

Dokumentation der Softwarearchitektur des Projekts Graphvio

Abstract—In diesem technischem Report wird die Softwarearchitektur des Projekts Graphvio vorgestellt. Das Projekt wurde im Rahmen der Vorlesung Semantic Web Technologien (SWT) implementiert. Ziel der Implementierung ist es, Suchanfragen in einer Graphdatenbank auszuwerten und die Ergebnisse über eine Webapplikation darzustellen.

Index Terms—Semantic Web, Syntactic Web, Graph Database, Web UI

I. INTRODUCTION

Bei der Applikation Graphvio werden vorhandene Datensätze von Streaminganbietern wie Amazon Prime, Disney Plus, Netflix und Hulu in eine lokale Graphdatenbank abgelegt. Auf den Graph können dann per Rest-Schnittstelle Suchanfragen angewendet werden, die Ergebnisse im JSON-Format liefern. Die Interaktion des Anwenders mit der Applikation erfolgt dabei über das Webframework React.

II. ARCHITEKTUR ALLGEMEIN

Als Architektur wird ein Ansatz verfolgt, der sich stark an Microservices orientiert. Dadurch wird eine abgeschlossene Logik von Front- und Backendkomponenten erreicht, die jeweils von einem Subteam entwickelt wird. Diese gekapselten Einheiten können jeweils als isolierte Applikation der Containervisualisierung Docker betrieben werden. Durch eine zentrale Konfigurationsdatei des Containers können diese einfach zwischen den Teammitgliedern ausgetauscht und ausgeführt werden. Damit wird eine hohe Portabilität gewährleistet. Die Kommunikation der Teilsysteme wird durch eine sprachunabhängige REST-Schnittstelle angeboten. Dadurch wird es möglich einzelne Komponenten des Gesamtsystems auszutauschen, ohne größere Veränderungen an den anderen Microservices vornehmen zu müssen. Dazu werden modulare Reactkomponenten erzeugt, die über Texteingaben und weitere Suchmaskenelemente Parameter der REST-Anfrage setzen. Die zurückgelieferte Antwort der Anfrage im JSON-Format wird wiederum über modulare Reactkomponenten auf der Webapplikation dargestellt.

III. FILMVERWALTUNG

Mithilfe des Web-Frameworks Express ist es möglich, Endpunkte für eine leistungsfähige API zu erstellen. Dabei werden Client-Anfragen an passende Endpunkte des Node-Servers übergeben. Nach Bearbeitung dieser erfolgt eine Antwort mit den Daten im JSON-Format. JSON ist ein programmiersprachen unabhängiges Format, welches leicht lesbar, wenig Code erfordert und zusätzlich schnell

verarbeitet werden kann. Zusätzlich wird dabei wieder eine Austauschbarkeit einzelner Komponenten gefördert, da keine proprietären Datentypen verwendet werden müssen.

Die Funktionalität der Filmverwaltung umfasst folgende API-Zugriffe:

Filmsuche (/db/search-movies)

- Beschreibung: Bei der Filmsuche werden in der Datenbank passende Filme gesucht.
- Parameter:
 - Teiltitel
 - Provider
- Rückgabe: Liste mit Filmen und deren Metadaten

Personensuche (/db/search-persons)

- Beschreibung: Bei der Personensuche werden in der DBpedia Daten zu Filmmitwirkenden eines Filmes gesucht.
- Parameter:
 - Filmtitel
 - Erscheinungsjahr
- Rückgabe: Liste der Filmmitwirkenden und dazugehörigen Metadaten

Gemeinsamkeitensuche (/db/compare-movies)

- Beschreibung: Bei der Gemeinsamkeitensuche werden gemeinsame Merkmale einzelner Filme gesucht.
- Parameter: Liste mit Filmtiteln
- Rückgabe: Liste der Gemeinsamkeiten der Filme

Suche nach ähnlichen Filmen (/db/search-similar-movies)

- Beschreibung: Bei der Suche nach ähnlichen Filmen werden Filme gesucht, die Gemeinsamkeiten mit den angegebenen Filmen aufweisen.
- Parameter:
 - Liste mit Filmtiteln
 - Streamingdienstanbieter der zu vergleichenden Filme
- Rückgabe: Liste mit Filmen

Suche nach IMDB Daten eines Filmes (/imdb/search-imdbdata)

- Beschreibung: Bei der Suche nach einem Film werden die Bewertungen und ein hinterlegtes Bild des Filmes in der IMDB gesucht.
- Parameter: Titel des Filmes
- Rückgabe:
 - Liste mit Bewertungen des Filmes

- URL eines in der IMDB hinterlegten Bildes

Die Verwaltung der Daten erfolgt mittels einer Virtuoso-Datenbank. Diese enthält Filmdaten der Streamingdienstanbieter Netflix, Amazon-Prime, Disney+ und Hulu, welche aus Kaggle bezogen wurden. Die Datensätze enthalten Filmtitel der jeweiligen Provider, sowie dessen Metainformationen wie Filmmitwirkende, Genre und Erscheinungsjahr. Mithilfe einer in Virtuoso angelegten View ist es möglich mit SparQL-Statements auf die importierten Daten zu zugreifen.

Das Modul "Fetch SPARQL Endpoint" ermöglicht es innerhalb der Filmverwaltung eine Verbindung zwischen dem Node-Server und der Virtuoso-Datenbank sowie der DBpedia aufzubauen. Anschließend können SPARQL-Statements abgesetzt und die Ergebnisse asynchron Verarbeitet werden.

Die IMDB ist eine Datenbank zu Filmen, Fernsehserien, Videoproduktionen und Computerspielen sowie über Personen, die daran mitgewirkt haben. Sie bietet eine API zum Zugriff auf diese Daten. Dafür wird ein API-Schlüssel benötigt, mit diesem sind täglich 100 Anfragen kostenfrei verfügbar. Nach überschreiten dieser Begrenzung sind keine weiteren Anfragen an diesem Tag möglich, die Anfragenbegrenzung kann jedoch kostenpflichtig erhöht werden.

Um die vorgesehene Funktionalität gewährleisten zu können, wurden mithilfe von Jest, einem JavaScript Testframework, Unit- und Integrationstests entworfen. Dadurch kann die interne Integrität, sowie die Anbindung an externe Systeme, sichergestellt werden. Die Tests wurden dabei mit dem Arrange-Act-Assert-Pattern entworfen. Dieses soll eine übersichtliche Struktur und Einheitlichkeit garantieren.

IV. FRONTEND

A. Funktionalitäten Frontend

Das Frontend wird über eine Single-Page Applikation (SPA) mit dem Webframework React realisiert. In der Kopfzeile der Basisdatei App.js wird ein Dropdown angeboten, über das einzelne Suchmöglichkeiten zu den Graphdaten angewählt werden können. Es stehen folgende Optionen zur Auswahl

- **Movie Search:** Angabe eines Textinhalts, der als Bestandteil in allen Filmtiteln gesucht wird
- **Movie Compare:** Angabe von mehreren Filmtiteln über eine Mehrfachauswahl, die nach übereinstimmenden Metadaten durchsucht werden
- **Movie Recommend:** Angabe von mehreren Filmtiteln über eine Mehrfachauswahl, die ähnliche Filme zu der Auswahl auflistet

Zusätzlich kann bei der Anzeige von Filmen auf die Filmbezeichnung geklickt werden, um weitere Informationen zu den Mitwirkenden zu erhalten.

B. Aufbau Frontend

Die einzelnen Seiteninhalte (Views) werden im Body der Seite gerendert und sind als modulare Elemente im Ordner View abgelegt. Die Ein- und Ausgabeelemente der einzelnen Views sind wiederum aus modularen Reactklassen aufgebaut. Diese Klassen sind im Ordner Base abgelegt. Für die Umsetzung wurden sowohl reactspezifische als auch allgemeine HTML-Elemente verwendet. Die formatspezifischen Eigenschaften dieser Elemente wurden in eigenen CSS-Dateien festgelegt. Für die REST-Abfragen werden eigene Klassenfunktionen in den Views angelegt. Die Aufrufe erfolgen dabei über das Package Axios. Das Package beinhaltet einen HTTP-Client, der asynchrone Aufrufe unterstützt und so das Rendering der Seite nicht negativ beeinflusst.

V. TESTING

A. Testen des Backend

B. Testen des Frontend

Das Frontend wird durch Unittests auf Fehler untersucht. Dazu wird zu jeder Komponente im Base-Ordner eine Testdatei mit Funktionalitäten aus der React Testing Library geschrieben. Die Tests werden über das Kommando "npm test" ausgeführt und erzeugen Klassenobjekte aus den Basiskomponenten mit spezifischen Parametern. Anschließend wird das aktuelle Abbild (Snapshot) mit dem Abbild des vorherigen Test verglichen.

VI. EASE OF USE

A. Maintaining the Integrity of the Specifications

The IEEEtran class file is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

VII. PREPARE YOUR PAPER BEFORE STYLING

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections ??-?? below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not number text heads— \LaTeX will do that for you.

A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, ac, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

B. Units

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: “Wb/m²” or “webers per square meter”, not “webers/m²”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.
- Use a zero before decimal points: “0.25”, not “.25”. Use “cm³”, not “cc”).

C. Equations

Number equations consecutively. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

$$a + b = \gamma \quad (1)$$

Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(??)”, not “Eq. (??)” or “equation (??)”, except at the beginning of a sentence: “Equation (??) is . . .”

D. \LaTeX -Specific Advice

Please use “soft” (e.g., `\eqref{Eq}`) cross references instead of “hard” references (e.g., (1)). That will make it possible to combine sections, add equations, or change the order of figures or citations without having to go through the file line by line.

Please don’t use the `{eqnarray}` equation environment. Use `{align}` or `{IEEEeqnarray}` instead. The `{eqnarray}` environment leaves unsightly spaces around relation symbols.

Please note that the `{subequations}` environment in \LaTeX will increment the main equation counter even when there are no equation numbers displayed. If you forget that, you might write an article in which the equation numbers skip from (17) to (20), causing the copy editors to wonder if you’ve discovered a new method of counting.

\BibTeX does not work by magic. It doesn’t get the bibliographic data from thin air but from .bib files. If you use \BibTeX to produce a bibliography you must send the .bib files.

\LaTeX can’t read your mind. If you assign the same label to a subsection and a table, you might find that Table I has been cross referenced as Table IV-B3.

\LaTeX does not have precognitive abilities. If you put a `\label` command before the command that updates the counter it’s supposed to be using, the label will pick up the last counter to be cross referenced instead. In particular, a `\label` command should not go before the caption of a figure or a table.

Do not use `\nonumber` inside the `{array}` environment. It will not stop equation numbers inside `{array}` (there won’t be any anyway) and it might stop a wanted equation number in the surrounding equation.

E. Some Common Mistakes

- The word “data” is plural, not singular.
- The subscript for the permeability of vacuum μ_0 , and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
- In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an “inset”, not an “insert”. The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
- Do not use the word “essentially” to mean “approximately” or “effectively”.
- In your paper title, if the words “that uses” can accurately replace the word “using”, capitalize the “u”; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones “affect” and “effect”, “complement” and “compliment”, “discreet” and “discrete”, “principal” and “principle”.
- Do not confuse “imply” and “infer”.
- The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the “et” in the Latin abbreviation “et al.”.
- The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

An excellent style manual for science writers is [?].

F. Authors and Affiliations

The class file is designed for, but not limited to, six authors. A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

G. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced.

H. Figures and Tables

a) *Positioning Figures and Tables:* Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. ??”, even at the beginning of a sentence.

TABLE I
TABLE TYPE STYLES

Table Head	Table Column Head		
	<i>Table column subhead</i>	<i>Subhead</i>	<i>Subhead</i>
copy	More table copy ^a		

^aSample of a Table footnote.

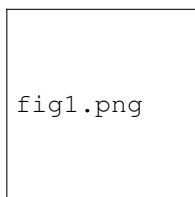


Fig. 1. Example of a figure caption.

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

ACKNOWLEDGMENT

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

Please number citations consecutively within brackets [?]. The sentence punctuation follows the bracket [?]. Refer simply to the reference number, as in [?]¹—do not use “Ref. [?]” or “reference [?]” except at the beginning of a sentence: “Reference [?] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [?]. Papers that have been accepted for publication should be cited as “in press” [?]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [?].

REFERENCES

- [1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955.
- [2] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
- [3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
- [4] K. Elissa, “Title of paper if known,” unpublished.
- [5] R. Nicole, “Title of paper with only first word capitalized,” *J. Name Stand. Abbrev.*, in press.
- [6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” *IEEE Transl. J. Magn. Japan*, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetism Japan, p. 301, 1982].
- [7] M. Young, *The Technical Writer’s Handbook*. Mill Valley, CA: University Science, 1989.

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove the template text from your paper may result in your paper not being published.