WIRTSCHAFTSUNIVERSITÄT WIEN
Vienna University of Economics and Business

Bachelor's Thesis

Title of Bachelor's Thesis (English)	Footology: A Football Ontology through Linked Open Terms Approach
Title of Bachelor's Thesis (German)	Footology: Eine Fußball-Ontologie durch den Linked Open Terms-Ansatz
Author (last name, first name):	Bacaj, Ardit
Student ID number:	11702367
Degree program:	Bachelor of Science (WU), BSc (WU)
Examiner (degree, first name, last name):	Dr. Fajar J. Ekaputra

I hereby declare that:

1. I have written this Bachelor's thesis myself, independently and without the aid of unfair or unauthorized resources. Whenever content has been taken directly or indirectly from other sources, this has been indicated and the source referenced.
2. This Bachelor's Thesis has not been previously presented as an examination paper in this or any other form in Austria or abroad.
3. This Bachelor's Thesis is identical with the thesis assessed by the examiner.
4. (Only applicable if the thesis was written by more than one author): this Bachelor's thesis was written together with

   The individual contributions of each writer as well as the co-written passages have been indicated.

14.10.2024
Date


Signature

WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

Bachelor Thesis

Footology: A Football Ontology through Linked Open Terms Approach

Ardit Bacaj

Date of Birth: 18.05.1997
Student ID: 11702367

Subject Area: Business Informatics

Studienkennzahl: H11702367

Supervisor: Fajar J. Ekaputra

Date of Submission: 14. October 2024

Department of Information Systems & Operations Management, Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria

DEPARTMENT FÜR WIRTSCHAFTSINFORMATIK
UND OPERATIONS MANAGEMENT DEPARTMENT
OF INFORMATION SYSTEMS AND OPERATIONS
MANAGEMENT

Contents

List of Figures

List of Tables

Abstract

Football is one of the most popular sports globally, and like many other domains, it increasingly integrates artificial intelligence (AI) technologies. Such integration increases the need for structured data representation to improve the quality of the result. Ontologies that facilitate the representation of knowledge of a domain could become essential in managing the complexity of football data. However, existing ontologies related to football often fail to cover a wide area that is needed and focus primarily on limited aspects such as player and team. This thesis aims to reduce the gap above by developing a football ontology following the Linked Open Terms (LOT) Methodology, where we engage key stakeholders such as ontology developers, domain experts, and users to identify the requirements. The developed ontology facilitates data representation from various football aspects, providing a standardized framework that supports querying and football analytics. Our initial evaluation shows that the ontology outperforms existing ontologies due to its broader representation of football knowledge. The ontology will serve as a foundation for future research that could extend the ontology to integrate real-time data, allowing for live match updates, enhancing decision-making during games, and supporting dynamic analysis based on current football events.

1 Introduction

Football is one of the most popular and passionately followed sports in the world [5]. Nowadays, football makes use of various technologies and as technology continues to improve, new opportunities such as Artificial Intelligence (AI) arise [8]. In recent years Artificial Intelligence has gained popularity, which has also found its way into football and has been revolutionizing football in many ways, for example player recruitment, statistical analysis, performance analysis, decision-making processes [8].

Most AI Applications in football are mainly based on sub-symbolic approaches, utilizing machine learning, deep learning and data-driven techniques [7]. However, the latest trends in AI shows a new trend of integrating sub-symbolic with symbolic approaches, which is not yet widely applied in Football. With introduction of symbolic AI, new paths emerge, e.g., the ability to represent (football) knowledge, relationships, and concepts in a structured, semantically meaningful, and human-readable way. Furthermore, ontologies can help with integration and representation of complex relationships within football data [1], which provides good support in advanced analytics and much more. Existing ontologies, like Sport-Ontology[1] and Semweb-football[2], concentrate mostly on team and player attributes, but they don't fully cover wider topics like tactical insights and match events. This gap highlights the need for a standardized football ontology that can capture the complexity of football knowledge.

This thesis aims to provide a comprehensive football ontology, offering insights into –among others– player, teams, and matches. As briefly described in the introduction, an ontology is a formal representation of knowledge about a specific domain, in this case the domain is about football, which contains concepts like player, team, matches, etc. where each concept has definition of its attributes, as well as relationship between the concepts or classes [1]. Introducing ontologies and knowledge graphs into the world of football is challenging because there was a lack of standardized and widely accepted ontologies tailored to football-specific knowledge across the world wide web. Although there is a large amount of football- related data accessible across the internet, there is apparently still a deficit of football ontologies that represent this data as knowledge graphs.

The main research question of this bachelor thesis to address the gap on the lack of availabilities of ontologies for football is *"How to develop a football ontology following the Linked Open Terms (LOT) method?"* In order

[1] `https://github.com/costinbusioc/Sport-Ontology`

[2] `https://github.com/sLeeNguyen/semweb-football`

to support this main research question, there are several sub-research questions that explore the requirements for building an ontology for football. The sub-research questions are:

- RQ1: What are the requirements for building an ontology for football?
- RQ2: What is the current state of the art on the ontology to represent football data?
- RQ3: How to evaluate the developed football ontology?

The thesis will follow the Linked Open Terms (LOT) methodology, which contains four main steps, the ontology requirements specification, ontology implementation, ontology publication, and ontology maintenance. In the first step, an analysis of the key stakeholders such as ontology developers, domain experts, and ontology users was carried out. Another important step is to search through ontology knowledge sources and look for already existing football ontologies that can be helpful to consider for this ontology development or reuse some of the concepts, more information about this step are shown in Section 3.2.2 and Section 4.3.

The thesis resulted in the development of a comprehensive football ontology that contains player and team statistics, match events, and other relevant football specific information, distinguishing it from other existing ontologies that only include limited aspects of football. The ontology provides a structured framework that can support analytics in football and enables flexible querying and reasoning over football data.

This thesis is structured as follows: In Section 2, the related work on ontology development in football is discussed. Section 3 reviews the methodology used to develop the ontology. Section 4 shows the results, including the ontology itself and its evaluation. The evaluation of the ontology is shown in Section 5. Finally, the discussion of this thesis is presented in Section 6.

2 Related Work

The goal of this section is to highlight the research gap in the development of football ontologies. To be specific, existing ontologies cover only smaller parts of the football domain instead of its broad aspects. Relevant works categorized by topics to show the need for a comprehensive football ontology will be reviewed.

2.1 Ontology Engineering Frameworks

Ontology engineering frameworks provide structured methodologies for developing ontologies. There are several frameworks available, each one has its own strengths and weaknesses. Below are three ontology engineering frameworks:

1. **Linked Open Terms (LOT) Methodology:** The Linked Open Terms (LOT) methodology, which is a lightweight approach focused on aligning ontology development with industrial practices [14]. The main steps in the LOT methodology are:

    - **Ontology Requirements Specification:** This step involves collecting and defining the requirements that the ontology should fulfill. Techniques such as Competency Questions, natural language statements, and tabular information are used to describe ontology requirements. This process ensures that the ontology meets the specific needs of the domain it is designed for.
    - **Ontology Implementation:** During this phase, the ontology is built using a formal language based on the requirements identified. This involves creating classes, properties, and relationships that represent the domain knowledge accurately. The ontology development team collaborates closely with domain experts to ensure the ontology's accuracy and relevance.
    - **Ontology Publication:** The aim of this step is to provide online access to the ontology. The ontology is made available both as human-readable documentation and as a machine-readable file from its URI. This ensures that the ontology can be easily accessed and reused by other researchers and practitioners.
    - **Ontology Maintenance:** Ontologies need to be updated periodically to remain relevant and accurate. The maintenance phase involves updating the ontology during its life-cycle based on new information, feedback from users, and changes in the domain knowledge.

    The LOT methodology has been applied in at least 18 projects, demonstrating its flexibility and effectiveness in various industrial contexts. By involving different roles such as ontology developers, domain experts, and ontology users, the LOT methodology ensures a comprehensive approach to ontology development [14].

2. **Ontology Development 101:** Ontology Development 101 is a simplified methodology for creating ontologies. The steps include:

   - **Determine the domain and scope of the ontology:** Definition of what the domain of the ontology will cover, what the usage of the ontology should be, who will maintain the ontology.
   - **Consider reuse:** Identifying existing ontologies that could be reused or adapted
   - **Enumerate important terms in the ontology:** Identifying terms that are important for the ontology creation.
   - **Define the classes and the class hierarchy:** This process considers the definition of concepts and the hierarchy of the ontology, followed by defining properties and relationships.
   - **Create instances** for each class.

   Comparing this methodology with LOT, Ontology Development 101 lacks the industrial orientation and maintenance found in LOT. Ontology Development 101 is simple and effective for small projects, but may not be robust for complex domains like football [12].

3. **Methontology:** The Methontology framework is a comprehensive step-by-step approach for developing ontologies and has been applied in both academic and industrial settings. The following steps are:

   - **Specification:** This step includes the purpose and scope of the ontology.
   - **Conceptualization:** During this phase, the structure of the ontology is defined, including classes, properties, and relationships.
   - **Formalization:** Conversion of the conceptual model into a formal model using a knowledge representation language like OWL.
   - **Implementation:** Building the ontology in a specific tool.
   - **Evaluation:** This steps ensures that the ontology meets the initial requirements.

   Methontology can be more resource-intensive and time-consuming compared to LOT, particularly for lightweight or industrial-focused projects. This methodology is best suited for projects which require highly detailed, structured ontology from scratch [2].

For this thesis, the LOT Methodology was selected because of its lightweight, flexible nature and focus on reusing existing ontologies, making it suitable for this football domain. In contrast to the other two methodologies presented, LOT balances scalability and flexibility, allowing for quick implementation and ongoing updates, which are important for a domain like football.

2.2 Data-Driven Approaches in Sports Analytics

Several works have applied data-driven approaches to evaluate player actions and tactics in football, providing valuable insights for the development of football ontologies:

1. **Valuing Player Actions:** In the Paper by Decroos et. al [3], they propose a novel framework that evaluates football players based on their actions during games. The framework considers various actions such as passes, crosses, dribbles, take-ons, and shots. By collecting player performances, this approach provides a detailed analysis of player contributions to the game. Although the primary focus is on player performance, the methodology indirectly supports football ontology development by providing a structured way to represent and analyze player actions [3].

2. **Tactical Analysis:** In the Paper by Yamanaka et. al [18] they focus on the use of objective data and mathematical models for tactical analysis based on the location of players and the ball. Their framework involves filming competitive matches and analyzing team tactics, including offensive and defensive phases of the game. The data is presented in numeric form, representing the dynamically changing locations of players and the ball in spatial coordinates. This approach provides a detailed understanding of team tactics, which can be integrated into a broader football ontology to enhance tactical insights [18].

The data-driven approaches described above help to construct a football ontology by providing systematic methodologies for analyzing player actions and team tactics. These frameworks offer a basis for representing football-related data in a formalized way. By integrating these methods into the football ontology, this thesis intends to create a more comprehensive representation of the football domain.

2.3 Ontology Development and Existing Ontologies in Sports

Ontology development within the sports domain provide insights into methodologies and applications that are relevant to football ontology development:

1. **Ontology for Football using Oriented Programming:** A method for constructing football ontologies using oriented programming was proposed. This method involves creating ontologies using Python, OWL 2, and various tools such as Owlready, NLP, and Protege. The ontology is visualized using OntoGraph tools available as a plugin in Protege. This approach demonstrates the effectiveness of oriented programming in developing complex ontologies for sports domains, highlighting the importance of using advanced tools and programming languages in ontology development [1].

2. **Event Detection using Graph Convolutional Networks (GCNs):** This approach focuses on automatic event detection from football videos using GCNs. The method involves modeling players and the ball in each frame of the video as a graph. By experimenting with different pooling methods for modeling the temporal context around each action, the authors demonstrate the potential of graph-based modeling techniques for football entities. This approach provides valuable insights for building a robust football ontology that can support event detection and analysis [16].

3. **Ontology for Ice Hockey:** An example of ontologies in sports was illustrated by Keskisarkka et. al [9]. The authors develop an ontology for ice hockey, which, despite being a different sports domain, provides valuable insights for football ontology development. The authors use a methodology that aligns with the LOT methodology, highlighting the benefits of using ontologies for knowledge representation and capturing information in sports analytics. The ontology development process includes steps such as describing use cases, specifying competency questions, formalizing the ontology, and validating it. This approach emphasizes the importance of a structured methodology in developing comprehensive sports ontologies [9].

4. **Semweb-Football Ontology:** One example of a football related ontology, is semweb-football [3], which is a simple ontology developed by "sLeeNguyen" and provides a valuable foundation for understanding

[3] `https://github.com/sLeeNguyen/semweb-football`

how ontology structures can be employed to represent football-related concepts. Analyzing such examples not only helps in comprehending the practical implementation of ontological structures but also underlines the need for more comprehensive and specialized football ontologies.

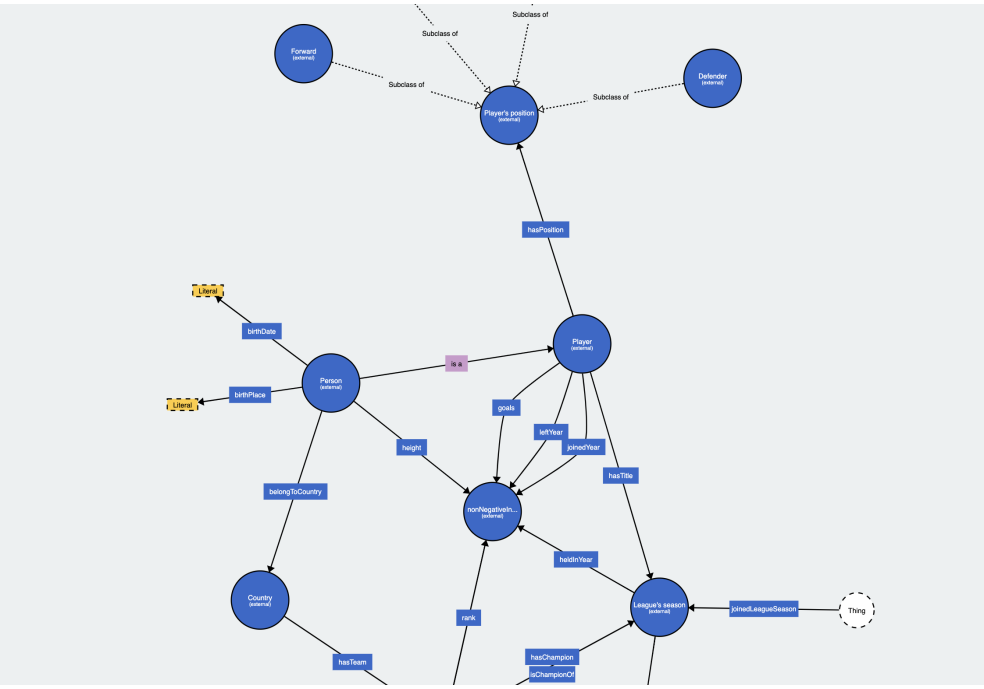

Figure 1: Ontolgy Example: Semweb-football

In the example showed in Figure 1, there are several basic concepts that can be seen such as Person, Player, Player's Position, League's Season, Country and more concepts that are not shown in the screenshot in Figure 1. Examples like these are important, because they provide a better understand how to structure and organize the football ontology. The information derived from other existing ontologies can be used for developing a comprehensive football ontology.

5. **Sport-Ontology:** Another interesting ontology was found on GitHub by the user "costinbusioc" called "Sport-Ontology"[4]. This ontology not only focuses on football, but also includes tennis players. This ontology was helpful with the structure of the sports domain in general as well, which two or three classes were reused and adapted to the Footology.

[4] `https://github.com/costinbusioc/Sport-Ontology`

The ontologies mentioned above provide useful information and insights, but they either focus on specific aspects of football, such as player actions or team data, or focus on different sports. This thesis aims to develop a comprehensive football ontology that integrates player statistics, team performance, and match events. In summary, the existing ontologies and frameworks provide useful approaches for specific aspects of the football domain. But there is a gap in the development of a comprehensive football ontology that cover the broad aspects of football. The ontology in this thesis will address this gap by integrating data-driven approaches and symbolic representations.

3 Methodology

In Section 2.1, the process of LOT methodology was discussed. In this chapter, the aim is to briefly describe how the LOT methodology is going to be carried out by discussing each step. As mentioned before, the four main steps or iterations in LOT are ontology requirements specification, ontology implementation, ontology publication and ontology maintenance. Each of these steps contain several sub-steps which will be discussed below.

3.1 Ontology Requirements Specification

This phase is a critical step in developing an ontology because it establishes the foundation on which the ontology will be built. The Ontology requirements specification includes identifying the key requirements, collecting input from stakeholders, and understanding the specific needs that the ontology should meet. This process begins with a use case specification, which includes identifying key stakeholders.

3.1.1 Use Case Specification

In the initial phase of ontology development, a comprehensive literature review on football analytics and existing ontologies is underway. This crucial step serves as the foundation for understanding the landscape of football-related data and ontology structures.

Defining key stakeholders is essential to the success of the ontology. The stakeholders include:

1. **Ontology Developer:** The ontology developer will actively engage in the creation process, including planning, information gathering, and implementation.

2. **Domain Experts:** Individuals with specialized knowledge in the football domain and football analytics, who can provide expert insights and validate the ontology's relevance.

3. **Ontology Users:** Users who will interact with the final product and provide feedback on its usability and functionality.

Domain experts are going to be identified and engaged who can provide valuable insights into statistical modeling in the football domain and the collection of sports data. To ensure the relevance of the ontology, the plan is to connect with university colleagues, who took the same courses in ontology development, who will act as test users. These colleagues will provide valuable insights and constructive feedback on the ontology's usability and functionality. The interaction of (test) users is an important step of refining and enhancing the ontology, ensuring that it aligns with real world expectations.

As the use case specification makes progresses, these interactions and engagements with stakeholders will play a vital role in shaping the design of the ontology, functionality, and effectiveness. The valuable insights and feedback collected from stakeholders will serve as a cornerstone for optimizing the ontology to align with the needs and expectations of end-users, ensuring its relevance and practicality.

3.1.2 Data Exchange Identification

This chapter focuses on the systematic acquirement of relevant information about football data. The methodology will involve an extensive search and analysis of various knowledge sources containing sports related data. The method will begin with identifying important knowledge sources that contain ontologies and football related data, such as Wikidata, DBpedia, GitHub, and Linked Open Vocabularies. These knowledge resources are going to be used for identifying existing classes, relationships, attributes, and data structures related to football. There is also going to be a research of existing ontologies, which can be helpful as a starting point of this bachelor thesis. This process is expected to be challenging due to the potential lack of existing football ontologies, but it will be essential for identifying gaps in this field and providing a new, more comprehensive football ontology.

3.1.3 Purpose and Scope Identification

The first important step is to highlight and explain what the football ontology is for and to identify common purposes and scopes. The purpose is to create a standard way to represent football knowledge in the form of an

ontology, which can be used to represent the data in a knowledge graph. This step is important for many reasons. First, it ensures that the ontology is designed with a clear focus, which would be helpful in meeting the specific needs of its intended users, such as analysts or researchers. A well-defined purpose and scope also contribute to more efficient development, because they provide a clear overview what the ontology should include and exclude. Furthermore, by stating the scope and purpose potential misunderstandings or misalignments between stakeholders can be avoided.

3.1.4 Functional Ontological Requirements Proposal

In this phase of the methodology, the focus will be on defining the functional requirements for the football ontology. A structured approach will be employed. First, a set of competency questions (CQs) will be created. These competency questions will be gathered and derived from a review of existing literature, research findings, and insights into football data.

After the set of competency questions are created, interviews with domain experts are going to be conducted in order to review and prioritize the CQs. They will provide helpful feedback which will be crucial for refining the scope of the ontology, ensuring that it focuses on the important aspects of football. Additionally, discussions with the thesis supervisor will be held to review the competency questions and integrate any additional feedback. This process is important to ensure that the football ontology is comprehensive and helpful for its intended users.

3.1.5 Functional Ontological Requirements Completion

Functional Ontology Requirements Completion is a pivotal stage in the ontology development process, where the proposed requirements undergo careful examination in order to make sure they are aligned with the ontology's objective and integrated with its elements. The process involves several steps in validating each requirement and creating connections with ontology elements, resolving conflicts, and verifying with domain experts.

At the start, all the proposed requirements from the Function Ontology Requirements Proposal undergo validation to confirm their relevance. Each requirement is carefully examined, which is a crucial step to ensure that it reflects the intended functionality of the ontology and remains with the project goals.

Subsequently, the validated requirements are linked to specific elements of the ontology, such as concepts (classes), the attributes of each entity, the relationship between entities, and other relevant components. During this

step, each requirement addresses the ontology's intended elements, which leads to more clarity.

3.1.6 ORSD Formalization

This step focuses on formalizing the Ontology Requirements Specification Document (ORSD). In the ORSD, all requirements, competency questions, intended use, and more are organized and documented. The ORSD will serve as the definition document ensuring that all aspects of the ontology are clearly defined and aligned with the thesis' goals. The document will use a template provided by the LOT methodology to ensure consistency and completeness.

3.2 Ontology Implementation

This section focuses in the creation and implementation of the ontology from conceptualization to realization, detailing the steps involved in its development, deployment, and utilization. In this phase of the project, theoretical aspects and ideas gained from the requirements specification section are going to be used for the ontology conceptualization, encoding, and ontology evaluation.

3.2.1 Ontology Conceptualization

In the ontology development, the conceptualization serves as the foundational stage where football knowledge is turned into formal representation. Within the LOT Methodology, the ontology conceptualization plays a significant role in defining the scope, the concepts, and specifying the relationships needed to develop the ontology.

During this phase, the concepts found and identified in the earlier phases are refined and structured for example categorizing fundamental elements of football such as players, teams, matches and structuring the interaction between these elements within the ontology. These concepts will have relationships with other concepts in a way that reflects the real-world dynamics of the football domain.

3.2.2 Ontology Reuse

This Process will handle the reuse of existing sports or football related ontologies aimed at enhancing efficiency, consistency, and interoperability. This phase begins with a research on existing ontologies using available ontologies in repositories such as DBpedia, Wikidata, GitHub, and other sources. The

goal is to find ontologies with a similar domain that include relevant concepts, which can be integrated to this ontology.

By using and building upon existing ontologies, the established knowledge can be leveraged, which avoids redundancy and improves the quality of the ontology

3.2.3 Ontology Encoding

In the encoding phase of the ontology, the focus will be on translating the conceptualized elements of the football ontology into a formal, machine-readable format. This will be integrated with the ontology conceptualization phase to enhance the efficiency. The encoding will be done by using OWL (Web Ontology Language), with the help of WebVOWL, which is a user-friendly interface for creating and visualizing ontologies, allowing for adjustments and refinements during the conceptualization phase. The ontology will be encoded directly within WebVOWL, whill will be downloaded as .ttl (Turtle) files.

3.2.4 Ontology Evaluation

In the ontology development process, the evaluation phase is crucial for ensuring the quality and overall effectiveness of the ontology. The method for the evaluation will include multiple approaches, ensuring that it meets the predefined goals and objectives.

The evaluation will begin with testing the ontology using specific test data in order to answer the competency questions. A dataset will be created which represents the classes and properties of the ontologies and then mapping this data onto the ontology structure using tools like Ontotext Refine and GraphDB. Ontotext Refine is a practical tool that is used for data transformation and linking or mapping [13] and GraphDB is a powerful graph database engine that supports graph data management.

Additionally, the ontology will be evaluated for potential errors using the OOPS! (Online Ontology Pitfall Scanner) tool. OOPS! uses a set of predefined heuristics to analyze ontology structures and detect errors in ontlogy design [15].

Finally, evaluation from domain experts will be carried out, where the domain experts will review ontology's structure, completeness, and usability. Interviews and feedback sessions will be held, which will provide insights on potential improvements and validate that the ontology effectively captures the necessary aspects of the football domain.

3.3 Ontology Publication

Ontology publication is crucial for understanding how to share and distribute the ontology so it can be accessed, reused, and integrated by other researchers and developers. In this chapter, it will be discussed how the ontology is going to be published. First of all, a release candidate is going to be proposed, then, an HTML documentation will be created. Subsequently, the the ontology will be published online.

3.3.1 Propose Release Candidate

After the ontology evaluation is done and the issues are identified, I will share the ontology with some of my colleagues at the WU Wien, who also took the courses K2 and K3 in Semantic AI Technologies for Knowledge Management and have knowledge about ontologies, who will give me feedback about the created ontology.

3.3.2 Documentation

The documentation process is an important step to ensure transparency, maintainability, and easy to understanding for future users and developers. The documentation process will involve generating a detailed HTML page using WIDOCO (WIzard for DOCumenting Ontologies), a tool specifically designed to simplify and automate the documentation of ontologies [6]. Within the HTML page created by WIDOCO, a short description will be provided as well as an overview of the ontology and all of the related essential details such as latest version, author information, license, and other relevant metadata. This documentation will also include a description about for each class, outlining their domains, ranges, and specific roles within the ontology. This structured approach ensures that the ontology is easily understandable and usable by others,

3.4 Ontology Maintenance

The maintenance of the ontology is an important step to keep the ontology up-to-date and remove as many errors as possible. This chapter is divided into two steps: Bug Detection and New Requirements.

1. **Bug Detection:** The ontology will be monitored addressing the issues and bugs that may occur in the ontology, however, after submission of the bachelor thesis and the creation of the ontology, a detailed bug detection will likely not be carried out. Instead, the testing of the

ontology will be done during the ontology creation phase to minimize potential issues.

2. **New Requirements:** The plan in this phase would be to stay informed about the football domain and constantly update the ontology based on new information in order to keep the ontology up to date.

4 Results

The results of the ontology development process are presented in this section, including the specification of use cases, purpose and scope, and the completion of functional requirements. Each subsection shows the outcomes that contributed to the development of a comprehensive football ontology.

4.1 Ontology Requirements Specification

The ontology development follows a structured approach to ensure the effectiveness and clarity of the ontology. The focus of this chapter are the main steps that were carried out to define, refine, and document the requirements for the football ontology. The sections below describe the use case specification, the process of data exchange identification and scope definition, the proposal and completion of functional ontology requirements, and the formalization of the Ontology Requirements Specification Document (ORSD).

4.1.1 Use Case Specification

During the use case specification phase, the following outcomes were achieved. A comprehensive review of existing literature on football analytics and ontologies was conducted. This review included academic papers, industry reports, and existing ontologies related to football.

As for the stakeholder identification, key stakeholders were successfully identified and engaged. Finding **domain experts** in the football domain was not easy. Thanks to my supervisor of this thesis, two domain experts, where provided to me.

University colleagues who also took courses in ontology development were involved as test users. Their feedback on the ontology's functionality were valuable. They provided feedback on the ontology structure, which helped identify areas where the ontology could be simplified.

4.1.2 Data Exchange and Scope Definition

In the process of data exchange identification, the objective is to systematically acquire relevant information about football data. During this phase, a comprehensive search through several knowledge sources such as DBpedia, Wikidata, GitHub, or Linked Open Vocabularies was conducted. This step was quite challenging, while those knowledge resources offered a wealth of information and various classes related to football, there is only a small amount of comprehensive football related ontologies published. Despite high amount of classes and data concerning football, the absence of comprehensive football ontologies highlights a significant gap in the existing resources. This scarcity underscores the necessity for the development of a dedicated football ontology, which will contribute to the advancement of knowledge representation in the football domain.

The main objective of this bachelor thesis is to develop a comprehensive and standardized framework for representing football knowledge through the creation of an ontology. The developed ontology will serve as a foundation for organizing football data into a knowledge graph, promoting a systematic representation. With the finished football ontology, the aim is to facilitate seamless integration with various football-related applications.

The scope of this ontology focuses on the key concepts within the football domain. Defining the scope is crucial as it shows the range of the ontology. The scope of the ontology is the following:

- **Player Attributes:**
  - Basic Information: Name, Date of Birth, Nationality, and preferred Playing Position
  - Performance Metrics: Metrics: Goals, Assists, Tackles, Dribbles
- **Team:**
  - Basic Information: Name of Team, Country, League, Formation
  - Ranking: Current ranking in the league, Ranking in tournament, Number of titles
- **League:**
  - Scope: Only Top 5 Leagues of Europe (Bundesliga, Premier League, La Liga, Serie A, and Ligue 1), Table (Ranking), Number of teams, Start-End Date, Results

  - Tournaments: Coverage of major tournaments such as FIFA World Cup, UEFA Euro Cup, Champions Leauge, Europa League, Conference League

- **Match:**
  - Match Metadata: Date, Kick-Off times, Details such as League Game or Cup Game and which teams played against each other
  - Match Events: Goals, Fouls, Yellow and Red Cards, Pass Accuracy

- **Referee:**
  - All referees from that specific league (Top 5 leagues) including assistant referees, Name, Age, Nationality, the matches they officiate

- **Coach:**
  - All team (club) coaches of the top 5 leagues, Coaching history, Current team, Coaching style (tactical approach, preferred formation)

All those concepts or entities are going to have a specific relationship between each other. Table 1 shows the basic concepts that this ontology will represent, but there are still more concepts which are not shown in this table.

Concept	**Attributes**
Person	Name, Age, Birthday, Height, Nationality
Player	Team, Strong Foot, Awards, Titles
Team	Team Name, Coach, League, Ranking, Titles, Stadium
League	League Name, Country, Number of Teams, Title
Knock Out Tournament	Structure, Number of Teams

Table 1: Basic Concepts

By defining the scope and purpose of the ontology, it aims to address the diverse needs of users within the football community. A clearly defined scope mitigates the risk of errors and misunderstandings, promoting a shared understanding among stakeholders involved in the football community.

To summarize this sub-chapter, the purpose of this ontology is to create a standardized representation of football knowledge, while the scope encompasses players, teams, and associated attributes. This approach aligns with the goal of enhancing user experience, promoting efficient development, and minimizing errors in football data representation.

4.1.3 Functional Ontological Requirements Proposal and Completion

During the phase of the requirements specification, the focus was on defining functional requirements for the football ontology. To do this, a structured approach has been done by coming up with competency questions (CQs) to guide the ontology's development.

In the beginning, a set of 30 competency questions were derived based on existing literature, research findings, and some of my own ideas about football data. These questions were designed to cover a wide range of topics, from player details to match events and more, trying to capture the core essence of football-related knowledge.

The next step was to conduct an interview with domain experts to refine the scope of this project. The first interview was held on Zoom with the first domain expert. Each competency question was examined and was prioritized by the importance or priority, which was set to either high, medium, or low. The next interview with the second domain expert was similar to the first interview, where all 30 competency questions were examined. At the end of both interviews, a more refined understanding and clearer scope of the ontology were gained. In a personal meeting with the supervisor, the competency questions were again discussed and suggestions for improvement were made.

Through these interviews, a valuable perspective from experts who work with football analytics and ontology development was gained. Their input did not only validate the initial competency questions but also helped to design the ontology to better meet the needs of potential users.

Overall, these interviews were a crucial step in defining the functional requirements of the football ontology. The collaboration with experts improved this project, making sure that the ontology would be robust and user-friendly for people in the football-community.

In the following, Table 2 shows some of the competency questions that were created within this project. The single competency questions are listed in its corresponding column, followed by the Answer. The last column shows the priority, that were assessed by the domain experts during the interview. For example, Medium / Low means that the first domain expert prioritized the competency question as Medium, and the second domain expert decided that the priority of the CQ is "Low". The assessments of the priorities were a crucial step in finding the correct scope of the ontology, as the input came from domain experts who actively work with football-related data.

Getting in touch with domain expert played a crucial role in this process, with valuable input coming from all three domain experts, validating the

Competency Question	**Answer**	**Priority**
What team is Player X currently playing for?	Manchester City	Low / Medium
How many goals did Player X score in the season 22/23?	23	High / High
What playing position is Player X currently playing?	Striker	High / High
What League ranking does Team X have in Season 23/24?	2	High / Medium
What is the Pass Accuracy of Player X?	90%	High / Low

Table 2: Competency Questions

relevance of the competency questions created in the Functional Ontology Requirements Proposal stage. Valuable inputs of those domain experts ensured that the requirements were relevant and important enough to be used in the ontology.

Steps have also been conducted for resolving any conflicts or inconsistencies among the requirements. The supervisor pointed out the lack of quality in the scope definition during the second meeting, which were not clear at the beginning. He provided very helpful input, which was then used to improve the scope identification to ensure clarity and consistency for the later process of the ontology development.

During the completion phase, documentation and reporting are maintained to capture the validation process and outcomes. For the competency questions, a template for requirements was used provided by the LOT GitHub Repository. Another crucial template is MODA template, also provided by the LOT GitHub Repository, which lists all the entities (or concepts), their attributes and the respective relationships between them. Furthermore, an ORSD document has been created with all the documentation and process of the ontology development. More detailed information about the ORSD document and the templates will be discussed in Section 4.1.4.

By finishing the Functional Ontology Requirements Proposal and Completion step, the ontology development includes a robust basis for the ontology's further development and advancement. With validated requirements integrated in the ontology, this project slowly moves towards the goal of delivering a robust and effective ontology.

4.1.4 ORSD Formalization

The Ontology Requirements Specification Document is a highly important document for reporting every step that has been done in the requirements specification, including the creation of competency questions, MODA document, description of intended use and so on. The primary focus of ORSD is to document the requirements identified during the early stages of the project into a structured format.

The first template that was used for the ORSD, was the template domain requirements, where a set of competency questions was identified and captured in this document, as mentioned in Section 3.1.4. This, and other ORSD documents, can be found on the Footology GitHub Repository[5]. By creating the competency questions, valuable insights from domain experts were gained, which was crucial in finding the right scope and constraints of the ontology.

After completing the domain requirements stage, another important document needs to be drafted, the MODA document. The MODA document highlights all the entities used in the ontology, followed by the attributes and later the relationship of the entities identified. This is a crucial step for providing a foundation for the conceptualization stage, because the MODA consists of the relations that can be used later in the ontolgoy development. Table 3 shows a small excerpt the MODA document, this is one of the three excel folders provided by MODA, in this case this represents the relations between the entities. In the first column, the Material entity is listed, followed by the second column named "Material Relation", which connects the first entity with the second entity shown in column three "Target entity".

After all the information about the requirements, obtained from the purpose and scope identification, ontological requirements proposal, and ontological requirements completion have been gathered, the Ontology Requirements Specification Document (ORSD) is created. This document includes all functional and non-functional requirements identified and the information associated to them [14].

Figure 2 shows the first page of the ORSD document created for this football ontology. This template was downloaded from the LOT GitHub Repository, which was used as a starting point for this project. The Figure 2 shows the purpose, the scope, and the implementation language used in the ontology development. The document goes beyond that and shows more information like intended end-users, functional and non-functional requirements and can be accesses in the Footology GitHub Repository[6]

[5] `https://github.com/arditb1997/footology/tree/main/ORSD`

[6] `https://github.com/arditb1997/footology/tree/main/ORSD`

Ontology Requirements Specification Document	
1	**Purpose**
	The purpose of our ontology is to organize football-related knowledge, offering a structured framework for easy access and analysis. By organizing player details, team attributes, match data, and more, this ontology aims to provide a user-friendly resource for football enthusiasts, analysts, and researchers.
2	**Scope**
	Player Attributes: - Basic Information: Name, Date of Birth, Nationality, and preferred Playing Position - Performance Metrics: Goals, Assists, Tackles, Dribbles - Career Path: Clubs played before, National Team **Team:** - Basic Information: Name of Team, Country, League, Formation - Ranking: Current ranking in the league, Ranking in tournament, Number of titles, **League**: - Scope: Only Top 5 Leagues of Europe (Bundesliga, Premier League, La Liga, Serie A, and Ligue 1), Table (Ranking), Number of teams, Start-End Date, Results, - Tournaments: Coverage of major tournaments such as FIFA World Cup, UEFA Euro Cup, Champions Leauge, Europa League, Conference League, **Match:** - Match Metadata: Date, Kick-Off times, Details such as League Game or Cup Game and which teams played against eachother, - Match Events: Goals, Fouls, Yellow and Red Cards, Substitutions, Shots, Shots on target, Possesion Percentage, Fouls commited, Corners, **Referee:** - All referees from that specific league (Top 5 leauges) including assistant referees, Name, Age, Nationality, the matches they officiate, **Coach:** - All team (club) coaches of the top 5 leagues, Coaching history, Current team, Coaching style (tactical approach, pereferred formation)
3	**Implementation Language (optional)**
	The ontology will be implemented using OWL (Web Ontology Language), a

Figure 2: ORSD Document

Material entity	Material Relation	Target entity
Player	playsFor	Team
Team	includesPlayer	Player
Player	hasPosition	Position
Position	isPositionOf	Player
Team	partOf	Tournament
Tournament	hasTeam	Team
Team	competesIn	Match
Match	isWonBy	Team
Team	playsIn	Stadium
Stadium	isHomeOf	Team
Coach	manages	Team
Team	managedBy	Coach
Referee	officiates	Match
Match	officiatedBy	Referee
Match	hostedBy	Stadium

Table 3: MODA Document

4.2 Ontology Conceptualization

After the first version of the MODA document was created, the first attempt was to sketch an ontology on paper. Based on the sketches, a first digital conceptualization has been approached. Although, the first attempt for developing the ontology using the tool Draw.io, which is a powerful tool for building diagramming applications, and the worlds most widely used browser-based end-user diagramming software [4]. It quickly became clear that working with WebVOWL would be a better option, since I had more experience creating ontologies with this tool. WebVOWL is a convenient web application for the visualization and representation of ontologies. It is based on open web standards and implements the Visual Notation for OWL Ontologies (VOWL). JSON Files are used to generate the visualizations into which the ontologies need to be converted [10]. WebVOWL is a free ontology development tool that can be accessed online [7].

First of all, the key classes were created in WebVOWL by simply double clicking on the empty project. The key classes Team, Player, Coach, Referee, Stadium, Match, League, and Performance Stats were created and the relationships (object properties) between each class. For each class, datatype properties are created as shown.

[7] `https://service.tib.eu/webvowl/`

Figure 3 shows the first version of the ontology created in WebVOWL based on the MODA document and is split into two separate images for a clearer view. The ontology is structured with the core concepts such as Player, Team, Match, League, and Tournament serving as central nodes. These classes are connected by specific object properties such as "playsFor" for a player's participation in a Team or "manages" which involves the coach managing a team. Attributes or datatype properties are defined for each entity. For example an entity "Player" is associated with the attribute "Height, Player Name, Player Position, etc.". The transition to WebVOWL enabled a more seamless and effective mapping of the relationships between the various classes in the ontology. This user-friendly interface greatly enhanced the development process, ensuring a clearer, more accurate representation of the domain.

The first version was a good starting point as it contained the main classes that needed to be in the ontology. Following an evaluation of the initial version, the next step was to improve and extend the ontology. Some relationships between certain entities were removed as they did not fit in the ontology. For example the relationship "participates in" between the class "Player" and "Match" was removed, because the player already is connected to the "Team" class and "Team", in turn, is connected to "Match".

Adding more datatype properties was another step in advancing the ontology, enriching the depth of the model by adding specific attributes to the classes defined within the ontology. For example attributes such as Team Ranking, Goals Scored, or Assists. Incorporating such attributes provides a comprehensive dataset for analytical purposes. The MODA document was updated simultaneously during this step, ensuring the ORSD document to align with the ontology and avoid any errors.

In the final version of the football related ontology shown in Figure 4, a network of classes and sub classes is represented, each connected through a series of relationships, capturing the hierarchical structure of football data. Sub-classes such as "Left Wing" or "Right Wing" have been added to ensure a more detailed representation of the Player's playing position. The "center" of the ontology is the "Player" class, which is connected to the "Team" class, and indirectly to the "Match" class through their teams, illustrating their participation in the football's competitive events. The relationship between "Player" and "Performance Stats" is highly important, as it represents the individual performance of each player such as goals scored, assists, fouls, and so on. This class changes frequently throughout the season based on the player's performance in the team. Other classes or attributes like player information or the team a player is playing for, rarely change, for example once in a season even do not change at all. The relation between "Player"

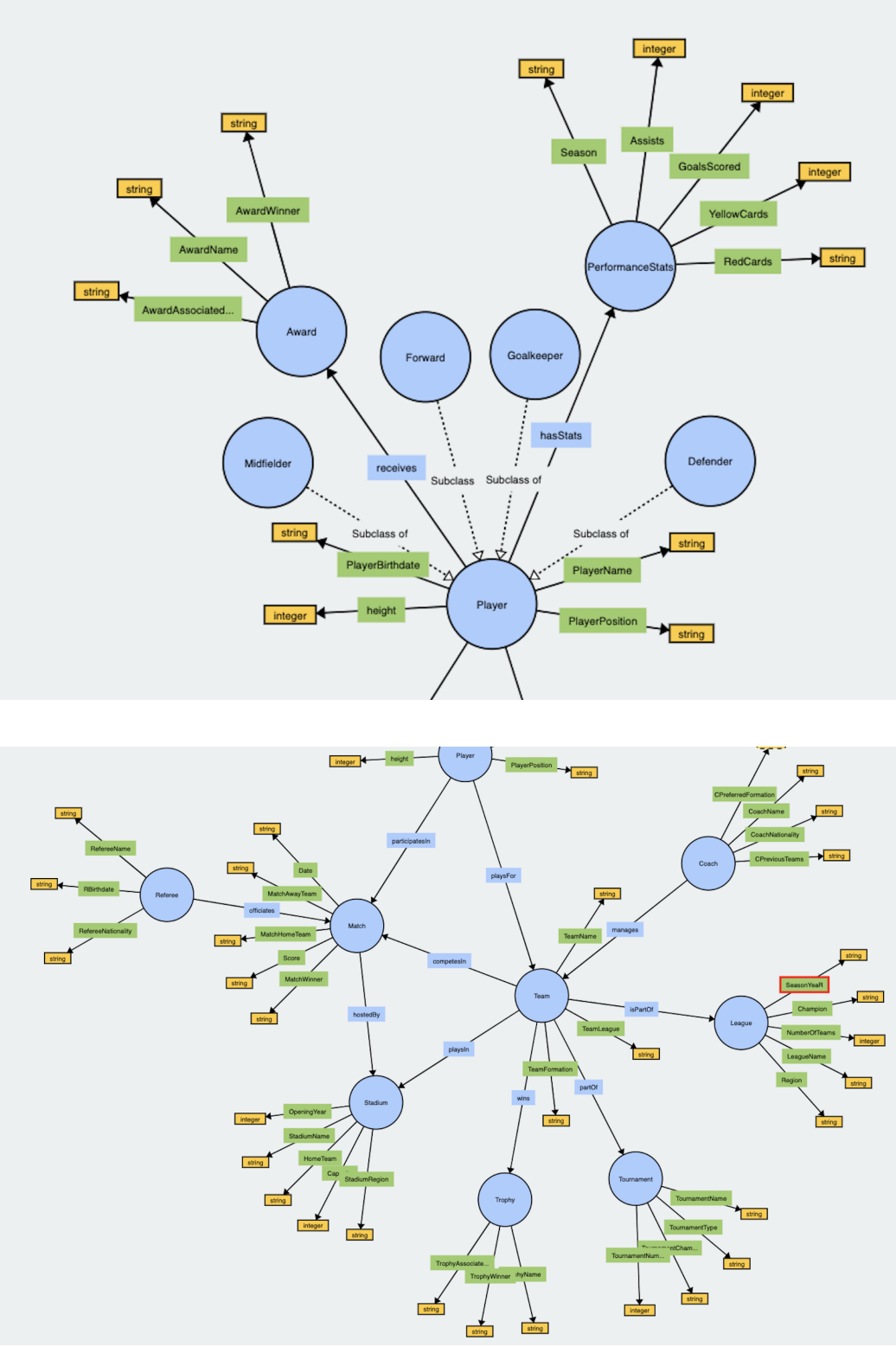

Figure 3: Ontology Version 1

to "Match" and "Performance Stats" to "Match" are important in order to track the performance statistics of each player in a specific match.

The "Team" class plays another significant role in this ontology, including associations with classes such as "Match", "Tournament", "Trophies", reflecting the activities of each team. Furthermore, "Tournament" contains of two sub classes, "Knock Out Tournament", which is tournament type similar to the Champions League, which goes from group stages, to knockout stages up to a final game, and "League", which represents a typical league type such as Premiere League. Also classes such as "Stadium" and "Coach" are linked with the "Team" class, showing which stadium a team is playing in and the name of coach managing a team.

4.3 Ontology Reuse

In this chapter, the focus is on leveraging existing ontologies within the football domain to enhance the efficiency of ontology development. During the conceptualization phase, an approach was used to explore available ontologies that are similar to the football domain, which included a research of various resources such as DBpedia, Wikidata, GitHub, and Linked Open Vocabularies. This approach was also done in the beginning of this project, which was explained earlier in Section 3.1.2 "Data Exchange Identification".

The research has shown that there are few public football related ontologies, which could be used as a reference point for this project. While these knowledge resources offer a lot of information and numerous classes related to football, the absence of comprehensive football-related ontologies becomes clear. This highlights a significant gap in existing resources and emphasize the necessity for the development of a football ontology.

An important consideration in developing the ontology is to reuse the semweb-football ontology created by "sLeeNguyen", which is described in Section 2.3. The semweb-football ontology provided a valuable foundation such as the classes "Player", "Team", or "Player Position". However, this ontology lacked coverage on broader topics like match events and player statistics. Although the possibility of implementing direct subclass or subproperty relationships between the two ontologies was explored, these were not implemented due to structural differences and the extended scope of the new ontology.

Overall, the chapter "Ontology Reuse" highlights the importance of leveraging existing assets to enrich the development of a football ontology. By joining and reusing components from existing ontologies, the project aims to build on established knowledge and frameworks while taking into account specific requirements and goals in the football sector.

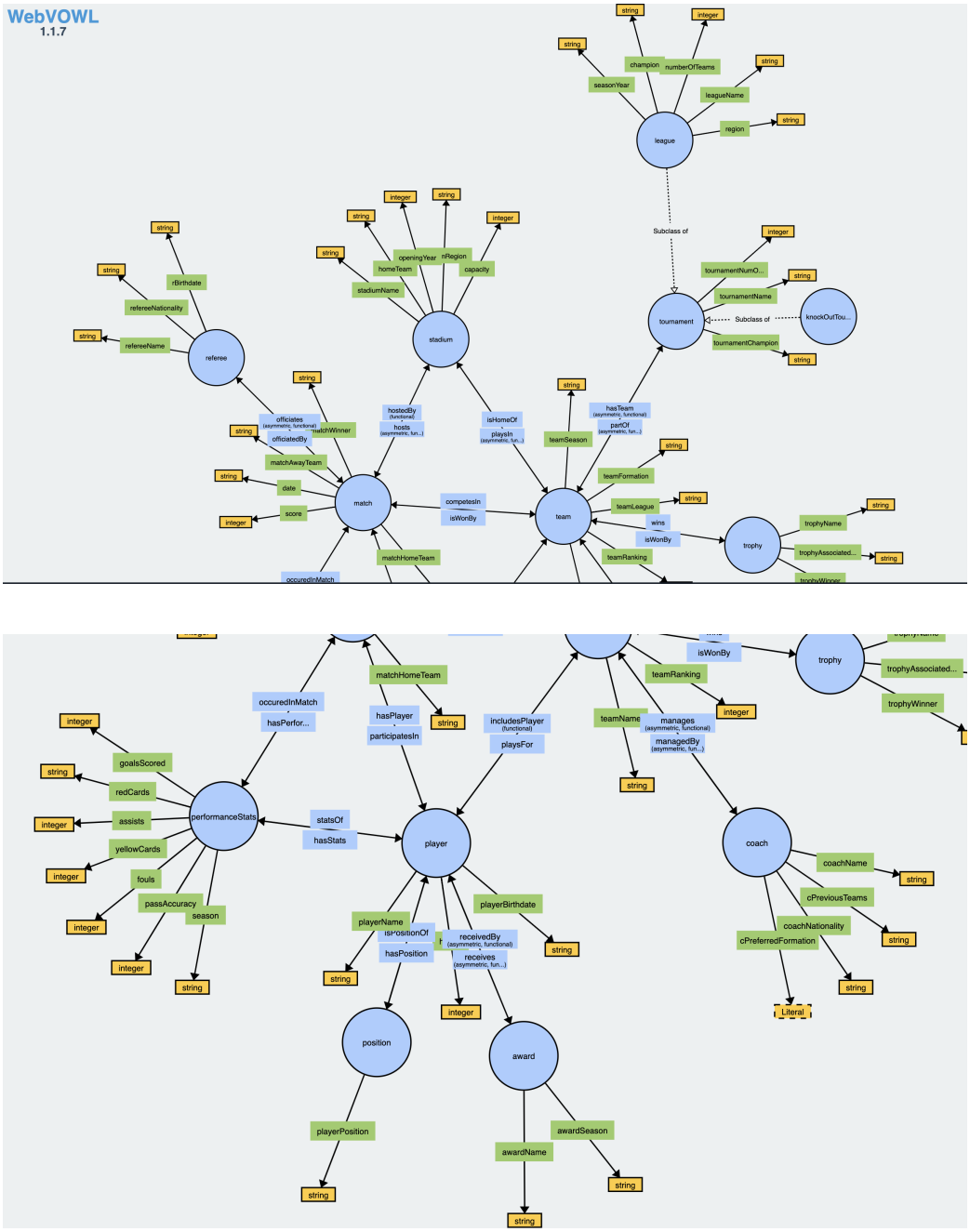

Figure 4: Ontology Final Version

4.4 Ontology Encoding

In Section 3.2.1 "Ontology Conceptualization", the process of ontology encoding was integrated and have been done simultaneously. During this process, the fundamental elements of the football ontology were not only designed, but also encoded into a machine-readable format using OWL technology. This approach ensured that conceptualization and encoding phase were executed together, improving the efficiency in ontology development.

By integrating ontology encoding within the conceptualization phase, the timeline of the ontology's development process was optimized, allowing concurrent adaptions and refinement of concepts and machine-readable representation. The encoding process was done with the help of WebVOWL, which was useful in both, conceptualizing and encoding, since it is a tool for visual representation of ontologies. The ontology created in WebVOWL can then easily be downloaded as a Turtle (.ttl) file.

In general, the use of WebVOWL for ontology encoding within the ontology conceptualization phase ensured higher efficiency and iterative refinement, which is done much easier with the help of WebVOWL than sketching and refining ontologies on paper.

4.5 Ontology Publication

The publication of the ontology is essential to ensure that it is accessible to the community for reuse and collaboration. The ontology was uploaded on GitHub[8], offering both a machine-readable OWL format and human-readable documentation. With the publication of this ontology, other developers and researchers are able to explore, use, or extend the ontology for various football-related applications

4.5.1 Propose Release Candidate

As a preliminary step before public release, the ontology is going to be presented to a university colleague, who also visited the course "Applications of Semantic AI in Knowledge Management", which means that he is aware of the basics and fundamentals of ontologies. The ontology was shown and demonstrated to him. This peer review served as a good validation point, because it gave an external perspective of the ontology. He provided positive feedback on design and functionality of the ontology, meaning that no additional improvement was necessary according to him.

[8] `https://github.com/arditb1997/footology`

4.5.2 Documentation

A detailed documentation of the ontology is highly important for transparency, maintenance, and better understanding by other users and developers. In order to achieve this, an HTML page was created via WIDOCO (**WI**zard for **DOC**umenting **O**ntologies).

WIDOCO is a helpful tool for simplifying the documentation process of ontologies. The tool automatically creates an HTML page based on user input, representing the ontology, structure, classes, properties, and relationships in a clear web-format [6]. The HTML document contains metadata such as latest version, author, or the license. Furthermore, it shows a short introduction to the ontology, followed by a list of classes and namespaces that were declared in the ontology. The ontology itself is also visualized with WEBVowl integrated into the HTML page. A short description of the ontology is also included and eventually information about each of the classes, such as domain, range, or description of the class. Figure 5 displays a screenshot from the WIDOCO HTML documentation page, detailing the classes with their respective IRIs, brief descriptions, and the domains and ranges of each class. The WIDOCO page can be accessed on GitHub[9].

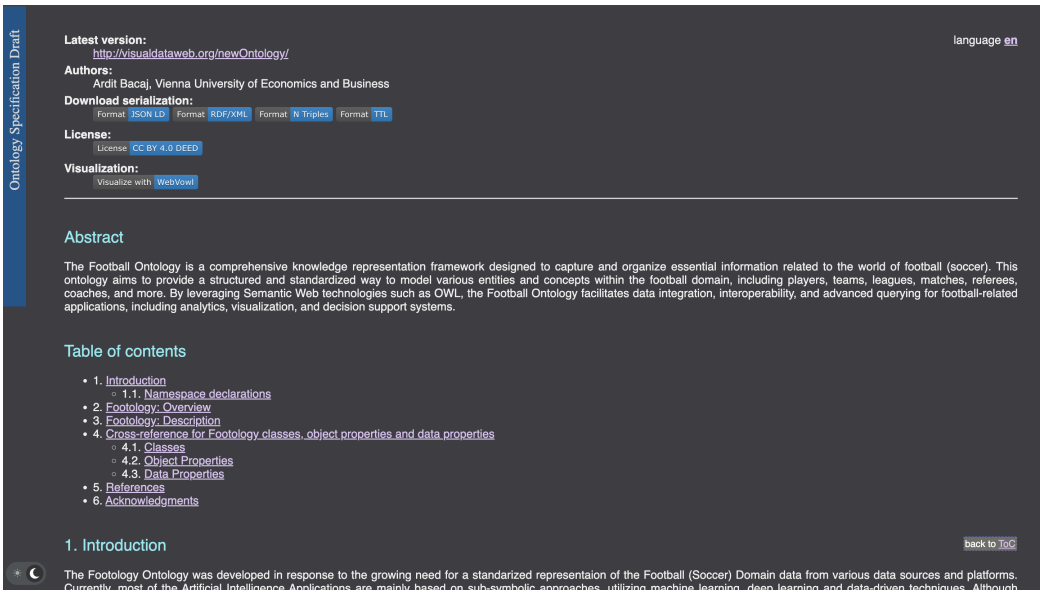

Figure 5: Screenshot of HTML page created using WIDOCO

[9] `https://arditb1997.github.io/widoco-footology/`

4.6 Bug Detection and New Requirements

The ontology will be continuously monitored for bug detection. It is essential to identify issues and remove these issues to maintain the ontology functional. However, after completion and submission of the bachelor thesis, a bug detection process will not be actively pursued but the focus will be on rigorous testing during the ontology development phase, the aim of which is to minimize potential errors. Although the domain of football is constantly changing in terms of player statistics, team strategies, and more, the continuous update of the ontology will be not done by me but rather by community of users or future developers interested in this ontology, as it is accessible on GitHub [10].

5 Ontology Evaluation

Evaluating the ontology is a critical process in ontology development in order to ensure the quality, correctness, and suitability of the created ontology for the intended purpose. It plays a major role in ensuring that the ontology effectively captures and represents knowledge about the football domain while simultaneously meeting the requirements and goals.

In the process of developing a football related ontology, evaluation becomes even more important due to the complexity of the domain. To evaluate the ontology correctly, three different approaches were carried out.

1. Evaluation using Test Data
2. Evaluation using OOPS! Pitfall Scanner
3. Evaluation from Domain Experts

5.1 Evaluation using Test Data

To make sure that the ontology can answer the competency questions correctly, there is only one way testing this, creating test data for each individual class of the ontology and mapping the test data with the ontology using Ontotext Refine and GraphDB. Ontotext Refine is a practical tool that is used for data transformation and linking or mapping [13]. GraphDB is a powerful graph database engine that supports graph data management. GraphDB, also developed by Ontotext, enables efficient storage, organization, and management of large data sets using RDF (Resource Description

[10] `https://github.com/arditb1997/footology`

Framework), SPARQL (SPARQL Protocol and RDF Query Language) and OWL (Web Ontology Language).

The first step in the test data evaluation approach, was to create test data using Microsoft Office Excel. The first two classes "Player" and "Team" were taken as test classes for mapping purposes using Ontotext Refine. The test data was obtained from the website "Transfermarkt"[11], which provides a lot of information about football players, teams, and so on.

PlayerID	PlayerName	height	PlayerBirthdate	PlayerPosition	TeamID	AwardID	StatsID
1	Cristiano Ronaldo	188cm	05.02.85	Striker	7		1
2	Lionel Messi	170cm	24.06.87	Right Winger	8	1	2
3	Jude Bellingham	186cm	29.06.03	Midfield	2		3
4	Kai Havertz	193cm	11.06.99	Midfield	5		4
5	Manuel Neuer	193cm	27.03.86	Goalkeeper	1	5	5
6	Thomas Mueller	185cm	13.09.89	Striker	1		
7	Robert Lewandowski	185cm	21.08.88	Striker	3	3	
8	Jorginho	178cm	20.12.91	Midfield	5		
9	Erling Haaland	195cm	21.07.00	Striker	4	4	
10	Kylian Mbappe	178cm	20.12.98	Striker	6		

Figure 6: Player Dataset

Figure 6 illustrates an example of the "Player" class, which shows various attributes about the Player ID, Player Name, Height, Player Birthday, Position, Team ID, Award ID, and Stats ID. This excel file contains 10 random players whose information were extracted from Transermarkt.at. The Columns Team ID, Award ID, and Stats ID are important information that are needed in order to link players to teams, awards and performance stats. For instance, consider the player Erling Haaland, whose Team ID is listed as number 4. This Team ID corresponds to Manchester City, as indicated in the "Team" dataset Excel sheet where the Team ID for Manchester City is indeed 4. This cross-referencing mechanism ensures accurate and coherent data representation within the ontology.

The latest version of the ontology was uploaded via GraphDB as Figure 7 shows. GraphDB allows to inspect and search through the uploaded ontology. The below Figure represents the class hierarchy, offering an interface for exploring the different components, including classes, instances, and their relationships. This is essential for verifying structure of the ontology after updates or modifications. The class hierarchy view in GraphDB also easily enables users to access detailed information on each class

The next important phase is the mapping of the created data sets. As mentioned earlier, the mapping process will be done with the help of Ontotext Refine. The first step was to start Ontotext Refine and create a new project.

[11] `https://www.transfermarkt.at/`

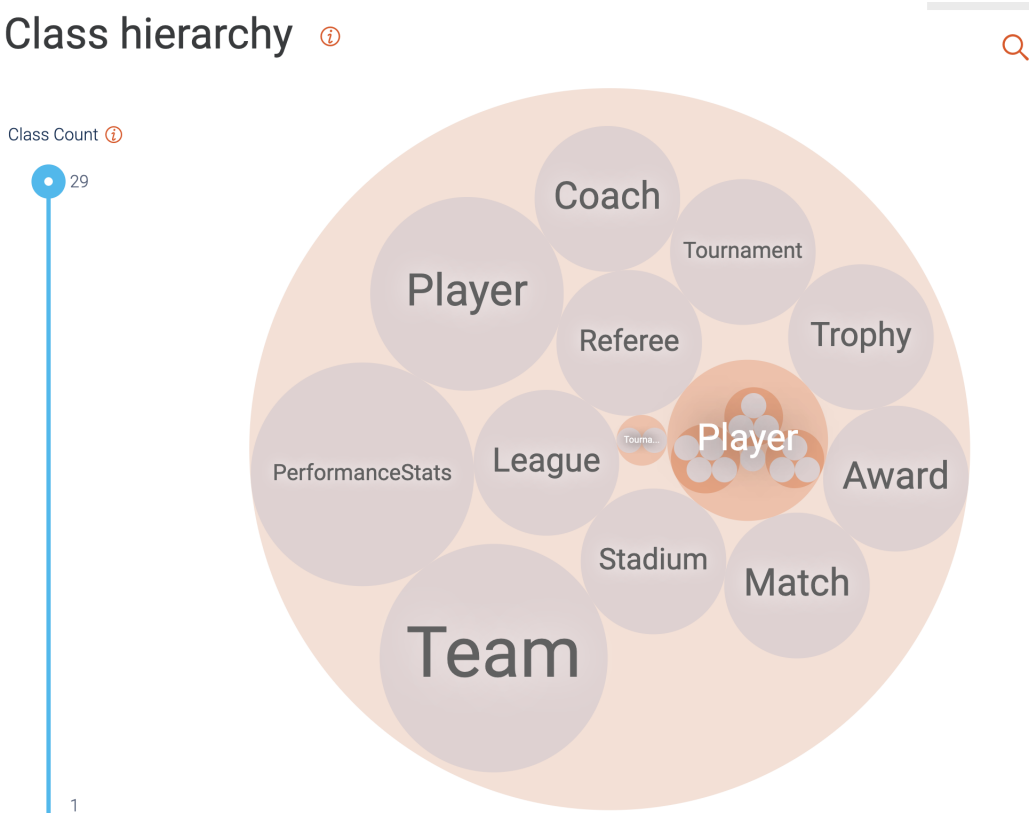

Figure 7: GraphDB Ontology

Then the Excel sheet (data set) was uploaded and the RDF mapping process was started.

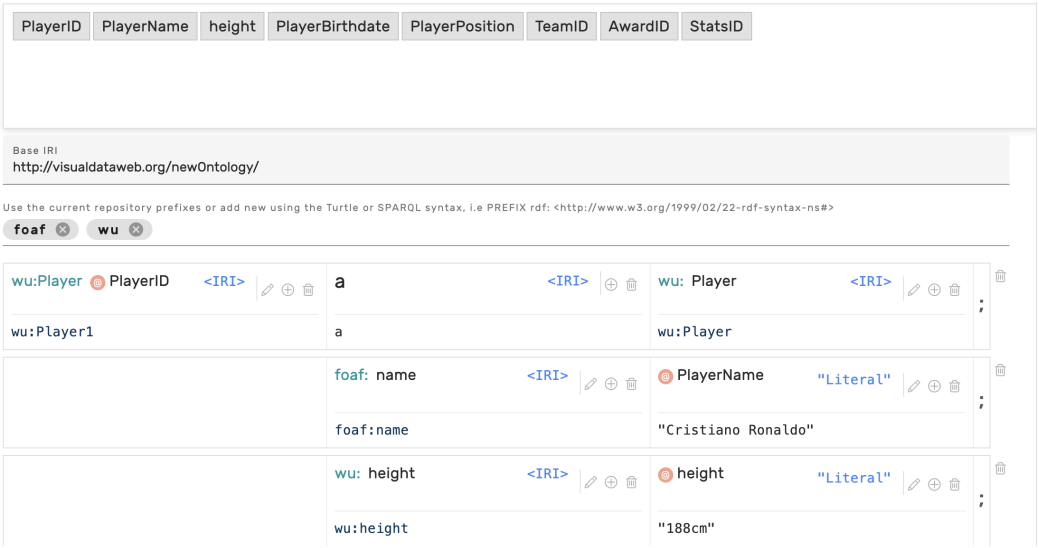

Figure 8: RDF Mapping

The process pictured in Figure 8 showcases the mapping of football player data into a structured RDF format using Ontotext Refine. In this example, the class "Player" was used and the mapping defines the key attributes associated with a player such as Name, Height, Birthday, Player Position (Position ID), Team ID, Award ID, and Stats ID, each linked to the corresponding ontology properties.

As a starting point, each player is identified as an instance of "wu:Player", with "wu" being a prefix for the Base IRI. The Base IRI (Internationalized Resource Identifier) in the context of ontologies is an important component used for constructing full IRIs (Internationalized Resource Identifiers) for elements within a dataset or ontology. The Base IRI is like a root address or namespace from which all other IRIs are determined [17]. This is essential in RDF to establish the type of data each entity represents. The name of the player is mapped using "foaf:name" property, which is included in FOAF (Friend of a Friend) vocabulary widely used in Semantic Web to describe persons and their relations to other persons and objects. The height of a player is mapped using "wu:height" which links each player with the height attribute. This pattern is used for each attribute or datatype property within this class. Once the mapping is completed, it can be saved as a Turtle (.ttl) file and uploaded to GraphDB, which will link the instances to the ontology.

This process was done with every class of the ontology, ensuring a complete ontology including instances that can now be queried through.

After the whole mapping process is completed, a series of SPARQL queries were employed to verify each competency question from the domain requirements described in Section 3.1.6 "ORSD Formalization". These queries are crucial for evaluating the correctness and accuracy of the ontology. In the following, a set of competency questions and their corresponding SPARQL queries are shown as an example:

1. CQ1: What team is Player X currently playing for?

```
SELECT ?playerName ?teamName
WHERE {
  ?player wu:playsFor ?team .
  ?player foaf:name ?playerName .
  ?team wu:TeamName ?teamName .
  FILTER(?playerName = "Erling Haaland") .
}
```

   "?player wu:playsFor ?team ." creates a relationship where the variable "?player" is linked to a "?team" through the "wu:playsFor" property. This indicates which team the player is currently part of. "?player foaf:name ?playerName ." assigns the player's name to the variable "?playerName" using the FOAF vocabulary. "?team wu:TeamName ?teamName ." links the team entity to its name, assigning it to the variable "?teamName" through the property "wu:TeamName". The FILTER clause is used to only show a specific player's information, in this case Erling Haaland. The result is "Manchester City".

2. CQ2: How many goals did Player X score in the season 22/23?

```
SELECT ?playerName ?goals ?season
WHERE {
  ?player foaf:name ?playerName .
  ?player wu:GoalsScored ?goals .
  ?player wu:Season ?season
  FILTER(?playerName = "Jude Bellingham") .
  FILTER(?season = "22/23" )
}
```

   "?player foaf:name ?playerName ." assigns the player's name to the variable "?playerName" using the FOAF vocabulary. "?player wu:GoalsScored ?goals ." retrieves the number of goals associated with the player, binding this number to the variable "?goals". "?player wu:Season ?season" connects the player to the specific season, linking the season identifier to the variable "?season". The FILTER clause again only shows a specific player and a specific season. The result is 23.

3. CQ4: What League ranking does Team X have in Season 23/24?

```
SELECT ?teamName ?ranking ?season
WHERE {
    ?team wu:TeamName ?teamName .
    ?team wu:TeamRanking ?ranking .
    ?team wu:TeamSeason ?season .
    FILTER(?teamName IN ("FC Bayern Munich")) .
}
```

"?team wu:TeamName ?teamName ." This line identifies the team name. "?team wu:TeamRanking ?ranking .". This associates the team identified by "?team" with its ranking, which is stored in the variable "?ranking". "?team wu:TeamSeason ?season .". This links the team to the specific season during which the ranking is applicable, with this season stored in "?season". The result is 2.

The SPARQL queries described above are just a few examples that have been formulated to demonstrate the capabilities of the ontology. These questions are crucial for evaluating the effectiveness of the ontology in retrieving specific information, as it ensures that all elements are correctly integrated and functional. The results of the SPARQL queries underlines a successful implementation of the ontology. These SPARQL queries demonstrate how the ontology addresses the competency questions outlined earlier in 4.1.3. More competency questions and their corresponding queries can be found on the Footology GitHub Repository[12].

5.2 Evaluation using OOPS! Pitfall Scanner

Evaluation in general is a crucial step for ensuring quality, coherence, and usability. Ontology developers must handle difficulties when designing ontologies that may contain anomalies or errors. Therefore it is highly important to evaluate ontologies for error detection. In this case, the Online Ontology Pitfall Scanner (OOPS!) serves as a helpful tool for identifying potential pitfalls within ontologies.

For this evaluation process the OOPS! tool was utilized for the football ontology. There are two ways to upload the ontology on OOPS! for error detection. The first option is to enter a URI of the ontology and the second option is to directly copy and paste the RDF code into the input field. The football ontology was converted into a RDF format and the RDF code was eventually uploaded into the OOPS! tool.

[12] https://github.com/arditb1997/footology

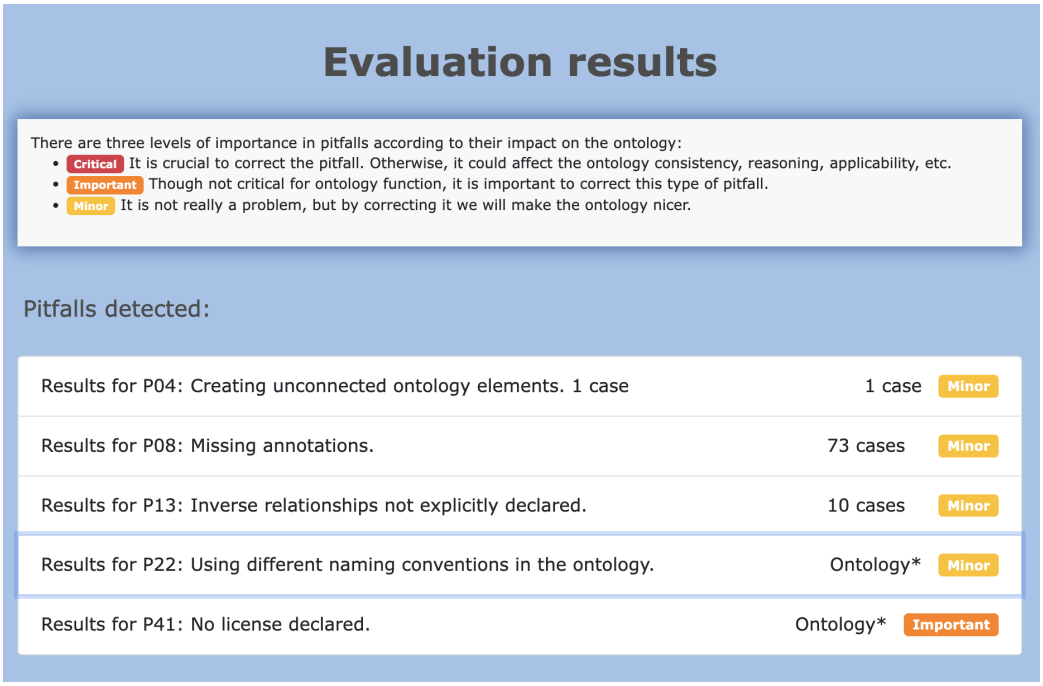

Figure 9: OOPS! Pitfall Scanner Results

The Pitfall Scanner illustrated in Figure 9 shows a couple of minor errors and one important error. The first minor error found in one case is an unconnected ontology element was found. One class was created isolated with no relation to the rest of the ontology. The second error was the lack of annotations in classes, datatype properties and so on. This ensures human readability of each class, datatype, etc. Another pitfall that OOPS! detected was the lack of inverse relationships. This pitfall appears when any relationship does not have an inverse relationship (owl:inverseOf) defined within the ontology. Another problem found by OOPS! is the naming convention in the ontology. The ontology elements are not named following the same convention like CamelCase or use of delimiters such as "-" or "_". All of the previous mentioned errors were minor, but there is one error highlighted as important, which is the lack of license, which was added later using Protege.

During the evaluation process using OOPS!, the errors and pitfalls were handled and fixed with another powerful tool called Protege. Protege is a well-known and widely-used ontology editor. It provides a user-friendly interface for creating, editing, and visualizing ontologies [11]. With protege, all detected pitfalls were able to be fixed, as OOPS! specifically pointed out the exact error class or property, ensuring that the ontology complies to the standards of ontology development.

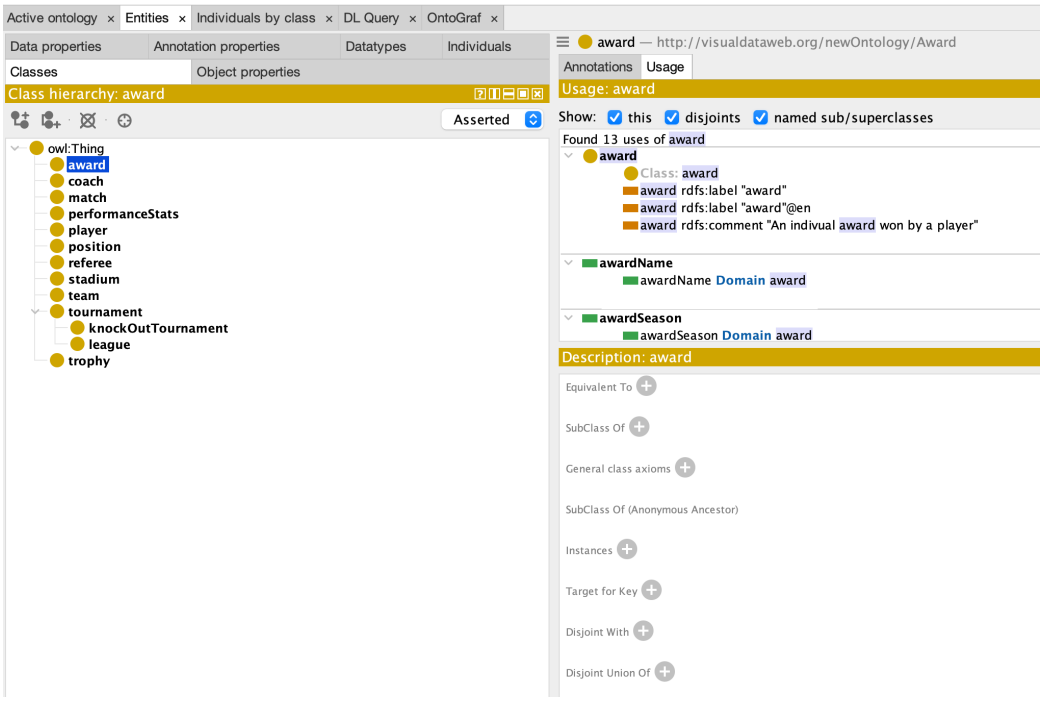

Figure 10: Ontology opened with Protege

The above Figure 10 shows the football ontology opened with protege. In this case, the "Classes" tab is opened, showing all classes of this ontology. Protege follows a hierarchical organization, typically presented in a tree-like format. On the right side there are several options available, such as annotations like label, comment, etc. and other possibilities such as adding sub-classes or instances.

The evaluation using OOPS! Pitfall Scanner was a very important step for ensuring a correct and functional ontology without missing annotations or any other pitfalls that can occur during ontology development, with the help of protege, the detected errors were fixed.

5.3 Evaluation from Domain Experts

In the final step of the evaluation phase, an interview was held with two domain experts, who were also interviewed in the beginning of this bachelor thesis in Section 4.1.3 in order to assess the completeness, correctness, and usability of the developed ontology.

During the interview, the domain expert suggested to include an additional class "Position", which would be a more structured representation of player position, instead of directly linking the positions to the players as it was before. The "Position" class would include all the sub-classes such as forward, defender, or midfielder, providing a more organized representation of player positions within the ontology.

This evaluation with feedback provided by the domain experts helped to identify potential improvements for the ontology, ensuring that it fits to the needs and expectations of its end users in the football domain.

6 Discussion

The discussion of the ontology involves reflecting the outcomes of the development and how it is compared with other existing ontologies. A deeper understanding of the challenges faced was gained, especially in stakeholder engagement, ontology structuring, and data integration. This Section will focus on achievements and the limitations observed during the project, as well as potential approaches for future improvement.

6.1 Insights gained from the Ontology Development

Developing the ontology using the LOT methodology was a challenging project, where several key insights were gained. One important learning was the importance of stakeholder engagement, particularly in defining functional requirements through competency questions. The feedback of domain experts helped refine the ontology scope and ensured its relevance to real-world scenarios. Another interesting insight was integrating the conceptualization and encoding phases. With utilizing tools such as WebVOWL, real-time adjustments and iterative refinements were made, resulting in well-structured ontology. One of the challenges were during the data exchange identification and encoding phases, such as the scarcity of comprehensive football ontologies, underlined the need of creating a dedicated and detailed ontology in this domain.

6.2 Comparison with Existing Ontologies

To evaluate the effectiveness of Footology, a comparison with other existing ontologies in the sports domain was conducted. In Table 4, the similarities and differences between Footology and other existing ontologies are shown. The two ontologies, "Semweb-football" and "Sport-Ontology", from Section 2.3 were used as a comparison ontology.

Feature	**Footology**	**Semweb-Football**	**Sport-Ontology**
Main Domain	Football-specific	Football-specific	General Sports
Scope	Comprehensive, covering players, teams, matches, and detailed performance stats.	Focuses on football but primarily on players and teams without deep integration of match data or stats.	Limited to general sports events and teams with minimal focus on football-specific details
Detail Level	High, with detailed player stats, match events, and performance statistics.	Medium, with a focus on player attributes but less on match-specific data.	Medium, general information about sports events, teams, and matches.
Key Classes	Player, Team, Match, PerformanceStats	Player, NationalTeam, League, LeagueSeason	SportsEvent, SportsTeam, Match, Tournament
Object Properties	Rich set of properties including hasPlayer, hasPerformanceStats, participatesIn, and others tailored to football.	Properties like hasPosition, isChampionOf, hasTeam, focused but lacks thorough coverage in match-specific contexts.	Basic properties like hasHomeTeam, hasAwayTeam, hasParticipants, which are more generic.

Table 4: Comparison of Ontologies

Footology offers a more comprehensive and detailed representation of football compared to the other two ontologies. While "Semweb-Football" focuses mainly on players and teams with less match-specific data, and "Sport-Ontology" focuses more on general sports aspects, Footology covers all key aspects of football, including detailed player stats, match events, and performance metrics. This makes Footology more complete and better suited for football-specific applications.

6.3 Conclusion and Future Work

This bachelor thesis handled the development and evaluation of a football related ontology, applying the Linked Open Terms (LOT) methodology. The thesis began with a literature review, identifying gaps in existing football ontologies and resources underlining the need for a structured representation of football related knowledge. The first step of the LOT methodology discussed the requirement specification, leading to a completed ORSD document, which

served as a foundation for the implementation and development process. The development process contained the conceptualization and evaluation phases to ensure it met the functional requirements derived from competency questions and expert feedback. The publication of the ontology on GitHub and its documentation via WIDOCO ensures that it remains accessible to other developers and researchers.

As for future work, no further updates are planned, but the ontology will be accessible on GitHub for the community to expand the scope, refining its detail level, or integrating it with other relevant information. By uploading the Footology online, it serves as a resource for researchers, other developers, or enthusiasts who want to build upon this ontology or use it as a reference in their own projects.

References

[1] Adel AL-Zebari, Shahab Zebari, and Karwan Jacksi. *Football Ontology Construction using Oriented Programming.* Journal of Applied Science and Technology Trends, 2020.

[2] Oscar Corcho, Mariano Fernandez-Lopez, Asuncion Gomez-Perez, and Angel Lopez-Cima. Building legal ontologies with methontology and webode.

[3] Tom Decroos, Lotte Bransen, Jan Van Haaren, and Jesse Davis. *Actions Speak Louder than Goals: Valuing Player Actions in Soccer.* Association for Computing Machinery, 2023.

[4] draw.io. About draw.io. `https://www.drawio.com/about`, Accessed: 2024-03-25.

[5] Sport for Business. The worldâ€™s most watched sports, 2024. Accessed: 2024-09-24.

[6] Daniel Garijo. Widoco: a wizard for documenting ontologies. In *International Semantic Web Conference*, pages 94–102. Springer, Cham, 2017.

[7] Eleni Ilkou and Maria Koutraki. *Symbolic Vs Sub-symbolic AI Methods: Friends or Enemies?* 2020.

[8] Rathi Keshav, Koul Aditya V., Somani Priyam, and Dr. Manu K. S. *Applications of Artificial Intelligence in the Game of Football: The Global Perspective.* ERM Publications, 2020.

[9] Robin Keskisarkka, Huanyu Li, Sijin Cheng, Niklas Carlsson, and Patrick Lambrix. *An Ontology for Ice Hockey.*

[10] Steffen Lohmann1, Vincent Link, Eduard Marbach, and Stefan Negru. *WebVOWL: Web-based Visualization of Ontologies.* Knowledge Engineering and Knowledge Management, 2015.

[11] M.A. Musen. The protege project: A look back and a look forward. *AI Matters*, 1(4), June 2015. Association of Computing Machinery Specific Interest Group in Artificial Intelligence.

[12] Natalya F. Noy and Deborah L. McGuinness. Ontology development 101: A guide to creating your first ontology.

[13] Ontotext AD. Ontotext refine. Accessed: 2024-04-30.

[14] Maria Poveda-Villalon, Alba Fernandez-Izquierdo, Mariano Fernandez-Lopez, and Raul Garcia-Castro. *LOT: An industrial oriented ontology engineering framework.* Engineering Applications of Artificial Intelligence, 2022.

[15] María Poveda-Villalón, Asunción Gómez-Pérez, and Mari Carmen Suárez-Figueroa. OOPS! (OntOlogy Pitfall Scanner!): An On-line Tool for Ontology Evaluation. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 10(2):7–34, 2014.

[16] Aditya Sangram Singh Rana. *Event Detection in Football using Graph Convolutional Networks.* arXiv e-prints ADS Bibcode: 2023arXiv230110052S, 2023.

[17] W3C. Rdf 1.1 concepts and abstract syntax, 2014. Accessed: 2024-05-02.

[18] Akira Yamanaka, Hiroshi Otsuka, Tatsuya Deguchi, Ken Okihara, Dota Otsuka, Shinya Uchida, Jorge Diaz-Cidoncha Garcia, Selina Khoo, David Stroud, and Masahiro Sugiyama. *Tactical Analysis Through Objective Data in Football.* Insight - Sports Science, 2020.