



Studienarbeit - Entwurf

Ausarbeitung InfluxDB und Wetterdaten

Erstellt von:

Henrik Mertens
Hatzfelder Str 25
33104 Paderborn

Prüfer:

Prof. Dr. Ulrich. Reus

Eingereicht am:

10. Mai 2022

Inhaltsverzeichnis

| | |
|--|----------|
| Abkürzungsverzeichnis | IV |
| Abbildungsverzeichnis | V |
| Tabellenverzeichnis | VI |
| Listingverzeichnis | VII |
| 1 Einleitung | 1 |
| 1.1 Zielsetzung | 1 |
| 1.2 Aufbau und Vorgehensweise | 1 |
| 2 Grundlagen | 2 |
| 2.1 Verbreitete Datenbank Management System (DBMS) | 2 |
| 2.2 Time Series Data | 3 |
| 2.3 Funktion Time Series Database | 3 |
| 2.4 Notizen | 3 |
| 2.5 Schrift | 4 |
| 2.5.1 Schriftgrößen | 4 |
| 2.5.2 Schrift Typen | 5 |
| 2.5.3 Schrift Ausrichtung | 5 |
| 2.6 Abbildungen | 5 |
| 2.7 Tabellen | 6 |
| 2.8 Zitate | 6 |
| 2.9 Abkürzungen | 7 |
| 2.10 Listen | 7 |
| 2.11 Quelltext | 8 |
| 3 Installation | 9 |
| 3.1 TeX-Distribution | 9 |
| 3.1.1 Windows | 9 |
| 3.1.2 Linux | 9 |
| 3.1.3 Mac-OS | 9 |
| 3.2 PDF-Viewer | 10 |
| 3.2.1 Windows | 10 |

| | | |
|----------|---------------------------------|-----------|
| 3.2.2 | Linux und Mac-OS | 10 |
| 3.3 | Hello World | 10 |
| 3.4 | Literaturverwaltung | 10 |
| 3.5 | Texteditor | 11 |
| 3.6 | PDF-Erzeugung | 11 |
| 4 | Zusammenfassung | 13 |
| | Anhang | 14 |
| | Quellenverzeichnis | 16 |
| | Ehrenwörtliche Erklärung | 18 |

Abkürzungsverzeichnis

AES Advanced Encryption Standard.

DBMS Datenbank Management System.

DWD Deutscher Wetterdienst.

TSDB Time Series Database.

Abbildungsverzeichnis

| | |
|---|----|
| Abbildung 1: DB-Engines Ranking | 2 |
| Abbildung 2: Das Logo der FHDW | 6 |
| Abbildung 3: Mendeley Referenzmanager | 11 |
| Abbildung 4: Sublime Texteditor | 12 |

Tabellenverzeichnis

| | |
|---|---|
| Tabelle 1: Die am häufigsten verwendeten PINs | 6 |
|---|---|

Listingverzeichnis

| | |
|--|---|
| Listing 1: Scanning for Wi-Fi Access Points on Android | 8 |
|--|---|

1 Einleitung

1.1 Zielsetzung

Das Ziel dieser Arbeit ist es eine Einführung in die Funktion von Time Series Database (TSDB) zu geben. Außerdem soll beispielhaft an InfluxDB gezeigt werden wie mit einer TSDB gearbeitet wird. Dazu werden die Wetterdaten vom Deutscher Wetterdienst (DWD) Importiert und ausgewertet.

1.2 Aufbau und Vorgehensweise

Im ersten Teil dieser Arbeit werden die Grundlagen von TSDB erklärt und Besonderheiten beschreiben. Im darauf Folgenden Kapitel wird dann exemplarisch an InfluxDB gezeigt wie mit einer TSDB gearbeitet wird.

Im letzten Kapitel werden die Inhalte dieser Arbeit zusammengefasst.

2 Grundlagen

2.1 Verbreitete DBMS

Aktuell gibt es wie in Abb. 1 zu sehen einige beliebte Multi-Model Datenbanken die als TSDB genutzt werden können. So können die Datenbanken MongoDB, Redis, Teradata und Couchbase mit Time Series Daten arbeiten. Die erste reine TSDB im Ranking ist InfluxDB auf Platz 29.¹

Abbildung 1: DB-Engines Ranking

394 Systeme im Ranking, Mai 2022

| Rang | | | DBMS | Datenbankmodell | Punkte | | |
|----------|----------|----------|------------------------------|---------------------------|----------|----------|----------|
| Mai 2022 | Apr 2022 | Mai 2021 | | | Mai 2022 | Apr 2022 | Mai 2021 |
| 1. | 1. | 1. | Oracle + | Relational, Multi-Model | 1262,82 | +8,00 | -7,12 |
| 2. | 2. | 2. | MySQL + | Relational, Multi-Model | 1202,10 | -2,06 | -34,28 |
| 3. | 3. | 3. | Microsoft SQL Server + | Relational, Multi-Model | 941,20 | +2,74 | -51,46 |
| 4. | 4. | 4. | PostgreSQL + | Relational, Multi-Model | 615,29 | +0,83 | +56,04 |
| 5. | 5. | 5. | MongoDB + | Document, Multi-Model | 478,24 | -5,14 | -2,78 |
| 6. | 6. | ↑ 7. | Redis + | Key-value, Multi-Model | 179,02 | +1,41 | +16,85 |
| 7. | ↑ 8. | ↓ 6. | IBM Db2 | Relational, Multi-Model | 160,32 | -0,13 | -6,34 |
| 8. | ↓ 7. | 8. | Elasticsearch + | Suchmaschine, Multi-Model | 157,69 | -3,14 | +2,34 |
| 9. | 9. | ↑ 10. | Microsoft Access | Relational | 143,44 | +0,66 | +28,04 |
| 10. | 10. | ↓ 9. | SQLite + | Relational | 134,73 | +1,94 | +8,04 |
| 11. | 11. | 11. | Cassandra + | Wide column | 118,01 | -3,98 | +7,08 |
| 12. | 12. | 12. | MariaDB + | Relational, Multi-Model | 111,13 | +0,81 | +14,44 |
| 13. | 13. | 13. | Splunk | Suchmaschine | 96,35 | +1,11 | +4,24 |
| 14. | 14. | ↑ 27. | Snowflake + | Relational | 93,51 | +4,06 | +63,46 |
| 15. | 15. | 15. | Microsoft Azure SQL Database | Relational, Multi-Model | 85,33 | -0,45 | +14,88 |
| 16. | 16. | 16. | Amazon DynamoDB + | Multi-Model | 84,46 | +1,55 | +14,39 |
| 17. | 17. | ↓ 14. | Hive + | Relational | 81,61 | +0,18 | +5,42 |
| 18. | 18. | ↓ 17. | Teradata + | Relational, Multi-Model | 68,39 | +0,82 | -1,59 |
| 19. | 19. | 19. | Neo4j + | Graph | 60,14 | +0,62 | +7,91 |
| 20. | 20. | 20. | Solr | Suchmaschine, Multi-Model | 57,26 | -0,48 | +6,07 |
| 21. | 21. | ↓ 18. | SAP HANA + | Relational, Multi-Model | 55,09 | -0,71 | +2,33 |
| 22. | 22. | 22. | FileMaker | Relational | 52,27 | -0,64 | +5,55 |
| 23. | ↑ 24. | ↑ 24. | Google BigQuery + | Relational | 48,61 | +0,63 | +10,98 |
| 24. | | | Databricks | Multi-Model | 47,85 | | |
| 25. | ↓ 23. | ↓ 21. | SAP Adaptive Server | Relational, Multi-Model | 47,78 | -0,58 | -2,19 |
| 26. | ↓ 25. | ↓ 23. | HBase + | Wide column | 43,19 | -1,14 | -0,05 |
| 27. | ↓ 26. | ↓ 25. | Microsoft Azure Cosmos DB + | Multi-Model | 40,22 | -0,12 | +5,51 |
| 28. | ↓ 27. | 28. | PostGIS | Spatial DBMS, Multi-Model | 31,82 | -0,23 | +1,98 |
| 29. | ↓ 28. | 29. | InfluxDB + | Time Series, Multi-Model | 29,55 | -0,47 | +2,38 |
| 30. | ↓ 29. | ↓ 26. | Couchbase + | Document, Multi-Model | 28,38 | -0,67 | -1,85 |

Quelle: <https://db-engines.com/de/ranking?msclkid=4f2a29e5d08811ec95ccd74f8f5146ab>

¹vgl. solid-IT-gmbh (2022)

- Ranking TSDB
- Graph InfluxDB vs MySQL

2.2 Time Series Data

Wie der Name schon sagt ist Time Series Data eine Reihe von Daten die über einen Zeitraum gesammelt worden sind. Es wird also nicht nur der Endwert aufgezeichnet sondern die Veränderung über einen Zeitraum. Diese Daten können z.B. Servermetriken, Anwendungs Leistungsüberwachung, Netzwerkdaten, Sensordaten, Ereignisse, Klicks, Marktgeschäfte und viele andere Arten von Daten sein. Time Series Data können gut daran erkannt, dass die Zeit eine Wichtige Axe bei der Darstellung der Werte ist.²

Manchmal ist es nicht notwendig alle Daten zu erfassen. Zum Beispiel wird in vielen Anwendungen nur der letzte Login gespeichert und mehr ist auch für die Funktion nicht notwendig. Allerdings können zusätzliche Informationen gewonnen werden wenn nicht nur die letzten Daten sondern die Veränderung aufgezeichnet werden. So kann zum Beispiel festgestellt werden wie oft und wann sich der Kunde einloggt und ob es dabei ein Muster gibt. Anhand dieser Daten können Kunden dann Kategorisiert werden.³

Eine Besonderheit von Time Series Data ist das sie sich nicht verändert. Wenn die Daten einmal erfasst wurden wird an ihnen nichts mehr geändert. Es werden nur neue Daten hinzugefügt⁴

2.3 Funktion Time Series Database

2.4 Notizen

sind dabei immer dann sinnvoll wenn Time Series Data gespeichert werden soll.⁵

²vgl. Dix, Paul (2021)

³vgl. Data-Science-Team (2020)

⁴vgl. Fangman, Sam (2019)

⁵vgl. Data-Science-Team (2020)

Diese Daten zeichnen sich besonders dadurch aus das die Zeit eine wichtiger Faktor für diese Daten ist.

- Eine Reihe von Daten - Timestamp als wichtiger Wert - Meistens werden nur neue Daten hinzugefügt. - Im vergleich zu einer MySQL Datenbank gibt es nur Inserts und keine Updates - Es wird die Veränderung über die Zeit Aufgezeichnet.- Time Series Data kann in einer SQL Datenbank gespeichert werden. Diese Datenbanken skalieren aber nur schlecht.

Die Meisten TSDB haben zusätzliche fUNKTIONEN Um mit den Daten zu arbeiten. - alte Daten löschen - durchgehende Abfragen

Daten: - Server Monitoring - Sensordaten - Wissenschaftliche Daten - Industrielle Daten / Events - Log Data

- MySQL -> Beim Login wird LastLogin geupdatete - TSDB -> Beim Login wird ein neuer Datenpunkt erstellt. -> Mehr Informationen werden beibehalten. Wie oft eingeloggt. Wann eingeloggt u.s.w.

- Autos Sammeln 4,000 GB pro Tag

Use Cases - Software Monitoring - Hardware Überwachung - IOT - Wetterdaten - Raumtemperatur - Positionsaufzeichnung von Autos,Containern,Paletten - Eventaufzeichnungen wie Benutzerinteraktion - Business intelligence Überwachen von Kennzahlen - alles weitere wo Zeitrelevante Daten anfallen.

Eine große Menge von Daten ist sehr stark an die Zeit gebunden in der sie erzeugt werden. So sind z.B. Sensordaten sehr eng mit der Aufzeichnungszeit verbunden. Weil sich die Werte sofort wieder verändern können.

2.5 Schrift

2.5.1 Schriftgrößen

Das ist sehr kleine Schrift

Das ist kleine Schrift

Das ist normale Schrift

Das ist große Schrift

Das ist größere Schrift
Das ist noch größere Schrift
Das ist riesige Schrift
Das ist noch riesigere Schrift

Das ist Script Schrift

Das ist Fußnoten Schrift

2.5.2 Schrift Typen

Das ist ein fetter Text

Das ist ein kursiver Text

Das ist ein unterstrichener Text

DAS IST EIN KAPITÄLCHEN TEXT

Das ist ein serifenloser Text

Das ist ein Schreibmaschinen Text

Das ist ein normaler Text

2.5.3 Schrift Ausrichtung

Quote Text (Der gesamte Text innerhalb der Umgebung wird von beiden Seiten eingerückt)

Zentrierter Text (Der gesamte Text innerhalb der Umgebung wird zentriert)

Linksbündiger Text (Der gesamte Text innerhalb der Umgebung wird linksbündig)

Rechtsbündiger Text (Der gesamte Text innerhalb der Umgebung wird rechtsbündig)

In einer Fußnote⁶

2.6 Abbildungen

In Abb. 2 sehen Sie das Logo der FHDW.

⁶können zusätzliche Ergänzungen, Präzisierungen, Textverweise usw. eingeführt werden.

Abbildung 2: Das Logo der FHDW

Quelle: Eigene Darstellung

2.7 Tabellen

In Tabelle 1 auf Seite 6 sehen Sie die am häufigsten benutzten PINs.

Tabelle 1: Die am häufigsten verwendeten PINs

| Rank | PIN | Percentage | Accumulated |
|------|------|------------|-------------|
| 1 | 1234 | 4.34% | 4.34% |
| 2 | 0000 | 2.57% | 6.91% |
| 3 | 2580 | 2.32% | 9.23% |
| 4 | 1111 | 1.60% | 10.83% |
| 5 | 5555 | 0.87% | 12.70% |
| 6 | 5683 | 0.70% | 12.39% |
| 7 | 0852 | 0.60% | 12.99% |
| 8 | 2222 | 0.56% | 13.55% |
| 9 | 1212 | 0.49% | 14.03% |
| 10 | 1998 | 0.43% | 14.46% |

Quelle: Eigene Darstellung

2.8 Zitate

Ein Zitat im Fließtext ist zu sehen bei Füller, Nüßer und Rustemeyer (2012).

Ein vergleichendes Zitat.⁷

⁷vgl. Maslennikov, Denis (2011), S. 5 ff.

Ein „wörtliches Zitat“⁸

Zitat einer Quelle mit mehreren Autoren.⁹

2.9 Abkürzungen

Bei der ersten Verwendung werden Abkürzungen ausgeschreiben: Advanced Encryption Standard (AES). Später wird dann automatisch nur noch die Kurzform benutzt: AES

2.10 Listen

Eine einfache List mit Punkten:

- Punkt 1
- Punkt 2
- Punkt 3

Eine einfache Liste mit Nummern:

1. Punkt 1
2. Punkt 2
3. Punkt 3

Eine einfache Liste mit römischen Nummern:

- I. Punkt 1
- II. Punkt 2
- III. Punkt 3

Eine einfache Liste mit Buchstaben:

- (a) Punkt 1
- (b) Punkt 2
- (c) Punkt 3

⁸Meier, Reto (2010), S. 13 f.

⁹vgl. Hocking, Christopher G. u. a. (2010), S. 32 ff.

2.11 Quelltext

Listing 1 auf Seite 8 zeigt einigen Quelltext.

Listing 1: Scanning for Wi-Fi Access Points on Android

```
registerReceiver(new RSSIBroadcastReceiver(),
    new IntentFilter(WifiManager.SCAN_RESULTS_AVAILABLE_ACTION));

WifiManager wifi = getSystemService(Context.WIFI_SERVICE);
wifi.startScan();

/* not thread safe */
public class RSSIBroadcastReceiver extends BroadcastReceiver {

    public void onReceive(Context context, Intent intent) {
        WifiManager wifi = getSystemService(Context.WIFI_SERVICE);
        List<ScanResult> scanResults = wifiManager.getScanResults();

        for (ScanResult scanResult : results) {
            RSSI rssi = new RSSI();
            rssi.bssi = scanResult.BSSID;
            rssi.signalLevel = scanResult.level;
        }
    }
}
```

3 Installation

3.1 TeX-Distribution

Für die Arbeit mit \LaTeX ist eine aktuelle TeX-Distribution erforderlich.

3.1.1 Windows

Unter Windows ist MiKTeX die Standard- \LaTeX -Distribution. Der MikTeX-Installer kann unter <http://miktex.org/download> heruntergeladen werden.

3.1.2 Linux

Die Standard- \LaTeX -Distribution unter Linux ist Tex Live, welche über die gängigen Software-Repositories installiert werden kann.

Unter Debian/Ubuntu kann die Installation der erforderlichen Pakete mittels der folgenden Befehle durchgeführt werden:

```
sudo apt-get install texlive-latex-base
sudo apt-get install texlive-latex-recommended
sudo apt-get install texlive-fonts-recommended
sudo apt-get install biblatex
sudo apt-get install biber
```

3.1.3 Mac-OS

Von der Tex-User-Group wird jährlich ein komplettes aktuelles MacTeX-Paket angeboten (<http://www.tug.org/mactex/index.html>), in dem alle relevanten Programme und Pakete enthalten sind.

3.2 PDF-Viewer

3.2.1 Windows

Als PDF-Viewer unter Windows bietet sich der freie Sumatra PDF Viewer an: <http://blog.kowalczyk.info/software/sumatrapdf/download-free-pdf-viewer-de.html>

3.2.2 Linux und Mac-OS

Die installierten Standard-PDF-Viewer unter Linux bzw. Mac-OS können problemlos genutzt werden.

3.3 Hello World

Nach der Installation sollte ein erster Test der Vorlage versucht werden. Dazu öffnen Sie ein Kommandozeilenfenster und wechseln in das Verzeichnis, in dem sich die \LaTeX -Quellen dieser Vorlage befinden. Anschließend müssen auf der Kommandozeile die Befehle

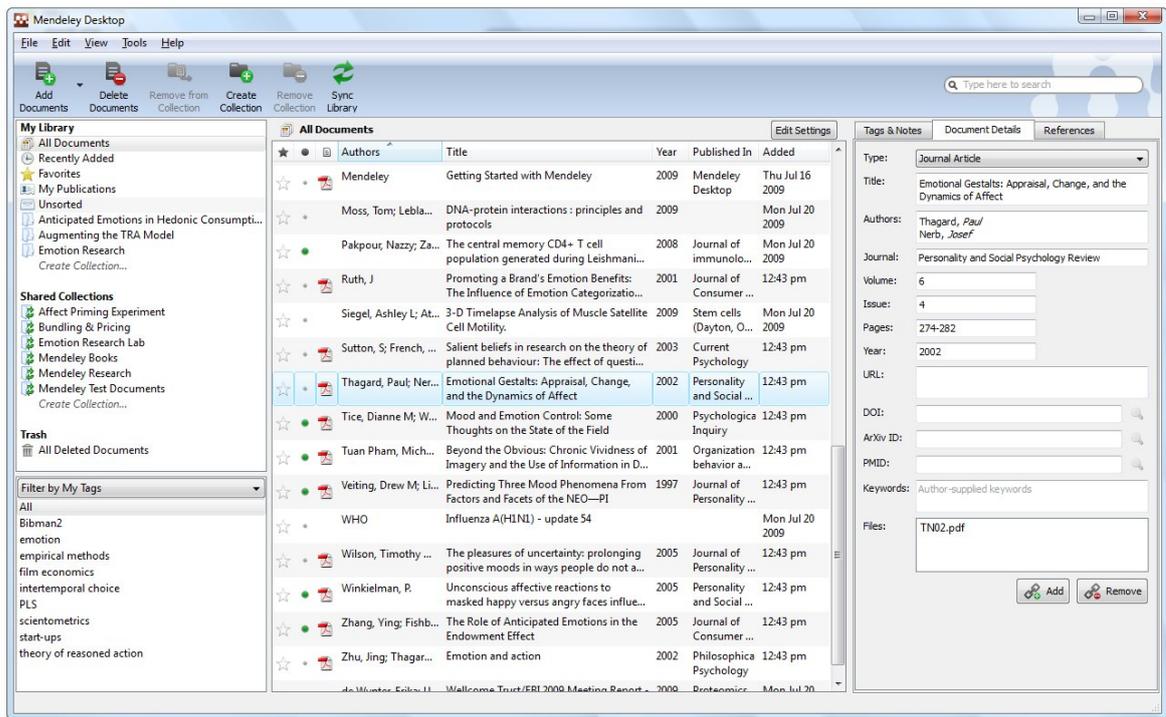
```
biber Thesis
pdflatex Thesis
```

einggegeben werden. Nun sollte eine neue Datei `Thesis.pdf` erzeugt worden sein. Falls nicht, sehen Sie bitte in den Ausgaben nach, die \LaTeX erzeugt hat. Diese sind recht umfangreich, auch wenn ein PDF-Dokument erzeugt werden konnte.

3.4 Literaturverwaltung

Für die Verwaltung von Quellen eignet sich das freie, Cloud-basierte Mendely: <http://www.mendeley.com/download-mendeley-desktop/>.

Abbildung 3: Mendeley Referenzmanager



Quelle: <http://dominique-fleury.com/?p=302>

3.5 Texteditor

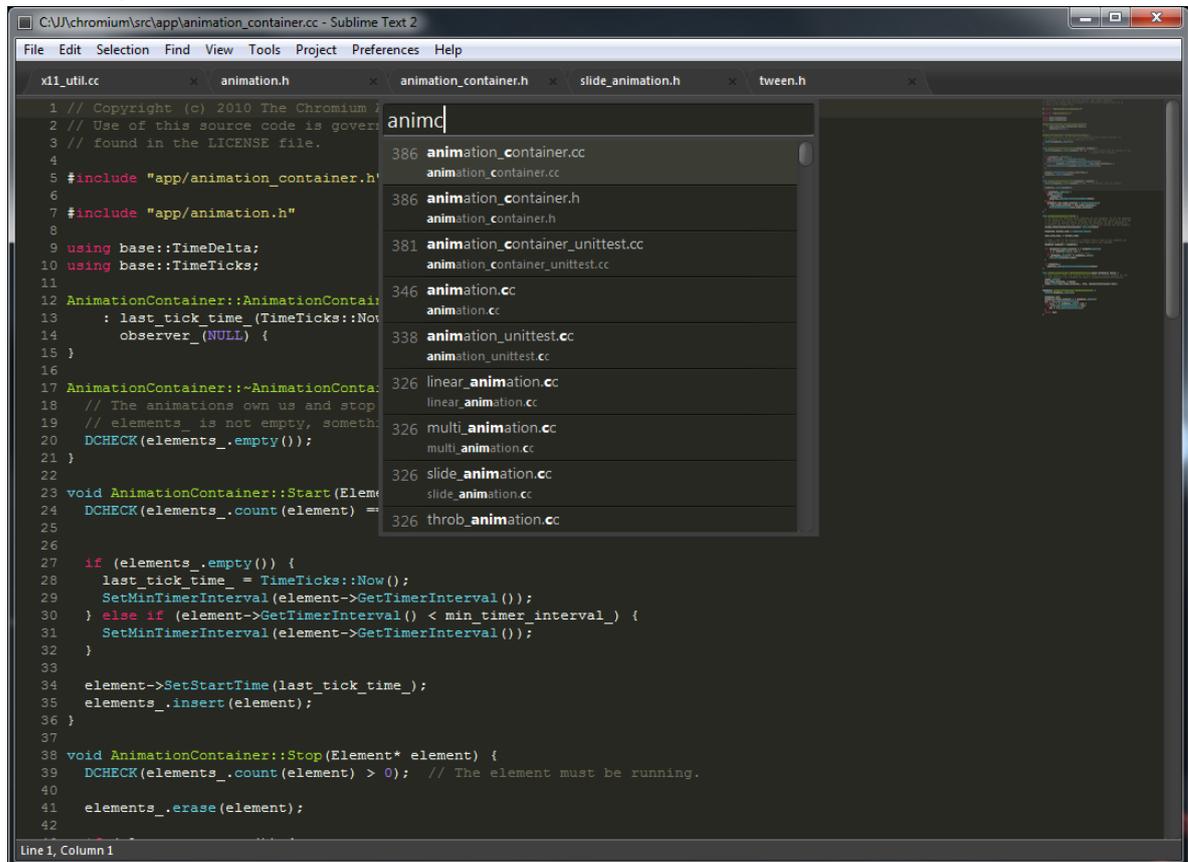
Als Texteditor für \LaTeX wird Sublime Text (<http://www.sublimetext.com>) empfohlen. Zur Arbeit mit Latex ist das Plugin *LaTeXTools* erforderlich (<https://github.com/SublimeText/LaTeXTools>).

3.6 PDF-Erzeugung

Für die Erzeugung des PDF-Dokuments inklusive Referenzen, Quellenverzeichnis und Glossar sind mehrere Programmaufrufe und -durchläufe erforderlich. Der vollständige Aufruf zur PDF-Erzeugung lautet:

```
pdflatex Thesis
biber Thesis
makeindex -s Thesis.ist -t Thesis.alg -o Thesis.acr Thesis.acn
makeglossaries Thesis
```

Abbildung 4: Sublime Texteditor



Quelle: http://www.sublimetext.com/screenshots/alpha_goto_anything2_large.png

pdflatex Thesis

pdflatex Thesis

4 Zusammenfassung

Dieses Dokument ist eine Hilfe, um die Formalien für eine Bachelor-Thesis an der FHDW bei der Verwendung von \LaTeX zu erfüllen und dabei möglichst viele Automatismen von \LaTeX zu nutzen. Eine Absprache mit dem betreuenden Professor ist dennoch ratsam.

Anhang

Anhangsverzeichnis

| | |
|--|----|
| Anhang 1: Gesprächsnotizen | 15 |
| Anhang 1.1: Gespräch mit Werner Müller | 15 |

Anhang 1 Gesprächsnotizen

Anhang 1.1 Gespräch mit Werner Müller

Gespräch mit Werner Müller am 01.01.2013 zum Thema XXX:

- Über das gute Wetter gesprochen
- Die Regenwahrscheinlichkeit liegt immer bei ca. 3%
- Das Unternehmen ist total super
- Hier könnte eine wichtige Gesprächsnotiz stehen

Quellenverzeichnis

Monographien

Füller, Matthias, Willi Nüßer und Thomas Rustemeyer (2012). „Context driven process selection and integration of mobile and pervasive systems“. In: *Pervasive and Mobile Computing* 8.3, S. 467–482.

Meier, Reto (2010). *Professional Android 2 Application Development*. Indianapolis, IN, USA: Wiley, S. 576.

Aufsätze in Sammelbänden und Zeitschriften

Hocking, Christopher G., Steven M. Furnell, Nathan L. Clarke und Paul L. Reynolds (2010). „A Distributed and Cooperative User Authentication Framework“. In: *Proceedings of International Conference on Information Assurance and Security (IAS)*, S. 304–310.

Internetquellen

Data-Science-Team (2020). *What the heck is time-series data (and why do I need a time-series database)?* URL: <https://datascience.eu/wiki/what-the-heck-is-time-series-data-and-why-do-i-need-a-time-series-database/> (besucht am 9. Mai 2022).

Dix, Paul (2021). *Why Time Series Matters for Metrics, Real-Time Analytics and Sensor Data*. URL: <http://get.influxdata.com/rs/972-GDU-533/images/why%20time%20series.pdf> (besucht am 10. Mai 2022).

Fangman, Sam (2019). *The Time Has Come for a New Type of Database*. URL: <https://medium.datadriveninvestor.com/the-time-has-come-for-a-new-type-of-database-47cf8df1667a> (besucht am 10. Mai 2022).

Maslennikov, Denis (2011). *Zeus-in-the-Mobile – Facts and Theories*. URL: http://www.securelist.com/en/analysis/204792194/ZeuS%5C_in%5C_the%5C_Mobile%5C_Facts%5C_and%5C_Theories (besucht am 20. Dez. 2012).

solid-IT-gmbh (2022). *DB-Engines Ranking*. URL: <https://db-engines.com/de/ranking> (besucht am 10. Mai 2022).

Ehrenwörtliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Studienarbeit - Entwurf selbständig angefertigt habe. Es wurden nur die in der Arbeit ausdrücklich benannten Quellen und Hilfsmittel benutzt. Wörtlich oder sinngemäß übernommenes Gedankengut habe ich als solches kenntlich gemacht. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Paderborn, 10. Mai 2022

Henrik Mertens