls example.glo</code>
6	<code>makeindex.exe -t example.alg -s example.ist</code> <code>-o example.acr example.acn</code>
7	<code>makeindex.exe -t example.ilg -o example.ind example.idx</code>
8	<code>pdflatex.exe example</code>
9	<code>pdflatex.exe example</code>

Table 6.2: Compilation steps for this document. The following abbreviations were used: table of contents (toc), list of figures (lof), list of tables (lot), list of algorithms (loa).

found at the Wikipedia page https://en.wikipedia.org/wiki/Comparison_of_TeX_editors. TeXstudio (<http://texstudio.sourceforge.net/>) is recommended.

6.3 Compilation

Modern editors usually provide the compilation programs to generate Portable Document Format (PDF) documents and for most L^AT_EX source files, this is sufficient. More advanced L^AT_EX functionality, such as glossaries and bibliographies, needs additional compilation steps, however. It is also possible that errors in the compilation process invalidate intermediate files and force subsequent compilation runs to fail. It is advisable to delete intermediate files (`.aux`, `.bbl`, etc.), if errors occur and persist. All files that are not generated by the user are automatically regenerated. To compile the current document, the steps as shown in Table 6.2 have to be taken.

6.4 Basic Functionality

In this section, various examples are given of the fundamental building blocks used in a thesis. Many \LaTeX commands have a rich set of options that can be supplied as optional arguments. The documentation of each command should be consulted to get an impression of the full spectrum of its functionality.

6.4.1 Floats

Two main categories of page elements can be differentiated in the usual \LaTeX workflow: *(i)* the main stream of text and *(ii)* floating containers that are positioned at convenient positions throughout the document. In most cases, tables, plots, and images are put into such containers since they are usually positioned at the top or bottom of pages. These are realized by the two environments `figure` and `table`, which also provide functionality for cross-referencing (see Table 6.3 and Figure 6.1) and the generation of corresponding entries in the list of figures and the list of tables. Note that these environments solely act as containers and can be assigned arbitrary content.

6.4.2 Tables

A table in \LaTeX is created by using a `tabular` environment or any of its extensions, e.g., `tabularx`. The commands `\multirow` and `\multicolumn` allow table elements to span multiple rows and columns.

Position		
Group	Abbrev	Name
Goalkeeper	GK	Paul Robinson
Defenders	LB	Lucas Radebe
	DC	Michael Duburrry
	DC	Dominic Matteo
	RB	Didier Domi
Midfielders	MC	David Batty
	MC	Eirik Bakke
	MC	Jody Morris
Forward	FW	Jamie McMaster
Strikers	ST	Alan Smith
	ST	Mark Viduka

Table 6.3: Adapted example from <https://en.wikibooks.org/wiki/LaTeX/Tables>. This example uses rules specific to the `booktabs` package and employs the multi-row functionality of the `multirow` package.

6.4.3 Images

An image is added to a document via the `\includegraphics` command as shown in Figure 6.1. The `\subcaption` command can be used to reference subfigures, such as Figure 6.1a and 6.1b.

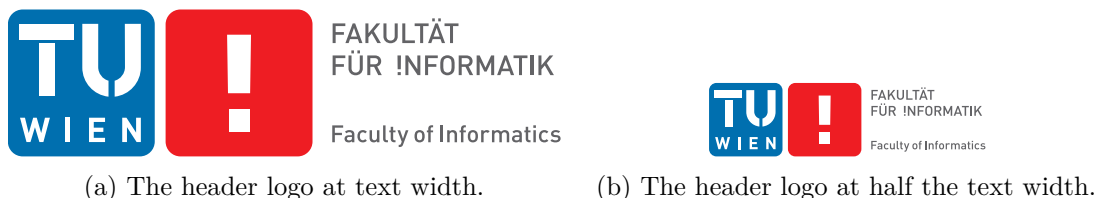


Figure 6.1: The header logo at different sizes.

6.4.4 Mathematical Expressions

One of the original motivation to create the T_EX system was the need for mathematical typesetting. To this day, L^AT_EX is the preferred system to write math-heavy documents and a wide variety of functions aids the author in this task. A mathematical expression can be inserted inline as $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$ outside of the text stream as

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$$

or as numbered equation with

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}. \tag{6.1}$$

6.4.5 Pseudo Code

The presentation of algorithms can be achieved with various packages, such as `algorithmic`, `algorithm2e`, `algorithmicx`, or `algpseudocode`. See <https://tex.stackexchange.com/questions/229355> for an overview. An example of the use of the `algorithm2e` package is given with Algorithm 6.1.

6.5 Bibliography

The referencing of prior work is a fundamental requirement of academic writing and well supported by L^AT_EX. The B_IB_TE_X reference management software is the most commonly used system for this purpose. Using the `\cite` command, it is possible to reference entries in a `.bib` file out of the text stream, e.g., as [Tur36]. The generation of the formatted bibliography needs a separate execution of `bibtex.exe` (see Table 6.2).

Algorithm 6.1: Gauss-Seidel

Input: A scalar ϵ , a matrix $\mathbf{A} = (a_{ij})$, a vector \vec{b} , and an initial vector $\vec{x}^{(0)}$

Output: $\vec{x}^{(n)}$ with $\mathbf{A}\vec{x}^{(n)} \approx \vec{b}$

```
1 for  $k \leftarrow 1$  to maximum iterations do
2   for  $i \leftarrow 1$  to  $n$  do
3      $x_i^{(k)} = \frac{1}{a_{ii}} \left( b_i - \sum_{j < i} a_{ij} x_j^{(k)} - \sum_{j > i} a_{ij} x_j^{(k-1)} \right);$ 
4   end
5   if  $|\vec{x}^{(k)} - \vec{x}^{(k-1)}| < \epsilon$  then
6     break for;
7   end
8 end
9 return  $\vec{x}^{(k)}$ ;
```

6.6 Table of Contents

The table of contents is automatically built by successive runs of the compilation, e.g., of `pdflatex.exe`. The command `\setsecnumdepth` allows the specification of the depth of the table of contents and additional entries can be added via `\addcontentsline`. The starred versions of the sectioning commands, i.e., `\chapter*`, `\section*`, etc., remove the corresponding entry from the table of contents.

6.7 Acronyms / Glossary / Index

The list of acronyms, the glossary, and the index need to be built with a separate execution of `makeindex` (see Table 6.2). Acronyms have to be specified with `\newacronym` while glossary entries use `\newglossaryentry`. Both are then used in the document content with one of the variants of `\gls`, such as `\Gls`, `\glspl`, or `\Glspl`. Index items are simply generated by placing `\index{<entry>}` next to all the words that correspond to the index entry `<entry>`. Note that many enhancements exist for these functionalities and the documentation of the `makeindex` and the `glossaries` packages should be consulted.

6.8 Tips

Since \TeX and its successors do not employ a What You See Is What You Get (WYSIWYG) editing scheme, several guidelines improve the readability of the source content:

- Each sentence in the source text should start with a new line. This helps not only the user navigation through the text, but also enables revision control systems (e.g. Subversion (SVN), Git) to show the exact changes authored by different users. Paragraphs are separated by one (or more) empty lines.

- Environments, which are defined by a matching pair of `\begin{name}` and `\end{name}`, can be indented by whitespace to show their hierarchical structure.
- In most cases, the explicit use of whitespace (e.g. `\hspace{4em}` or `\vspace{1.5cm}`) violates typographic guidelines and rules. Explicit formatting should only be employed as a last resort and, most likely, better ways to achieve the desired layout can be found by a quick web search.
- The use of bold or italic text is generally not supported by typographic considerations and the semantically meaningful `\emph{...}` should be used.

The predominant application of the L^AT_EX system is the generation of PDF files via the PDFL^AT_EX binaries. In the current version of PDFL^AT_EX, it is possible that absolute file paths and user account names are embedded in the final PDF document. While this poses only a minor security issue for all documents, it is highly problematic for double blind reviews. The process shown in Table 6.4 can be employed to strip all private information from the final PDF document.

	Command
1	Rename the PDF document <code>final.pdf</code> to <code>final.ps</code> .
2	Execute the following command: <pre>ps2pdf -dPDFSETTINGS#/prepress ^ -dCompatibilityLevel#1.4 ^ -dAutoFilterColorImages#false ^ -dAutoFilterGrayImages#false ^ -dColorImageFilter#/FlateEncode ^ -dGrayImageFilter#/FlateEncode ^ -dMonoImageFilter#/FlateEncode ^ -dDownsampleColorImages#false ^ -dDownsampleGrayImages#false ^ final.ps final.pdf</pre>
	On Unix-based systems, replace <code>#</code> with <code>=</code> and <code>^</code> with <code>\</code> .

Table 6.4: Anonymization of PDF documents.

6.9 Resources

6.9.1 Useful Links

In the following, a listing of useful web resources is given.

<https://en.wikibooks.org/wiki/LaTeX> An extensive wiki-based guide to L^AT_EX.

<http://www.tex.ac.uk/faq> A (huge) set of Frequently Asked Questions (FAQ) about T_EX and L^AT_EX.

<https://tex.stackexchange.com/> The definitive user forum for non-trivial L^AT_EX-related questions and answers.

6.9.2 Comprehensive TeX Archive Network (CTAN)

The CTAN is the official repository for all T_EX related material. It can be accessed via <https://www.ctan.org/> and hosts (among other things) a huge variety of packages that provide extended functionality for T_EX and its successors. Note that most packages contain PDF documentation that can be directly accessed via CTAN.

In the following, a short, non-exhaustive list of relevant CTAN-hosted packages is given together with their relative path.

algorithm2e Functionality for writing pseudo code.

amsmath Enhanced functionality for typesetting mathematical expressions.

amssymb Provides a multitude of mathematical symbols.

booktabs Improved typesetting of tables.

enumitem User control over the layout of lists (`itemize`, `enumerate`, `description`).

fontenc Determines font encoding of the output.

glossaries Create glossaries and list of acronyms.

graphicx Insert images into the document.

inputenc Determines encoding of the input.

l2tabu A description of bad practices when using L^AT_EX.

mathtools Further extension of mathematical typesetting.

memoir The document class on upon which the `vutinfth` document class is based.

multirow Allows table elements to span several rows.

pgfplots Function plot drawings.

pgf/TikZ Creating graphics inside L^AT_EX documents.

subcaption Allows the use of subfigures and enables their referencing.

symbols/comprehensive A listing of around 5000 symbols that can be used with L^AT_EX.

voss-mathmode A comprehensive overview of typesetting mathematics in L^AT_EX.

xcolor Allows the definition and use of colors.

List of Figures

6.1	The header logo at different sizes.	14
-----	---	----

List of Tables

6.1	<code>TeX/LaTeX</code> distributions for different operating systems. Recommended choice in bold	11
6.2	Compilation steps for this document. The following abbreviations were used: table of contents (toc), list of figures (lof), list of tables (lot), list of algorithms (loa).	12
6.3	Adapted example from https://en.wikibooks.org/wiki/LaTeX/Tables . This example uses rules specific to the <code>booktabs</code> package and employs the multi-row functionality of the <code>multirow</code> package.	13
6.4	Anonymization of PDF documents.	16

List of Algorithms

6.1	Gauss-Seidel	15
-----	------------------------	----

Index

distribution, 11

Glossary

editor A text editor is a type of program used for editing plain text files.. 5

Acronyms

CTAN Comprehensive TeX Archive Network. 11

FAQ Frequently Asked Questions. 11

PDF Portable Document Format. 6, 10, 11, 13

SVN Subversion. 9

WYSIWYG What You See Is What You Get. 9

Bibliography

- [Tur36] Alan Mathison Turing. On computable numbers, with an application to the entscheidungsproblem. *J. of Math*, 58:345–363, 1936.