Fernuniversität in Hagen

Fakultät für Mathematik und Informatik

Lehrgebiet Multimedia und Internetanwendungen

Exposé

Eine Autorenumgebung zur Unterstützung der Transformation und Repräsentation von emerging Argument Entity (eAE) Mindmaps in Aussagengraphen und darauf aufbauenden Argumentationsbäumen

eingereicht von: Alexander Flammer

Matrikelnummer: [redacted]

Studiengang: Praktische Informatik

betreut durch: Prof. Dr.-Ing. Matthias Hemmje

Alexander Duttenhöfer

Tauberbischofsheim, der 5. Januar 2022

Contents

List of Abbreviations

AIF	Argument Interchange Format
API	Application Programming Interface
DFG	Deutsche Forschungsgemeinschaft
eA	emerging Argument
eAE	emerging Argument Entity
eNE	emerging Named Entity
eNER	emerging Named Entity Recognition
eNER-eAE-IRS	emerging Named Entity Recognition and emerging Argument Entity Information Retrieval System
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
IETF	Internet Engineering Task Force
IR	Information Retrieval
IS	Information System
KM-EP	Knowledge Management Ecosystem Portal
MeSH	Medical Subject Headings
NER	Named Entity Recognition
OWL	Web Ontology Language
PoC	Proof of Concept
RecomRatio	Rationalizing Recommendations
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
RFC	Request for Comments
SPARQL	SPARQL Protocol And RDF Query Language

SPO	Subject-Predicate-Object
URI	Uniform Resource Identifier
URL	Uniform Resource Locator

List of Listings

List of Figures

Chapter 1

Introduction

This thesis addresses the challenge of defining a machine-readable data format for arguments in the medical domain. A solution is discussed in which a data format is specified and validated using a Proof of Concept (PoC) argument editor implementation. First of all the used concepts are introduced in the following sections. They explain the advantages of machine-readable medical facts, already existing possible solutions, which research methodology is used to derive and validate a data format as well as the structure of the thesis.

The World Wide Web was invented by Tim Berners-Lee in 1989 and soon became the standard for information exchange over the internet [1]. It enabled the usage of Uniform Resource Identifier (URI) to cross-reference documents and therefore making them accessible via the Hypertext Transfer Protocol (HTTP). As early as 1998 Berners-Lee recognized that the focus on human-readable documents restricted the web. To achieve interaction between humans and computers semantic information about the documents needs to be understandable by the machines.

In 2001 Berners-Lee specified his vision of a Semantic Web [3]. In such a Semantic Web the documents served by the web are enriched with semantic information. The semantic information is machine-readable and can therefore be used to build new services by computer systems. Nowadays Semantic Web technologies are used throughout the Web, the most well knowns being DBpedia [11], the Google Knowledge Graph [19] and the Open Graph Protocol [21] by Facebook. The Semantic Web technology stack enables the formalization and expression of information in a machine-readable way, therefore its standards and formats are relevant for this thesis.

1.1 Motivation

As early as 1971 Jablonski [9] used the term information explosion to describe the steadily increasing number of scientific publications in medicine. In doing so, he pointed out that the ever-increasing amounts of information are only insufficiently searchable and that, among other things, indexing could improve information retrieval.

The research project Rationalizing Recommendations (RecomRatio) of the Deutsche Forschungsgemeinschaft (DFG) studies the support of medical professionals in decision-making processes by the automated extraction of arguments from scientific literature. In the context of medical Information Retrieval (IR) systems, the aim is to support expert users in decision-making by providing them with arguments for and against a particular therapy or diagnosis, among other things. [4]

In the context of Christian Nawroth's dissertation, "Supporting Information Retrieval of Emerging Knowledge and Argumentation" [14], an emerging Named Entity Recognition and emerging Argument Entity Information Retrieval System (eNER-eAE-IRS) was developed that specifically addresses the problem of constantly emerging publications. This system identifies emerging Named Entity (eNE) in medical literature as well as the associated emerging Argument Entity (eAE). This is of particular importance, since medical catalogs such as Medical Subject Headings (MeSH) must be constantly updated due to constantly new scientific publications. The eNER-eAE-IRS does not rely on updating catalogs such as MeSH, it recognizes emerging technical terms as well as arguments before they are included in specialized catalogs. Therefore, it allows its users to access a pool of data that is as up-to-date as possible, thus addressing the problem of information explosion.

The eNER-eAE-IRS is a standalone system connected to the Knowledge Management Ecosystem Portal (KM-EP) by an HTTP based Application Programming Interface (API). The KM-EP system was created by Vu et al. [23] and allows searching and managing medical literature through a web interface. The output of eAE in KM-EP is implemented using Hypertext Markup Language (HTML) in tabular text form.

By using the eNER-eAE-IRS in cooperation with the KM-EP, even the latest eAE can be found by medical professionals. However, there is still the problem that the relevant information needs to be extracted manually by a specialist user. Neither can data be linked and further processed nor automatically evaluated or filtered.

1.2 Problem Statement

Arguments are a central point of human interaction, they influence opinion making and enable us to express a point of view. To be able to exchange them between humans without loss of information a dedicated data format needs to be used. Furthermore, in the context of IR Systems arguments have to be machine-readable. [13]

As described in the previous section, the eNER-eAE-IRS in combination with the KM-EP supports the IR process of specialist users by extracting eNE and their eAE. But the arguments are not machine-readable as they are presented as plain strings on a web page by the KM-EP. This is equivalent to the document-based Web as explained in the first section of this thesis. The eNER-eAE-IRS is able to produce mindmaps for a given search query including the found eNE and the corresponding eAE. Such mindmaps support the understanding of interdependencies of the represented data but cannot be processed by machines.

Therefore, the graphical mindmap representation needs to be transformed into a graph of statements, represented in a machine-readable document format. Once such a statement graph exists, further applications are possible.

For medical specialist users it is important to make well-informed decisions, hence they'll need to extract the relevant statements from a graph and form an argument tree. Such an argument tree could represent disadvantages and advantages of a specific therapy or other important information for decision-making.

Summarized, the following problems exist:

- P_1 eAE mindmaps cannot be transformed into a graph structure, therefore preventing the exchange of arguments in a standardized format.
- P_2 Statements cannot be expressed in argument trees to support decision-making.
- P_3 Statements cannot be annotated with metadata, furthermore statements about statements cannot be expressed (Reification).
- P_4 Statement graphs and argument trees need to be visualized for easier understanding of the interdependencies of the statements.
- P_5 eAE mindmaps cannot be used for logical queries, whereas a graph structure could be used to infer new statements from an existing statement graph or extract filtered statements.

This leads to my two hypothesis:

$H_\mathbf{1}$ *Statements, derived from mindmaps of the eNER-eAE-IRS, can be represented in a machine-readable graph structure and such a graph can be used to visualize, annotate, reificate and query a set of arguments*

$H_\mathbf{2}$ *A subset of statements of a statement graph can be used to create an argument tree*

The first problem requires the specification of a data format suitable for eAE and the definition of an editing process for arguments, which enables users to transform existing mindmaps into statement graphs both manually and automated. If such a graph representation can be achieved and H_1 is true, then the second hypothesis H_2 can build on statement graphs and create argument trees which can support the decision-making process.

1.3 Research Questions

This section outlines the research questions and related challenges. They are derived from the problem statement (especially the mentioned problems P_1 to P_5) and the hypothesis H_1 and H_2 in the previous section 1.2. The hypothesis H_1 mentions the representation of statements in a graph structure. This problem is addressed in the following research question.

Q_1 How can statements, from mindmaps of the eNER-eAE-IRS, be transformed into a graph data format? (addresses P_1 and P_3)

The challenges to solve Q_1 include identifying similarities and differences of arguments based upon mindmaps of the eNER-eAE-IRS. Based upon the findings a vocabulary can be derived which can be used to formulate the analyzed statements. Furthermore, it deals with the challenge of designing a data format which addresses the problems introduced in the previous section 1.2. Existing data formats need to be evaluated and a specification of the data format created. Afterwards a PoC implementation can be used to validate the data format. To create and update arguments in the new data format a dedicated argument editor needs to be created which provides a manual as well as automatic transformation process.

The hypothesis H_2 discusses the possibility to create argument trees from statement graphs. It results in the following research question.

Q_2 How can a subset of statements of a statement graph be transformed into an argument tree representation? (addresses P_2)

The challenges derived from Q_2 include the selection and linking of statements from a graph to create an argument tree.

The following research questions address the remaining issues described in H_1.

Q_3 How can argument graphs and trees be visualized? (addresses P_4)

The challenges for Q_3 include research about available visualization frameworks and the implementation of a PoC visualization. Graphical editing should be evaluated, which allows to not only be able to create visualizations, but adapt the visualization without knowledge of the underlying data format.

Q_4 Is it possible to query a statement graph to filter the set of statements or infer new arguments? (addresses P_5)

Q_4 deals with the expected benefits of graph structures, which represent a well studied subject. Once a graph data structure has been implemented, can existing frameworks or algorithms be applied to query or filter a graph or even infer new statements?

1.4 Research Methodology

The research methodology according to Nunamaker, Chen, and Purdin [15] is intended for the development of information systems and is applied in this work. The methodology consists of four phases (see Figure 1.1) each of which represents a research strategy. In the following paragraphs, these are presented according to Nunamaker, Chen, and Purdin [15, pp. 94 sq.].

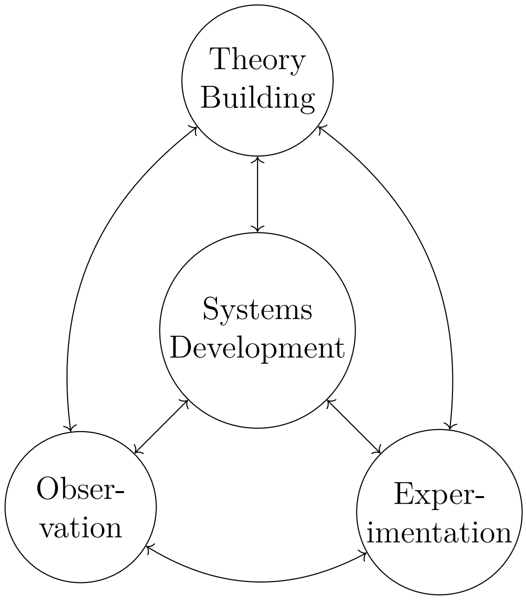

Figure 1.1: IS research process [15, p. 94]

The methodology does not enforce an order of execution for the phases, instead the insights of each phase may influence all other phases. In this thesis, the phases are ordered to structure the thesis. The phases are presented in the following paragraphs in the order of their execution during this thesis. Furthermore, the research objectives can be assigned to the individual phases.

Observation Surveys as well as case and field studies are research methods categorized in the Observation phase. These methods are especially useful if little is known about the research domain. Furthermore, the literature research belongs to this phase. Therefore, this phase is the first phase in this thesis as evaluating the State of the Art belongs to it.

Theory Building This research strategy is essentially concerned with the creation of new ideas, concepts, frameworks, and methods, models, or theories. Therefore, this phase contributes significantly to the creation of new knowledge and its theories can influence the Observation and Experimentation phases in particular. It is the second phase in this thesis. Based upon the results of the *Observation* phase models and concepts are created to answer the research questions.

Systems Development Systems Development deals with the implementation of an information system and is divided into the following phases:

1. concept design: Design of the software, especially the to-be-implemented use cases
2. constructing the architecture: Specification of the software architecture
3. prototyping: Creation of a prototype as PoC
4. product development: Extension of the prototype to a complete solution
5. technology transfer: Deployment of the software (esp. in companies and organizations)

In this thesis the *Systems Development* phase will be used to implement the concepts from the previous *Theory Building* phase. A PoC argument editor is implemented. The *Systems Development* phases *Product Development* and *Technology Transfer* are not conducted in this thesis.

Experimentation The Experimentation phase includes laboratory as well as field experiments and computer simulations. In this phase, the theories gained from the *Theory Building* phase are validated with the implemented PoC from the *Systems Development* phase.

1.5 Research Objectives

Based on the research questions from section 1.3 and the research methodology presented in section 1.4, the research objectives to be achieved are defined below and assigned to the corresponding phases according to Nunamaker, Chen, and Purdin [15].

As all research questions are directly or indirectly influenced by the eNER-eAE-IRS and its mindmaps, these need to be studied in an overarching research goal:

$O_{\mathbf{0.1}}$ Identify the structure and capabilities of mindmaps produced by the eNER-eAE-IRS *Observation*

Research Question Q_1 is about the transformation of arguments from mindmaps into a machine-readable data format. Therefore, the structure, commonalities and characteristics of medical arguments and their representation in mindmaps produced by the eNER-eAE-IRS need to be analyzed. Based on the methodology presented in [15], the following research objectives are derived for Q_1:

$O_{\mathbf{1.1}}$ Identify common aspects and characteristics of eAE based on output of the eNER-eAE-IRS for the Vitamin-D-Metabolism. *Observation*

$O_{\mathbf{1.2}}$ Define an abstract data model which is able to represent the Vitamin-D-Metabolism arguments. *Theory Building*

Before the concrete data format can be implemented, additionally to the observations gained by achieving $O_{1.1}$ and $O_{1.2}$, it is necessary to understand which data formats and representations are possible and already existing. Therefore, for research question Q_1 the following objectives are additionally identified:

$O_{\mathbf{1.3}}$ Discuss existing data formats which are capable of representing logical statements. *Observation*

$O_{\mathbf{1.4}}$ Find suitable frameworks for the creation of an argument editor. *Observation*

$O_{\mathbf{1.5}}$ Specify a data format for medical statements. *Theory Building*

$O_{\mathbf{1.6}}$ Define a process to transform mindmaps into the data format. *Theory Building*

$O_{\mathbf{1.7}}$ Analyze the use cases for the argument editor. *Theory Building*

$O_{\mathbf{1.8}}$ Implement a mindmap to data format transformation in an argument editor. *Systems Development*

$O_{\mathbf{1.9}}$ Evaluate the data format using the argument editor. *Experimentation*

After having defined and implemented a data format based on graphs for medical arguments, the methodology leads to Q_2 which results in the following objectives:

$O_{\mathbf{2.1}}$ How can arguments be arranged in a tree structure. *Observation*

$O_{\mathbf{2.2}}$ Define a process to create an argument tree from a statement graph. *Theory Building*

$O_{\mathbf{2.3}}$ Enhance the argument editor mentioned in $O_{1.7}$ with the functionality to create argument trees. *Systems Development*

$O_{\mathbf{2.4}}$ Evaluate the argument tree creation using the argument editor. *Experimentation*

Once Q_1 and Q_2 have been processed, an argument editor should exist which enables us to create and modify statement graphs and argument trees. Next, Q_3 deals with the visualization possibilities of the graph and tree structures. The following objectives are derived:

$O_{\mathbf{3.1}}$ Identify common visualization frameworks. *Observation*

$O_{\mathbf{3.2}}$ Define a visualization approach used by the argument editor. *Theory Building*

$O_{\mathbf{3.3}}$ Implement a visualization of statement graphs and trees for the argument editor. *Systems Development*

$O_{\mathbf{3.4}}$ Evaluate the implemented visualization. *Experimentation*

The last research question Q_4 deals with the possible applications of existing algorithms and frameworks to query graphs. It leads to the objectives:

$O_{\mathbf{4.1}}$ Identify querying possibilities and frameworks. *Observation*

$O_{\mathbf{4.2}}$ Define a query use case for the argument editor. *Theory Building*

$O_{\mathbf{4.3}}$ Implement the query capability for the argument editor. *Systems Development*

$O_{\mathbf{4.4}}$ Evaluate the query capability. *Experimentation*

The research objectives are derived from the research questions and their challenges described in section 1.3. As described in objectives $O_{1.6}$, $O_{2.3}$, $O_{3.3}$ and $O_{4.3}$, an argument editor is at the core of the thesis. It can be used to finally evaluate the designs made in the *Theory Building* phase.

1.6 Research Approach

The research approach of the thesis follows the methodology according to Nunamaker, Chen, and Purdin [15] by assigning the research objectives to the phases *Observation*, *Theory Building*, *Systems Development* and *Experimentation*. The following section presents the phases with their corresponding research objectives.

The phase *Observation* deals with research regarding the state of the art in science and technology. Furthermore, the fundamental principles used throughout this thesis are introduced, e.g. statements and graph and tree data structures. The research objectives processed in this phase are $O_{0.1}$, $O_{1.1}$, $O_{1.3}$, $O_{1.4}$, $O_{2.1}$, $O_{3.1}$ and $O_{4.1}$. They are dealt with in chapter 2 in detail.

In the following *Theory Building* phase new concepts and models are created to fulfill the research goals. The objectives $O_{1.2}$, $O_{1.5}$, $O_{1.6}$, $O_{1.7}$, $O_{2.2}$, $O_{3.2}$ and $O_{4.2}$ belong to this phase. This work is about the representation of statements in a machine-readable data format hence processes and uses cases need to be designed for the transformation process as well as a model and specification of the data format. These issues are discussed in chapter 3. Additionally, an evaluation concept of the to be implemented data format and argument editor needs to be established.

A PoC implementation is attempted during the *Systems Development* phase. The PoC should be used to evaluate the concepts and models created in the previous phase afterwards. Objectives $O_{1.8}$, $O_{2.3}$, $O_{3.3}$ and $O_{4.3}$ belong to this phase. They deal with the implementation of specific functionality of the PoC argument editor which is described in chapter 4. They discuss the implementation of the argument editor and how it can be used to transform mindmaps into statement graphs and these into argument trees. Furthermore, visualization and querying of these data structures are discussed.

Finally, the *Experimentation* phase evaluates the developed models, concepts and the PoC argument editor. The research objectives $O_{1.9}$, $O_{2.4}$, $O_{3.4}$ and $O_{4.4}$ are handled in this phase and described in chapter 5.

1.7 Outline of the Thesis

The structure of the thesis is based on the assignment of the research objectives to the phases according to Nunamaker, Chen, and Purdin [15] presented in the previous section.

Chapter 2 is associated to the *Observation* phase and describes the state of the art in science and technology regarding knowledge representation including additional relevant principles like graph and tree data structures.

The following *Theory Building* phase discusses the challenges regarding the creation of new models and concepts in chapter 3. A model and data format for the representation of medical statements is proposed and also the argument editor designed. Especially the use cases of the argument editor need to be established before the implementation can begin. In addition, an evaluation concept for the created models and concepts is presented.

In the following chapter 4, the *Systems Development* phase is represented. The implementation of the PoC argument editor is discussed. Additionally, used third-party software components and used techniques, frameworks and tools are presented.

The last phase *Experimentation* maps to chapter 5 and deals with the execution of the designed evaluation concept for the PoC argument editor. Therefore, the results of the evaluation are presented.

The thesis concludes with chapter 6, which provides a summary of the gained results and how they contribute to the research objectives and research goals. Afterwards, a conclusion is made based on the evaluation results of the previous chapter and the scientific contributions achieved in this thesis.

The presented outline of the thesis covers all chapters of this work and describes them using a short high-level summary. After this summary, the next chapter discusses the relevant state of the art in science and technology. The chapter begins with general principles necessary to be understood for this thesis and then processes the research objectives mapped to the *Observation* phase described in chapter 1.6. The state of the art acts as the foundation to provide models and concepts to answer the research questions described in chapter 1.3.

Chapter 2

State of the Art in Science and Technology

This chapter discusses the necessary observations for the research questions presented in section 1.3. First, the section Principles introduces and explains required principles used throughout this thesis. These principles are needed to understand the following observations. Afterwards, the research questions are handled sequentially in sections 2.2 until 2.6.

After all observation-related research objectives are answered, a summary discusses the solved and remaining challenges.

2.1 Principles

This section introduces basic concepts which are required to understand this work. General terms like mindmap, graph and tree are defined.

2.1.1 Mindmap

The term mindmap got coined mainly by Tony and Barry Buzan [6]. According to them, a mindmap represents a methodology to structure information related to a single topic. This is done by placing the main topic in the center and then use branches to mark related topics or information in separate nodes. Every node represents a more specific term or concept from its ancestor, hence the level of detail increases with the level of the branches. The assumption is that by structuring information from the abstract to the detailed level, the human brain is better able to memorize the represented information. Furthermore, it is a visual approach and therefore colors and pictures can be used to the liking of the user to achieve a better memory recall.

Mindmaps are widely used in the learning sector because many studies at least show a subjective advantage for learners to memorize new information or even that mindmaps can have a positive effect on memory recall.

2.1.2 Graph

According to [7] a graph is a data structure representing a set of objects with their relationships. It is called graph because it is graphically represented by nodes representing the objects and edges connecting the nodes representing the relationships. Relationships may either be symmetric or directed, hence there are two types of graphs: directed graphs and undirected graphs.

A directed graph is a pair

$$G = (V, E)$$

for which applies

$$V$$

is a finite set of elements called nodes

$$E \subseteq VxV$$

an edge is a pair of nodes, the edges are a subset of the cartesian product of the nodes

Furthermore, an edge from node v to w can be represented as (v,w).

A undirected graph is a directed graph (V, E) in which the relation E is symmetric:

$$(v, w) \in E \Rightarrow (w, v) \in E$$

2.1.3 Tree

A tree is a data structure used to represent hierarchical relationships. A tree starts with a root node which is linked to child nodes, whereas every node can have child nodes. A node may only have a single parent node. Other than that, there exist multiple special tree data structures optimized for various tasks, e.g. efficient search or modifications.

2.2 Argumentation Mining

This section is related to research question Q_1 as it introduces the required knowledge about argument extraction and representation. These are required to answer Q_1 as it deals with the transformation of such arguments into a data format.

2.2.1 Argumentation

2.2.2 eNER-eAE-IRS

2.2.3 emerging Argument

2.2.4 Text Annotation

2.3 Data representation and standards

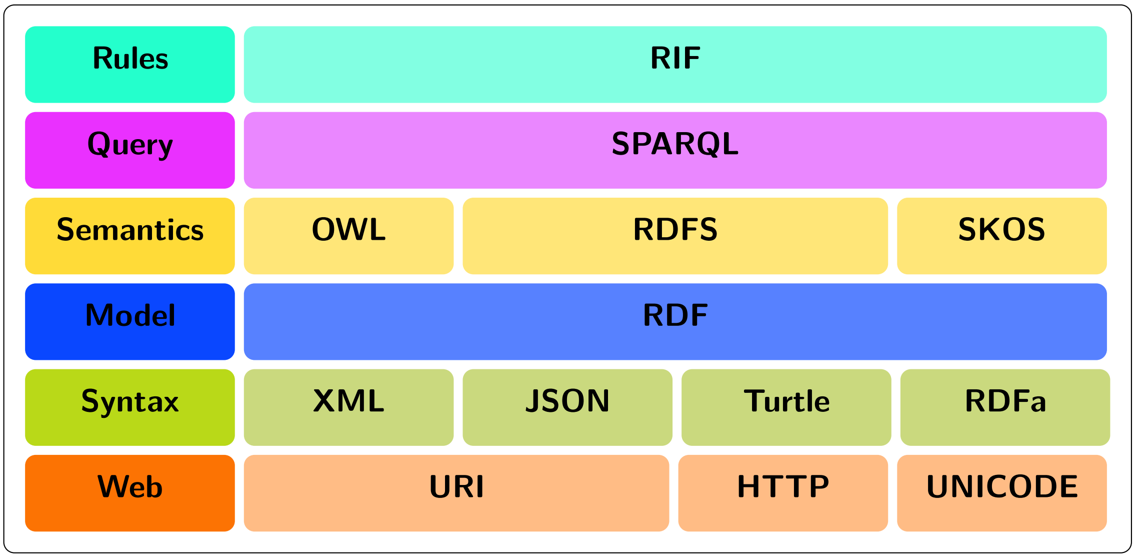

Figure 2.1: Semantic web technology stack

2.3.1 Uniform Resource Identifier

Uniform Resource Identifiers are specified by the Internet Engineering Task Force (IETF) in Request for Comments (RFC) 2396 [2].

2.3.2 Resource Description Framework

2.3.3 Resource Description Framework Schema

2.3.4 Web Ontology Language

2.3.5 Argument Interchange Format

2.4 Data Visualization

2.4.1 Graphviz

2.4.2 Sprotty

2.5 Inference and query languages

2.5.1 Logical inference

2.5.2 SPARQL

2.6 Editor Frameworks and Protocols

2.6.1 Language Server Protocol

2.6.2 Graphical Language Server Protocol

2.6.3 Eclipse RCP

2.6.4 Eclipse Theia

Chapter 3

Planning

3.1 Outline

The following outline represents a possible structure of the thesis and will be adapted in the course of working on the thesis.

1. Introduction
   1.1. Motivation
   1.2. Problem Statement
   1.3. Research Goal
   1.4. Research Methodology
   1.5. Research Objectives
   1.6. Research Approach
   1.7. Outline of the Thesis
2. State of the Art in Science and Technology
   2.1. Principles
      2.1.1 Mindmap
      2.1.2 Graph
      2.1.3 Tree
   2.2. Argumentation Mining
      2.2.1 Argumentation
      2.2.2 eNER-eAE-IRS
      2.2.3 emerging Argument
      2.2.4 Text Annotation

- 2.3. Data representation and standards
  - 2.3.1 Uniform Resource Identifier
  - 2.3.2 Resource Description Framework
  - 2.3.3 Resource Description Framework Schema
  - 2.3.4 Web Ontology Language
  - 2.3.5 Argument Interchange Format
- 2.4. Data Visualization
  - 2.4.1 GraphViz
  - 2.4.2 Sprotty
- 2.5. Inference and query languages
  - 2.5.1 Logical inference
  - 2.5.2 SPARQL
- 2.6. Editor Frameworks
  - 2.6.1 Language Server Protocol
  - 2.6.2 Graphical Language Server Protocol
  - 2.6.3 Eclipse RCP
  - 2.6.4 Eclipse Theia

3. Design
   - 3.1. Data model specification
     - 3.1.1 Statement graphs
     - 3.1.2 Mindmap to graph transformation
     - 3.1.3 Argument trees
     - 3.1.4 Graph to argument tree transformation
   - 3.2. Argument editor
     - 3.2.1 Requirements
     - 3.2.2 Use Cases
     - 3.2.3 System Architecture
     - 3.2.4 Evaluation concept
4. Implementation

4.1. Preparation
 4.1.1 Tools and Frameworks
 4.1.2 Test data
4.2. Argument editor implementation
 4.2.1 Data format support
 4.2.2 Subject-Predicate-Object (SPO) text annotation
 4.2.3 Statement graph creation
 4.2.4 Argument tree creation
 4.2.5 Graph visualization
 4.2.6 Tree visualization
 4.2.7 Graph queries

5. Evaluation
 5.1. Data format evaluation
 5.2. Editor evaluation
 5.3. Summary
6. Conclusion
 6.1. Results discussion
 6.2. Remaining challenges
 6.3. Future works

3.2 Schedule

The duration of the master thesis is six months. The expected start of the work is the 1st February 2022, therefore the 31st July 2022 would be the latest submission date.

The processing of the thesis is based on the presented methodology according to Nunamaker, Chen, and Purdin [15] and its phases, from which the following schedule is derived:

- 01.02.22 - 06.03.22 (Observation): The first phase begins with the literature review and the development of the research questions and objectives.
- 07.03.22 - 10.04.22 (Theory Building): Next, both a data format for arguments is specified and requirements for the argument IDE are established.

- 11.04.22 - 15.05.22 (System Development): Building on the results of the previous phases, the prototype application is implemented.
- 16.05.22 - 19.06.22 (Experimentation): The results of the work will be validated and evaluated.
- 20.06.22 - 31.07.22 (Completion): Last, the thesis is finalized and printed. A buffer for unforeseen delays is already included.

The schedule serves as a rough guide and must be adjusted in case of unforeseen events (e.g., illness or delays due to other problems). The last phase is the longest with a duration of six weeks and therefore offers enough buffer to have sufficient time for proofreading as well as printing, binding and submitting the thesis even in case of delays of the previous phases.

The described schedule is visualized in the following Figure 3.1 as a Gantt chart. Each phase is associated with the sections of the thesis and the targeted objectives.

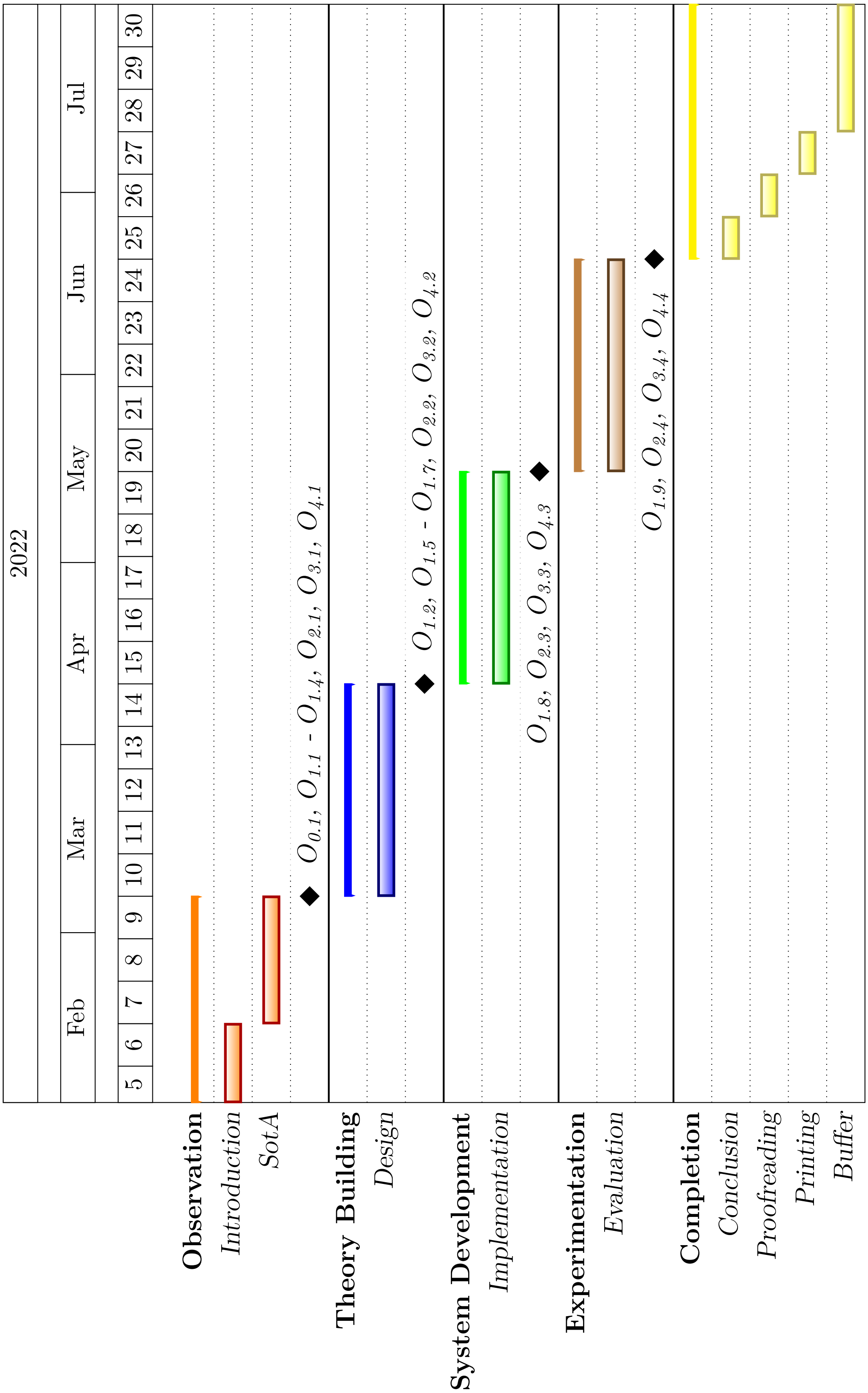

Figure 3.1: Thesis schedule

Index

Appendix

Bibliography

[1] Tim Berners-Lee. *The World Wide Web: A very short personal history.* 1998. URL: https://www.w3.org/People/Berners-Lee/ShortHistory.html/ (visited on 12/06/2021).

[2] Tim Berners-Lee, Roy T. Fielding, and Larry Masinter. *Uniform Resource Identifiers (URI): Generic Syntax.* RFC 2396. http://www.rfc-editor.org/rfc/rfc2396.txt. RFC Editor, Aug. 1998. URL: http://www.rfc-editor.org/rfc/rfc2396.txt.

[3] Tim Berners-Lee, James Hendler, and Ora Lassila. "The Semantic Web: A New Form of Web Content That is Meaningful to Computers Will Unleash a Revolution of New Possibilities". In: *ScientificAmerican.com* (May 2001). URL: https://www.scientificamerican.com/article/the-semantic-web/.

[4] Universität Bielefeld. *Rationalizing Recommendations (RecomRatio).* URL: http://ratio.sc.cit-ec.uni-bielefeld.de/projects/recomratio/ (visited on 11/16/2021).

[5] Dan Brickley and Ramanathan Guha. *RDF Schema 1.1.* W3C Recommendation. https://www.w3.org/TR/2014/REC-rdf-schema-20140225/. W3C, Feb. 2014.

[6] Tony Buzan and Barry. Buzan. *The mind map book.* English. BBC Books London, 1993, 320 p. ISBN: 0563863738.

[7] Ralf Hartmut Güting and Stefan Dieker. *Datenstrukturen und Algorithmen, 3. Auflage.* Leitfäden der Informatik. Teubner, 2004. ISBN: 978-3-519-22121-0.

[8] Steven Harris and Andy Seaborne. *SPARQL 1.1 Query Language.* W3C Recommendation. https://www.w3.org/TR/2013/REC-sparql11-query-20130321/. W3C, Mar. 2013.

[9] Stanley Jablonski. "The biomedical information explosion: from the index-catalogue to MEDLARS". In: *Bulletin of the Medical Library Association.* Vol. 59. PubMed, 1971, pp. 94 98.

[10] Holger Knublauch and Dimitris Kontokostas. *Shapes Constraint Language (SHACL).* W3C Recommendation. https://www.w3.org/TR/2017/REC-shacl-20170720/. W3C, July 2017.

[11] Jens Lehmann et al. "DBpedia - A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia". In: *Semantic Web Journal* 6 (Jan. 2014). DOI: 10.3233/SW-140134. URL: http://svn.aksw.org/papers/2013/SWJ_DBpedia/public.pdf.

[12] Murray Maloney et al. *W3C XML Schema Definition Language (XSD) 1.1 Part 1: Structures*. W3C Recommendation. https://www.w3.org/TR/2012/REC-xmlschema11-1-20120405/. W3C, Apr. 2012.

[13] Raquel Mochales and Marie-Francine Moens. “Argumentation mining: The detection, classification and structure of arguments in text”. In: Jan. 2009, pp. 98 107. DOI: 10.1145/1568234.1568246.

[14] Christian Nawroth. “Supporting Information Retrieval of Emerging Knowledge and Argumentation”. PhD thesis. Hagen, 2021. DOI: 10.18445/20210315-115252-0. URL: https://ub-deposit.fernuni-hagen.de/receive/mir_mods_00001694.

[15] Jay F. Jr. Nunamaker, Minder Chen, and Titus D.M. Purdin. “Systems Development in Information Systems Research”. In: *Journal of Management Information Systems* 7.3 (1990), pp. 89 106. DOI: 10.1080/07421222.1990.11517898. eprint: https://doi.org/10.1080/07421222.1990.11517898. URL: https://doi.org/10.1080/07421222.1990.11517898.

[16] *OWL 2 Web Ontology Language Document Overview (Second Edition)*. https://www.w3.org/TR/2012/REC-owl2-overview-20121211/. W3C, Dec. 2012.

[17] Iyad Rahwan and Chris Reed. “The Argument Interchange Format”. In: May 2009, pp. 383 402. ISBN: 978-0-387-98196-3. DOI: 10.1007/978-0-387-98197-0_19.

[18] Guus Schreiber and Yves Raimond. *RDF 1.1 Primer*. W3C Note. https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/. W3C, June 2014.

[19] Amit Singhal. *Introducing the Knowledge Graph: things, not strings*. 2012. URL: https://blog.google/products/search/introducing-knowledge-graph-things-not/ (visited on 12/06/2021).

[20] *SPARQL 1.1 Overview*. W3C Recommendation. https://www.w3.org/TR/2013/REC-sparql11-overview-20130321/. W3C, Mar. 2013.

[21] *The Open Graph protocol*. URL: https://ogp.me/ (visited on 12/06/2021).

[22] Henry Thompson et al. *W3C XML Schema Definition Language (XSD) 1.1 Part 2: Datatypes*. W3C Recommendation. https://www.w3.org/TR/2012/REC-xmlschema11-2-20120405/. W3C, Apr. 2012.

[23] Binh Vu et al. “A Metagenomic Content and Knowledge Management Ecosystem Platform”. In: Dec. 2019. DOI: 10.1109/BIBM47256.2019.9043355.

[24] David Wood, Richard Cyganiak, and Markus Lanthaler. *RDF 1.1 Concepts and Abstract Syntax*. W3C Recommendation. https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/. W3C, Feb. 2014.

Eidesstattliche Erklärung

Name:	Alexander Flammer
Matrikelnummer:	[redacted]
Studiengang:	Praktische Informatik
Modul:	Masterthesis

Ich erkläre, dass ich die vorliegende Abschlussarbeit mit dem Thema

Eine Autorenumgebung zur Unterstützung der Transformation und Repräsentation von emerging Argument Entity (eAE) Mindmaps in Aussagengraphen und darauf aufbauenden Argumentationsbäumen

selbstständig und ohne unzulässige Inanspruchnahme Dritter verfasst habe. Ich habe dabei nur die angegebenen Quellen und Hilfsmittel verwendet und die aus diesen wörtlich, inhaltlich oder sinngemäß entnommenen Stellen als solche den wissenschaftlichen Anforderungen entsprechend kenntlich gemacht. Die Versicherung selbstständiger Arbeit gilt auch für Zeichnungen, Skizzen oder graphische Darstellungen. Die Arbeit wurde bisher in gleicher oder ähnlicher Form weder derselben noch einer anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht. Mit der Abgabe der elektronischen Fassung der endgültigen Version der Arbeit nehme ich zur Kenntnis, dass diese mit Hilfe eines Plagiatserkennungsdienstes auf enthaltene Plagiate überprüft und ausschließlich für Prüfungszwecke gespeichert wird.

Ort, Datum　　　　　　　　　　Unterschrift