Data-Driven Analytics for Decision Making in Game Sports

Inauguraldissertation

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Philipp German Seidenschwarz

Basel, 2021

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Heiko Schuldt
Erstbetreuer

Prof. Dr. Malte Helmert
Zweitbetreuer

Prof. Dr. Arnold Baca
Externer Experte

Basel, den 21. September 2021

Prof. Dr. Marcel Mayor
Dekan

If my mind can conceive it, and my heart can believe it – then I can achieve it.

— Muhammad Ali

Zusammenfassung

Wettkampfdiagnostik und eine gute Entscheidungsfindung im Sport sind wichtig, um die Gewinnchancen zu maximieren. In den letzten Jahren ist die Menge und die Qualität der Daten, die für die Analyse zur Verfügung stehen, durch technische Fortschritte, wie z.B. von Sensortechnologien oder Computer Vision Technologie enorm gestiegen. Die datengetriebene Analyse von Sportler- und Mannschaftsleistungen ist jedoch sehr anspruchsvoll. Ein Grund hierfür ist die sogenannte semantische Lücke der Sportanalytik. Das bedeutet, dass die Konzepte der Trainer nur selten in den Daten repräsentiert sind. Darüber hinaus stellt der Sport im Allgemeinen und der Spielsport im Besonderen aufgrund der dynamischen Eigenschaften und der multifaktoriellen Einflüsse auf die Leistung eines Sportlers, wie z.B. die zahlreichen Interaktionsprozesse während eines Spiels, eine große Herausforderung dar. Dies erfordert verschiedene Arten von Analysen, wie z.B. qualitative Analysen und damit anekdotische Beschreibungen der Leistungen bis hin zu quantitativen Analysen, mit denen Leistungen mittels Statistiken und Indikatoren beschrieben werden können. Hinzu kommt, dass Trainer und Analysten unter einem enormen Zeitdruck arbeiten und Entscheidungen sehr schnell getroffen werden müssen.

Um die anspruchsvolle Aufgabe von Spielanalysten und Trainern zu erleichtern, stellen wir einen generischen Ansatz vor wie ein Data Analytics System (DAS) zur effizienten Unterstützung der Entscheidungsprozesse in der Praxis konzipiert und gestaltet werden kann. Wir stellen zunächst ein theoretisches Modell vor und zeigen einen Weg, wie man die semantische Lücke der Sportanalytik überbrücken kann. Dadurch wird sichergestellt, dass DASs relevante Informationen für die Entscheidungsträger liefern. Außerdem zeigen wir, dass DASs sowohl qualitative und quantitative Analysen als auch Visualisierungen kombinieren müssen. Zusätzlich stellen wir verschiedene Abfragetypen vor, die für ein ganzheitliches Retrieval von Sportdaten erforderlich sind. Des Weiteren zeigen wir ein Modell für die nutzerzentrierte Planung und Gestaltung der User Experience (UX) eines DAS.

Nach der Vorstellung der theoretischen Grundlagen präsentieren wir SportSense, ein DAS zur Unterstützung der Entscheidungsfindung im Spielsport. Seine generische Architektur erlaubt eine schnelle Anpassung an die individuellen Eigenschaften und Anforderungen verschiedener Spielsportarten. SportSense ist neuartig in Bezug auf die Tatsache, dass es Rohdaten, Ereignisdaten und Videodaten in einem System kombiniert. Darüber hinaus unterstützt es verschiedene Abfragetypen einschließlich eines intuitiven skizzenbasierten Retrievals und kombiniert nahtlos qualitative und quantitative Analysen sowie verschiedene Datenvisualisierungsoptionen. Des Weiteren stellen wir die beiden Anwendungen SportSense Football und SportSense Ice Hockey vor, die sportartspezifische Konzepte enthalten und (umfassende) Taktikanalysen abdecken.

Abstract

Performance analysis and good decision making in sports is important to maximize chances of winning. Over the last years the amount and quality of data which is available for the analysis has increased enormously due to technical developments like, e.g., of sensor technologies or computer vision technology. However, the data-driven analysis of athletes and team performances is very demanding. One reason is the so called semantic gap of sports analytics. This means that the concepts of coaches are seldomly represented in the data for the analysis. Furthermore, sports in general and game sports in particular present a huge challenge due to its dynamic characteristics and the multi-factorial influences on an athlete's performance like, e.g., the numerous interaction processes during a match. This requires different types of analyses like, e.g., qualitative analyses and thus anecdotal descriptions of performances up to quantitative analyses with which performances can be described through statistics and indicators. Additionally, coaches and analysts have to work under an enormous time pressure and decisions have to be made very quickly.

In order to facilitate the demanding task of game sports analysts and coaches we present a generic approach how to conceptualize and design a Data Analytics System (DAS) for an efficient support of the decision making processes in practice. We first introduce a theoretical model and present a way how to bridge the semantic gap of sports analytics. This ensures that DASs will provide relevant information for the decision makers. Moreover, we show that DASs need to combine qualitative and quantitative analyses as well as visualizations. Additionally, we introduce different query types which are required for a holistic retrieval of sports data. We furthermore show a model for the user-centered planning and designing of the User Experience (UX) of a DAS.

Having introduced the theoretical basis we present SportSense, a DAS to support decision making in game sports. Its generic architecture allows a fast adaptation to the individual characteristics and requirements of different game sports. SportSense is novel with respect to the fact that it unites raw data, event data, and video data. Furthermore, it supports different query types including an intuitive sketch-based retrieval and seamlessly combines qualitative and quantitative analyses as well as several data visualization options. Moreover, we present the two applications SportSense Football and SportSense Ice Hockey which contain sport-specific concepts and cover (high-level) tactical analyses.

Acknowledgements

This thesis is the result of four challenging and exciting years during which I received support of so many people. I would like to take the chance to express my special thanks to some particular people.

At first, I would like to thank my advisor Prof. Dr. Heiko Schuldt, not just for giving me the opportunity to do my PhD, but also for his endless support and valuable feedback he gave me during this time.

I would like to also thank Prof. Dr. Arnold Baca from the University of Vienna for his willingness, time, and effort to review my thesis.

A special thanks goes to my colleagues from the StreamTeam project. First, I would like to thank my mentor Martin Rumo. I really appreciated the fruitful discussions we had on many exciting topics and which finally also contributed to this thesis. Second, my thanks also go to my colleague Lukas Probst for his unprecedented patience and support during the last years.

I would like to thank all my colleagues from the Databases and Information Systems research group from the University of Basel and in particular Alexander Stiemer for his willingness to help and his technical support.

Another big thanks goes to Prof. Dr. Marcel Jacomet from the Bern University of Applied Sciences and to PD Dr. Silvio Lorenzetti from the Swiss Federal Institute of Sport Magglingen for providing me with the best working conditions during the last four years.

Finally, I would like to thank my family for their endless support during my whole studies without which none of this would have been possible.

This work has been partly supported by the Hasler Foundation in the context of the project StreamTeam, contract no. 16074, which is also thankfully acknowledged.

The video material which is part of the screenshots in Figure 4.7 and Figure 4.9 was provided courtesy of the Swiss Football Federation (SFV). The American football playing field used in Figure 4.4(a) is designed by Freepik: http://www.freepik.com. The basketball court used in Figures 4.4(b) and 4.5(a) is designed by jemastock/Freepik: http://www.freepik.com.

Contents

Figures

Tables

Definitions

Examples

Acronyms

AJAX	Asynchronous JavaScript and XML
DAS	Data Analytics System
DotA 2	Defense of the Ancients 2
EPV	Expected Possession Value
GPS	Global Positioning System
HTTP	Hypertext Transfer Protocol
IMU	Inertial Measurement Unit
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
REST	Representational State Transfer
UEFA	Union of European Football Associations, French: Union des Associations Européennes de Football
UI	User Interface
UX	User Experience
XML	Extensible Markup Language

Symbols

δ	Time range (delta)
aid	Agent identifier
AID	Set of agent identifiers
cid	Competition or match identifier
CID	Set of competition or match identifiers
cs	Continuous state
$csid$	Continuous state identifier
e	Event data item
$edid$	Event data item identifier
$EDID$	Set of event data item identifiers
i	Indicator
ID	Set of identifiers
imd	IMU data item
p	Phase
pd	Physiological data item
pid	Phase identifier
pr	Profile
$prid$	Profile identifier
q_{sp}	Query specification
Q_{sp}	Set of query specifications
r	Raw data item
R	Set of raw data items
$rdid$	Raw data item identifier
$RDID$	Set of raw data item identifiers
s	Statistic
S	Set of statistics
sid	Sensor identifier
SID	Set of sensor identifiers
td	Spatio-temporal tracking data item
tid	Team identifier
TID	Set of team identifiers
ts	Timestamp
TS	Set of timestamps

Part I

Introduction and Background

1 Introduction

How can you not be romantic about baseball?

— Billy Beane

Sport plays an important role in today's society. Among other reasons, sport encourages important values (e.g., fairplay and respect [Ghi15]), promotes good health [KDJ+01], integrates people, and is undoubtedly a very important economic factor [Kla08]. As far as the latter is concerned, this becomes more evident when we take a look at numbers in the following examples. The size of the global sports analytics market had a value of around 885.0 million U.S. dollars in 2020. Furthermore, this number is expected to increase at a compound annual growth rate of 21.3% from 2021 to 2028 to reach as much as 3,442.6 million U.S. dollars in 2028 [Res21]. Another example which demonstrates the economic power of sports is the money distribution in the UEFA[1] Champions League competition. Participating teams get 2.7 million euro for each win during the group stage and 900,000 euro for each draw [Goa20]. This puts coaches and players under considerable pressure to win their matches. Therefore, good decision making especially before and during a match is essential, as this will improve their chances of winning.

Decision making is an elementary process not only in sports but in all areas of life. In a military context like war, for instance, parties have to decide whether to attack or retreat. Fire fighters have to think about how to save the maximum number of people inside a burning house. Investment bankers have to decide how they should invest their clients' money. So there are various examples of decision making in different contexts with all having in common that it is desirable to make decisions that will lead to the most positive outcome. Taking the wrong decision in a war can lead to severe injuries or even death. The same holds for the fire fighter scenario. If bankers make the wrong decision, it might

[1] Union of European Football Associations, French: Union des Associations Européennes de Football (UEFA)

not endanger lives but it will have an effect on their clients' assets. However, making right decisions is not always easy and straightforward. It is a complex process and, depending on the context, influenced by a myriad of factors. At this point data-driven analytics could consitute a possible solution. Nowadays, huge amounts of data are generated in many areas of life. However, making use of the data and deriving the right decisions is very challenging. That is why in many areas Data Analytics Systems (DASs) are applied.

DASs are software tools which are used to gather and analyze data, e.g., from a company. The goal of using a DAS is to gain insights from the data which are not obvious for the user, because of the huge amount, the heterogeneity, and/or perhaps even the velocity of the produced data. Users can analyze and visualize the data with a DAS. The resulting reports from these analyses will then provide the users with all the necessary information to facilitate their decision making process. Such systems are widely applied in business and organizational settings. Nevertheless, they could also be useful when applied in other contexts where decision making is particularly complex like, for instance, in sports. Currently, this is not the case or at least not on a comparable level. The reason for that is simple. In the business setting most of the Key Performance Indicators (KPIs)[2] are already adequately defined. This allows for the relatively straightforward development and application of a DAS. Sports, on the other hand, has not yet reached that level of professionalism. This is because competitive sport can be a chaotic and dynamic environment [Rob20] and it is very difficult to represent athletes' and teams' performances adequately in statistics or indicators in many sports.

The development of DASs to support decision making in sports would be interesting from at least two different points of view. First, a lot of (tactical) decisions have to be made by athletes, analysts, and coaches. Because sports is very dynamic, these decisions mostly need to be made in a very short amount of time. The following examples illustrate this point clearly. In Tour de France stages, to have the best chance of winning or doing well, it is important how the athlete and his team position themselves within the peloton. In Formula 1, the tyre choice before and during races can decide between victory or a final rank outside the points. Ice hockey coaches have to decide quickly which block should play against the opponent's one to have the highest chances of scoring a goal. For all of these decisions there is not much time available. Therefore, a system supporting the coach or the analyst with insights gained from data

[2] KPIs can be defined as "quantifiable measures used to evaluate the success of an organization, team, employee, or athlete, in meeting objectives for performance" [HKB+21].

would definitely be helpful. Second, sports are dynamic systems which are very attractive from an analytical and technical point of view. Big amounts of data from different sources and of different formats are generated and gathered and therefore build a challenging opportunity for accessing, retrieving, and visualizing information with DASs. That is why we will focus on the development of DASs for sports in this thesis.

In the remainder of this chapter, we will briefly present milestones in the area of sports analytics (see Section 1.1), describe the challenges of performance analysts in sports practice (see Section 1.2), list the contributions we make in this thesis for data-driven analytics for decision making in game sports (see Section 1.3), and outline the content of this thesis (see Section 1.4).

1.1 Milestones of Sports Analytics

There are two major milestones in the area of sports analytics which we want to highlight in this section. First, the beginnings of performance analysis in football with Charles Reep and second, the spectacular story of Billy Beane in baseball analytics. Additionally, we will highlight the problem of the probably impossible transferability of the "Moneyball" approach to other sports.

Before we take a look at these two famous examples in the area of sports analytics it is important to first define what the term "sports analytics" actually means. Thus, we introduce the following definition of Link [Lin18] who describes sports analytics as

> "the process of searching, interpreting, and processing information in sports-related performance data using information systems and mathematical methods of data evaluation with the aim of achieving competitive advantages."

Having clarified the terminology we can now continue with the introduction of two milestones in the field of sports analytics. Charles Reep is seen as the pioneer of performance analysis in football [Pol02]. The first match he analyzed dates back to 1950 in the United Kingdom. His work was mainly based on a pencil and a notebook, where he manually recorded all the on-the-ball actions of a football match. This type of analysis is called "notational analysis". For that he developed his own system based on notes and symbols to represent the whole course of the match. Reep made first analyses for the team of Wolverhampton Wanderers regularly when the team began to dominate English football in the

1950's [Pol02]. The team applied a more direct style of play which was the result of Reep's mathematical analyses. He found out that chances of scoring a goal are higher when less passes are played. Over the years Charles Reep analyzed more than 2,000 matches [Mar19; Pol02] and published several scientific articles on the results of his analyses.

Baseball provides us with another highlight in the area of sports analytics. In 2002 Oakland Athletics manager Billy Beane achieved something extraordinary with his team. Making use of so called "sabermetrics"[3] he and his team managed to reach the playoffs in 2002 and 2003. This is extraordinary because the team had one of the smallest budgets in the MLB[4] at that time. To offset the financial disadvantage the Oakland Athletics had to find a different solution to become competitive. They used statistical analyses and found new indicators for the offensive performance of players. With that approach they found players who were undervalued by the market and also by other teams and could sign them for much less money. With Michael Lewis' book "Moneyball", telling the story of Billy Beane and the Oakland Athletics, the sports world finally became aware of the potential lying in data-driven analytics.

Although it seems very attractive, it is not easy or maybe even impossible to apply the "Moneyball" approach to other sports with more complex (tactical) patterns. The reason for that is simple. Baseball has a clear structure, a strict segmentation of the game, and less player interactions compared to other game sports like, for instance, football or basketball (see Section 3.2.1). Thus, the influence and the contributions of an individual player to the overall team performance cannot be clearly represented by simple statistics in more complex sports. Consequently, data-driven analyses are mainly limited to the physical level with a strong focus on monitoring the internal load of athletes, whereas tactical analyses and analyses of players' contributions mostly remain video-based. In contrast to that, less complex sports like, for instance, baseball or American football, follow more data-driven approaches also for the tactical analysis. However, this requires clear definitions of the sport-specific events, moves, and patterns. This is the case in American football. Therefore, teams can use so called playbooks, where their strategies for the offensive plays as well as the defensive plays are worked out and listed in detail. After a match or during the breaks, the performance can be analyzed and evaluated based on the statistics.

[3] Or SABRmetrics, where SABR stands for the Society for American Baseball Research, founded in 1971. The term "sabermetrics" itself is commonly known and is generally used to describe any mathematical or statistical study of baseball [SAB20].

[4] Major League Baseball (MLB)

Furthermore, tactics can be adapted if necessary. Pat Kirwan writes in his book "Take Your Eye Off The Ball" [KS10]:

> "A coach's master playbook can contain about 1,000 plays – pretty much anything he would ever consider calling in a game. Every bomb, blitz, and blocking scheme is in there somewhere, along with every gadget play and goal-line scenario. And every call has its roots somewhere in that all-encompassing bible, which every coach is forever adding to and carrying with him from job to job."

In this respect other sports definitely need to catch up. Having clear definitions of sport-specific events, moves, and patterns would facilitate the development of indicators or even KPIs representing athletes' or teams' performances which in turn could be integrated into a DAS for the sport in question.

1.2 Challenges of the Performance Analyst

In sports practice, it is mostly the performance analyst who works with the data and assists the coach with the decision making process. This job is very demanding and these analysts face a lot of different challenges. In this section, we describe four of these challenges: (1) the time pressure, (2) the ever-changing information needs, (3) the need for high-level tactical analyses in many sports, and (4) the provision of qualitative and quantitative analyses.

According to the definition of sports from the GAISF[5] a sport should include an element of competition. Competition in turn is always linked to time pressure. Imagine a handball club which plays in one of Europe's top leagues and which competes in three different competitions: the national league, the national cup, and the EHF[6] Champions League. This means that in some phases of the season several matches per week have to be played. This is demanding not only for the athletes, but also the coaching staff and the performance analyst are exposed to a very tough program. Among other things, the analyst has to prepare the team for the next opponents by analyzing their strengths and weaknesses. As there is not much time available between matches, working efficiently is essential for the analyst.

Another challenge performance analysts have to face are the ever-changing information needs. Each coach has his/her own concepts and understanding of

[5] Global Association of International Sports Federations (GAISF)
[6] European Handball Federation (EHF)

the sport (i.e., a mental model) and thus specific individual information needs. If the head coach is fired and replaced, the analyst has to quickly adapt to the new coach and to the new mental model. The analyst's flexibility is required in such situations. Additionally, there are also situations in which other members of the coaching staff need the services of the analyst, for instance, if the assistant coach needs some specific insights about the opponent's key player, or if the physiotherapist wants to watch different angles of the scene during which the athlete got injured. Again, the performance analyst should then be able to work very flexibly.

Depending on the sport, a tactical analysis might be very complex. When we take football as a specific example, there is not only the need for simple tactical analyses like, for instance, where the most shots on goal are created by a team, but also for high-level tactical analyses. This includes highly complex patterns like, for instance, the analysis of pressing behaviour. Therefore, analysts should also be able to deal with and to satisfy coaches' needs for high-level tactical analyses.

Coaches are primarily interested in the video. It is "the most important source of information"[7]. That is because the whole context is maintained in the video. With a qualitative video analysis coaches further have the chance to reflect on a match after the emotions have passed [GC04]. On the other hand coaches would like their impressions to be confirmed with statistics, either to better argument with their athletes or with other club or association responsibles like, for instance, the sporting director or the general manager. So there is a need for both: first, a qualitative analysis focusing on the video and the whole context of specific situations like events or patterns of events and second, a quantitative analysis focusing on aggregated statistics based on certain events or other performance parameters. Both types of analyses need to be covered by the performance analysts work.

The development and application of DASs which can handle all the previously mentioned challenges would definitely support analysts and facilitate their work.

1.3 Contributions

We have seen that decision making in the context of sports is very complex. Additionally, coaches and analysts are facing lots of challenges and are exposed to

[7] Anonymous UEFA Pro License coach

a considerable pressure in their daily work. In order to address these challenges, we present a generic approach for the conception and design of DASs to support decision making in the complex context of sports with a particular focus on game sports. More precisely, this thesis makes the following contributions:

- We present an approach how to bridge the semantic gap of sports analytics. This is essential to develop DASs which can provide useful analysis results for coaches and analysts because their concepts will be represented in the data. For this, we show a way how to get access to domain knowledge, how to extract sport-specific concepts, and how to translate the concepts into a performance model and finally into a data model. We demonstrate our approach in the context of football.

- We introduce a model to describe the complexity of game sports and explain why more complex sports require a combination of three different analysis options compared to simpler sports. We furthermore show that all sports can profit from DASs which combine the different analysis options.

- We present different query types which DASs need to support for a holistic retrieval of sports data. Additionally, we show how sports data should be modeled to allow for a performant retrieval process.

- We present a generic model for User Experience (UX) design of DASs for game sports in order to allow coaches and analysts an intuitive and flexible information retrieval. This allows the users a very time-efficient working and therefore optimizes the decision support in sports practice.

- We present an implementation of the DAS SportSense. SportSense unites raw data, event data and video data in one system. Furthermore, the system supports four different query types and enables coaches and analysts to work efficiently and flexibly. The system covers high-level tactical analyses, combines both qualitative and quantitative analyses, as well as several data visualization options. With that we show that SportSense covers all challenges which coaches and analysts are exposed to in their daily work.

- We present two applications of the DAS SportSense for game sports analysis, namely SportSense Football and SportSense Ice Hockey. With that we show that SportSense can easily be adapted to other game sports due to its generic architecture.

– We present the results of user studies on the SportSense Football system which we conducted with twelve football analysts and coaches.

– We present the results of a quantitative performance evaluation of SportSense Football where we show the retrieval performance and the scalability of the system.

1.4 Structure of Thesis

This thesis consists of four main parts. The first part is thought to motivate the reader with a short Introduction (see Chapter 1) and to give some background information for a better understanding of the rest of the thesis. Concerning the latter Chapter 2 provides the reader with some domain-specific insights from two different perspectives: a sport-scientific and a computer-scientific perspective.

The second part constitutes the core of this thesis and consists of two chapters. In Chapter 3 we introduce different approaches to conceptualize and design DASs to support decision making in game sports. For this, we present how to bridge the semantic gap of sports analytics. This will allow DASs to provide users with information which really is relevant in sports practice. Furthermore, we show that DASs need to combine different analysis options and need to provide different query types for a holistic information retrieval. Moreover, we present a generic model for the UX design of DASs for game sports to optimize the working with such a system. Chapter 4 describes the implementation of the DAS SportSense for decision making in game sports, its generic architecture, and two sample use cases for sports practice, namely SportSense Football and SportSense Ice Hockey.

The third part of this thesis consists of two chapters to discuss the main contributions. In Chapter 5 the SportSense Football application is evaluated from two perspectives. User studies are conducted with domain experts from football (i.e., coaches and analysts) to check qualitatively whether the system is usable in football practice. Additionally, the performance of SportSense Football is evaluated quantitatively through measurements of the retrieval performance and the scalability of the system. Related work is discussed in Chapter 6. Here, we illustrate approaches from research and also present industrial solutions.

The last part summarizes this thesis and shares thoughts on improvements as well as on ideas for future work in Chapter 7.

2 Foundations

Information is a source of learning. But unless it is organized, processed, and available to the right people in a format for decision making, it is a burden, not a benefit.

— William Pollard

Due to the interdisciplinary characteristics of this thesis we now provide background information from two different perspectives: first, from a sport science perspective (see Section 2.1) and second, from a computer science perspective (see Section 2.2). This is important because both domains bring relevant expertise in relation to conducting and interpreting tactical analyses [GMB+21]. The sport-scientific part describes in detail the challenges and requirements which come with the performance analysis in sports and especially in game sports. Furthermore, we introduce a formalization of the semantic gap of sports analytics. The computer-scientific part depicts the whole process and challenges of sports analytics from the data generation, to data integration, data analysis, and finally up to the data visualization and user interaction.

2.1 The Sport Science Perspective

In this section we illustrate that the decision making process in sports practice is very complex and that the development of DASs to support that process is highly required. This section further emphasizes why we focus especially on game sports in this thesis. The following pages cover three important points. First, we show that sports not equals sports, but that there exist many different types of sports with different complexity levels and challenges concerning the (tactical) performance analysis of athletes and teams (see Section 2.1.1). Second, we shed light on the special requirements of coaches and analysts for the analysis process in game sports (see Section 2.1.2). Third, we present the semantic gap of sports analytics, which describes the problem that concepts of coaches are often not represented in the data for the analysis (see Section 2.1.3).

2.1.1 Types of Sports

There are various types of sports with all having different rules, characteristics, and demands. When analyzing athletes performances it becomes clear that each type is exposed to different challenges. In this section we introduce a categorization of sports and a subsequent subcategorization of game sports. Additionally, we present the challenges when analyzing the performance of athletes or teams and further clarify the need for data-driven analytics for decision making in game sports.

The program for the Olympic Games 2020 in Tokyo comprises 33 different sports with even more disciplines [Com20b]. They reach from A like Archery to W like Wrestling [Gam20]. The Olympic Winter Games 2022 in Beijing include seven sports, again with different disciplines, reaching from B like Biathlon to S like Snowboard [Com20a]. However, these sports picture only a very small part of the sports which are practised all over the world. SportsPodium, for instance, claims that there exist around 8,000 different sports worldwide [Spo17]. Such a large number makes it difficult to categorize sports. Nevertheless, there are some approaches in literature which try exactly that. In Table 2.1 we see a categorization of sports according to Engelhardt and Neumann [EN94], which we slightly modified by adding eSports as an additional category. Engelhard and Neumann differentiate five categories: strength-based sports, endurance-based sports, technics-based sports, duel sports, and game sports. One can discuss whether eSports is an own additional category of sports or if it would be better to classify eSports as technics-based sports because of the requirement of a technical device to play the games. When going further into detail one might classify even single eSports games into different categories as well like, e.g., FIFA 21 into game sports and Fortnite into duel sports. Others might argue that eSports is not a type of sports at all. In any case this is a topic which is still heavily discussed in sport science. Even the five categories of Engelhardt and Neumann might not allow for a clear classification of all different types of

Table 2.1 Different types of sports with examples based on Engelhard & Neumann [EN94].

Strength-based	Endurance-based	Technics-based	Duel Sports	Game Sports	eSports
Weightlifting	Marathon	Figure Skating	Boxing	Basketball	FIFA 21
100m Sprint	Rowing	Gymnastics	Judo	Tennis	DotA 2
Ski Jumping	Biathlon	Motor Sports	Fencing	Cricket	Fortnite

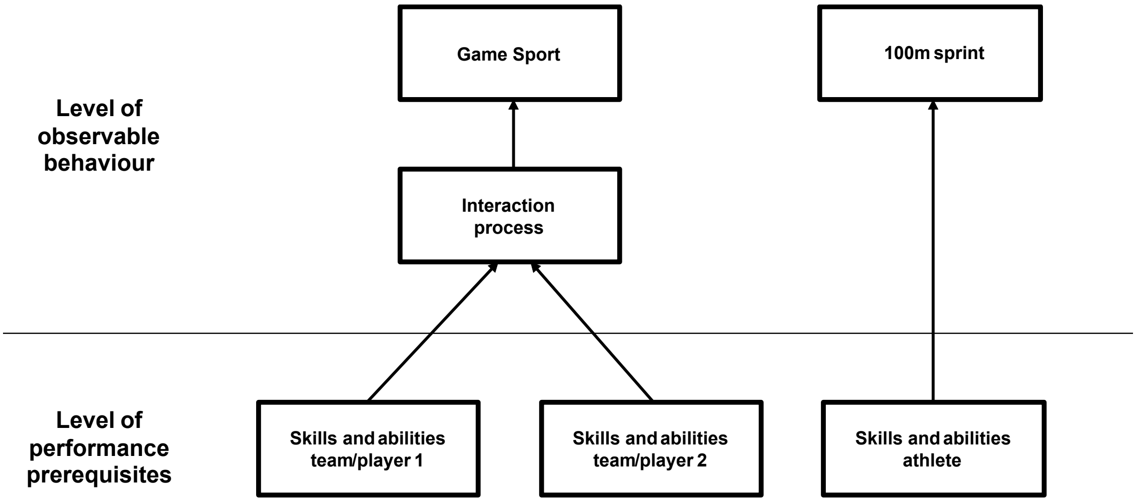


Figure 2.1 Differences between game sports and other sports according to Lames & McGarry [LM07].

sports as each sport can contain characteristics of one or several other categories as well. However, this classification is a good basis for further discussions and modelings.

When we now think about analyzing performances of athletes in different sports it soon becomes clear that there are differences between each individual sport, but also between whole categories. When comparing, for instance, a 100 meter sprint race to a basketball match it is quite obvious that the evaluation of individual performances is much more difficult in basketball than in the sprint. The reason for this is simple. In sprinting, it is the athlete's personal skills and abilities leading to the performance in the competition. In basketball, on the other side, the performance of an athlete is not only skill- and ability-related. Game sports like basketball involve complex player interactions [MZT+18] and therefore, the performance of an athlete is always influenced by the team mates as well as by the opponent's players. This agrees with the statement of Carmichael et al. [CTW00]:

> "In many sports, the isolation and identification of individual contributions is problematic, if not impossible, due to the continuously interactive nature of a match or game; between members of the same team and between them and their opponents."

It is therefore the interaction which makes the difference. This is also depicted by Lames and McGarry [LM07] (see Figure 2.1). In addition, interactions change during a match because players make adjustments [BKL05]. Thus, we can state that the analysis of athletes' performances in sports with lot of interac-

Table 2.2 Subcategories of game sports with examples based on Mitchell et al. [MOG13].

Target Games	Striking/Fielding Games	Net/Wall Games	Invasion Games
Golf	Baseball	Badminton	Basketball
Bowling	Softball	Tennis	Football
Billiards	Cricket	Volleyball	Ice Hockey
Croquet	Rounders	Table Tennis	Handball
Curling	Kickball	Squash	Rugby
		Racquetball	American Football

tion like, e.g., game sports is more challenging than the analysis in sports with less interaction like, e.g., strength-based sports. This is exactly what makes game sports particularly interesting for sports analytics and is a good reason to take a deeper look at this category in the following.

Mitchell et al. [MOG13] introduce subcategories of game sports based on the work of Almond [Alm86]. These subcategories together with a small excerpt of (slightly modified) examples are depicted in Table 2.2. We can see a differentiation between target games, striking and fielding games, net and wall games, and invasion games. The number of interactions per match is different when we compare sports of the four categories. Interactions get more numerous from target games, to striking and fielding games, to net and wall games, and finally to invasion games, where the number of interactions per match is highest. The element of interaction also has an influence on the tactical analysis of game sports as Decroos et al. [DVD18] confirm:

> “[...] many factors such as interactions among multiple different players across space and time, and features of the game state (e.g., score, field position, time left, and team quality) influence decision making and tactics.”

Interaction is not the only factor which complicates the analysis of an athlete’s or a team’s (tactical) performance. Multi-factorial influences make a detailed description of the complexity levels and the respective differences of the four categories of game sports very challenging.

To summarize this section we can state that the task of analysts and coaches to analyze athletes’ and teams’ performances is very demanding. Game sports and particularly invasion games present a huge challenge due to multi-factorial influences on an athlete’s performance and underline the importance of developing DASs to support decision making processes.

2.1.2 Types of Analyses

Having shown in the previous section that game sports are very challenging when it comes to the (tactical) performance analysis of athletes and teams we now highlight that the analysis process in sports practice requires the following three options: (1) qualitative analysis, (2) quantitative analysis, and (3) visualization.

As already introduced in Section 1.2 for most coaches and analysts the video is a very important source of information. The more complex relevant sport-specific patterns (e.g., events, plays) get, i.e., when the patterns are influenced by more factors, the less often they will occur in the course of a match and the more anecdotal the coaches' descriptions of the player or team performances will be. In these cases, a coach will more likely carry out a qualitative analysis by watching the respective video scenes. For the analysis it is here more important to consider the whole context which is represented in the video, instead of having aggregated statistics over a small sample which in turn do not fully consider and represent context information. In football, for instance, gaining possession in a dangerous area occurs only infrequently, but is very important [DVD18] and thus worth to be analyzed qualitatively. We define a qualitative analysis in Definition 2.1.

Definition 2.1 Qualitative analysis in sports.

A *qualitative analysis in sports* is defined as the subjective analysis of sport-specific events, movements, patterns, or plays, including all context information as they are contained, for instance, in videos.

In contrast, game sports also contain simpler patterns. These will occur significantly more often in the course of a match. For these cases, coaches and analysts will not always have time to analyze all situations qualitatively through watching the video scenes as these are just too numerous. Therefore, there is a higher need for a quantitative analysis, which means to calculate and analyze statistics aggregating the most important information of these patterns. This can be simple statistics, but also indicators or even KPIs which will help coaches to better evaluate and analyze the (tactical) performances of the players and the teams. In these cases it is also more negligible to not have the full context considered and represented for each situation. A defintion of a quantitative analysis is presented in Definition 2.2.

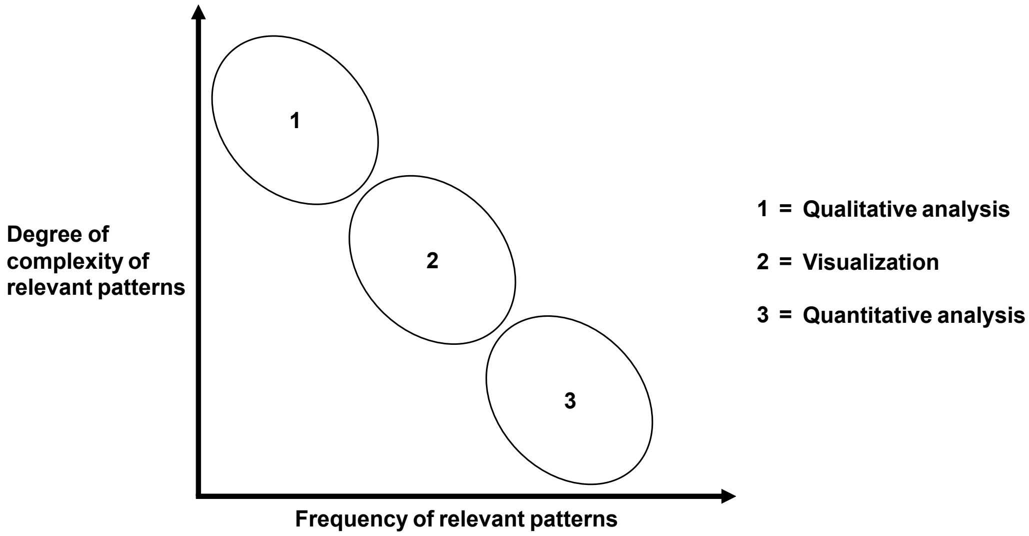


Figure 2.2 Differently preferred options for the analysis process.
Varying frequencies and complexity levels of relevant patterns cause differently preferred options for the analysis process: qualitative analysis, quantitative analysis, and visualization.

Definition 2.2 Quantitative analysis in sports.

A *quantitative analysis in sports* is defined as the objective analysis of sport-specific events, movements, patterns, or plays, in the form of aggregated statistics and indicators in which context information potentially get lost.

In between these two extremes there are patterns which occur frequently but not too often. Here, it makes sense to apply data-based visualizations. These visualizations in turn can accelerate and support the decision making processes as they can make, for instance, trends visible at first glance. Visualizations again can contain more context information compared to quantitative analyses. In the following, we present a definition of visualization (see Definition 2.3).

Definition 2.3 Visualization in sports.

A *visualization in sports* is a quantitative representation of sport-specific events, movements, patterns, or plays, which is analyzed qualitatively.

In Figure 2.2 we illustrate the preferred analysis options in game sports when complexity levels of patterns and their frequencies vary. Of course there are no discrete distinctions when a coach or analyst should use a specific option. This remains an individual decision of the coaches and analysts. These points get clearer with a specific example from football (see Example 2.1).

Example 2.1 Differently preferred options for the analysis process in football.

In football we observe less frequent, but very complex patterns like, for instance, chance creations via the wing players. There are also more frequent, but less complex patterns like, e.g., the pressing intensity in defensive transition phases. Furthermore, there are very frequent and very simple patterns like, e.g., pass sequences within one team. These three examples show the need for different options for the analysis process. The first example would be suitable for a qualitative analysis as coaches and analysts could take a closer look at the behaviour of the offensive players but also on how the defensive team was organized during these situations. This would not be represented adequately with aggregated statistics. For the second example it would be better to have a visualization which shows the intensity of pressing, e.g., with a line graph visualization. The last example of the pass sequences would be perfectly suitable for a quantitative analysis with aggregated statistics showing the passing behaviour of players.

2.1.3 The Semantic Gap of Sports Analytics

Turning now to the last point which we want to present from a sport science perspective, the semantic gap of sports analytics. In this section we clarify this phrase and explain what it actually represents from our understanding. Furthermore, we highlight the importance of bridging this gap. In this context, we also highlight the lack of collaboration between sports practice and research which contributes to a maintenance of the semantic gap of sports analytics.

According to a definition provided by Hein [Hei10], a semantic gap describes

> "the difference in meaning between constructs formed within different representation systems."

In the following, we will replace the term construct with concept to ensure a consistent terminology. A semantic gap so arises from a language problem between two parties which both describe a certain concept. The difference between these descriptions then represents the so called semantic gap. Of course, such gaps exist in various contexts and thus also in sports analytics as depicted in Figure 2.3.

We observe that coaches have their individual mental models of the sport in which they are engaged. A mental model can be described as a coach's coarse vision to reach a level of successful and ideally attractive play. This vision is

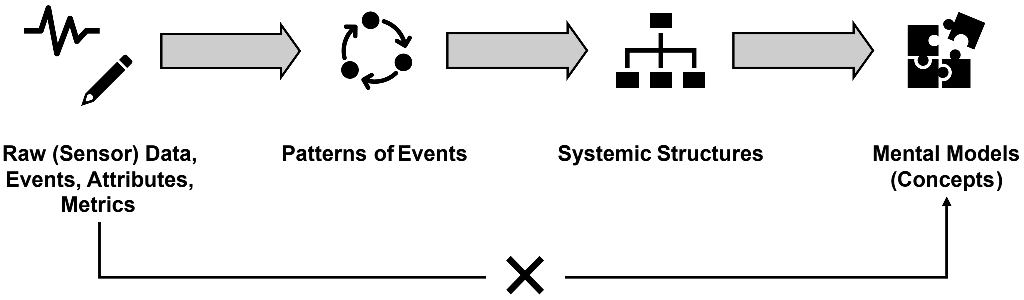


Figure 2.3 Semantic gap of sports analytics.
Concepts of coaches are often not represented in the data.

implemented by the team applying specific concepts of the coach. In a perfect scenario and to support a data-driven decision support using DASs, the coach's concepts would be represented in the data. This would be possible if the individual concepts can be represented as systemic structures which in turn consist of specific patterns of events. Patterns of events are certain sequences of single events (see Figure 2.3). These events can be captured either through manual tagging or automatically with corresponding systems (see Section 2.2.1.2). If the implementation of the previously described points is possible, e.g., with strict definitions, time-efficient analyses would be feasible and coaches could profit from the big amount of data which is by now available in the area of sports analytics (see Section 2.2). Unfortunately, this is mostly not the case and so the analysis often remains manually and video-based, which is a very tedious and time-intensive task. Consequently, the semantic gap, from our view, represents exactly that problem if concepts of coaches are not adequately represented in the available data for the analysis process. For a better understanding of the semantic gap of sports analytics we present Example 2.2.

Example 2.2 Pressing and the semantic gap of sports analytics.
The concept *Pressing* in football is one example revealing the semantic gap of sports analytics.

Football coaches have a certain idea how they want their teams to play in order to be successful and to win matches. This idea (i.e., the mental model) is put into practice with several specific concepts. *Pressing* can be one of these concepts and is often applied by teams in football practice. Depending on the current position of the ball on the football pitch, players should exert pressure on the ball-carrying player during defensive play phases with the goal to regain ball possession. However, pressing includes various factors and to further complicate the situation, these factors are not strictly defined and may even differ considerably between coaches. Therefore, pressing is a concept which will not be represented adequately in sports data.

We can find many other concepts which are also not represented in the data like in the case of pressing. Without having clear definitions for each concept, academia and also industry cannot create adequate analysis approaches and software solutions which would help coaches and analysts in sports practice. The language or jargon which is used by coaches is often not understandable for researchers and vice versa which further complicates the problem. That is why the analysis of many complex concepts still remains very time-intensive. In such cases coaches and analysts have to search for all scenes in the videos where the team applied relevant concepts. Therefore, it would be an enormous profit for coaches and analysts, if the semantic gap of sports analytics could be bridged. Among other things, the bridging of the semantic gap could be achieved through a closer collaboration between academia and sports practice. However, here occurs the next problem as Drust and Green [DG13] describe:

> "[...] the influence of the scientific information that is available has a relatively small influence on the day-to-day activities within the 'real world' of football."

This is confirmed by Nicholls et al. [NJB+19]. The authors state that academic literature and the results play only an insignificant part in sports practice. One reason is certainly the previously mentioned jargon of the parties involved. Coaches and analysts simply do not understand the results of researchers. Additionally, researchers often develop very complex models to measure team or player performances (see Section 6.1.1). A lack of academic background could be hindering for coaches to understand such scientific models [BKT+18]. On top, these models are often not transparent enough for coaches. Thus, the knowledge translation between research and practice should be optimized [San18]. On the other side, sports practitioners struggle in speaking clearly about their concepts and also in formulating them precisely which in turn would enable, e.g., researchers to represent these concepts in data or KPIs for the analysis in sports practice. An additional problem is that knowledge which is produced by scientists working inside a club often is not shared with the community but remains inside the clubs [Buc17]. To create potent DASs for decision making in sports it will be crucial to bridge or at least to narrow the semantic gap of sports analytics.

2.2 The Computer Science Perspective

Over the last years the amount of data which is available for the analysis in sports has increased enormously. This is due to technical developments like, e.g., sensor technologies which have improved noticeably in quality. But also other technologies like, for instance, computer vision showed impressive progresses and now allow, for instance, for a valid athlete and ball tracking. These developments enable analysts, coaches, and scouts to develop and follow more data-driven approaches for their analyses. However, sports data are not only interesting for practitioners but also for data scientists and other researchers. Applying machine learning or data mining approaches makes the detection and development of new KPIs and other interesting metrics possible. This in turn will allow coaches and analysts to draw better conclusions on player contributions to team performance. Additionally, data-based tactical analyses are facilitated. Some research approaches are presented in Section 6.1.

This section describes the whole data analytics pipeline in the context of sports data step-by-step. An example of such a pipeline is depicted in Figure 2.4. In Section 2.2.1 we present the different types of sports data which are generated and required for the analysis process. Afterwards, we present the data integration step together with its challenges in Section 2.2.2. The requirements during the data analysis step are focused in Section 2.2.3. The last step of the pipeline is presented in Section 2.2.4 and focuses on how to contextualize information through visualizations. Additionally, we introduce a layer model for the planning and design of DASs which later makes an intuitive and time-efficient analysis process with the system possible for users.

2.2.1 Data Generation

The commonality of all sports is that bodies or at least parts of bodies are moving in a certain space to achieve a sport-specific goal. The movements of human bodies moreover entail physiological responses. Different technologies allow the measurement of movements, corresponding external loads, as well as eventually accompanying physiological responses (i.e., internal loads). Additionally, important sport-specific moments of a match or competition can be captured. In this section we present three different types of data which play an important role in the area of sports analytics: (1) raw data, (2) event data, and (3) video data. We illustrate the characteristics of each type, how these data can be generated, as well as some use cases.

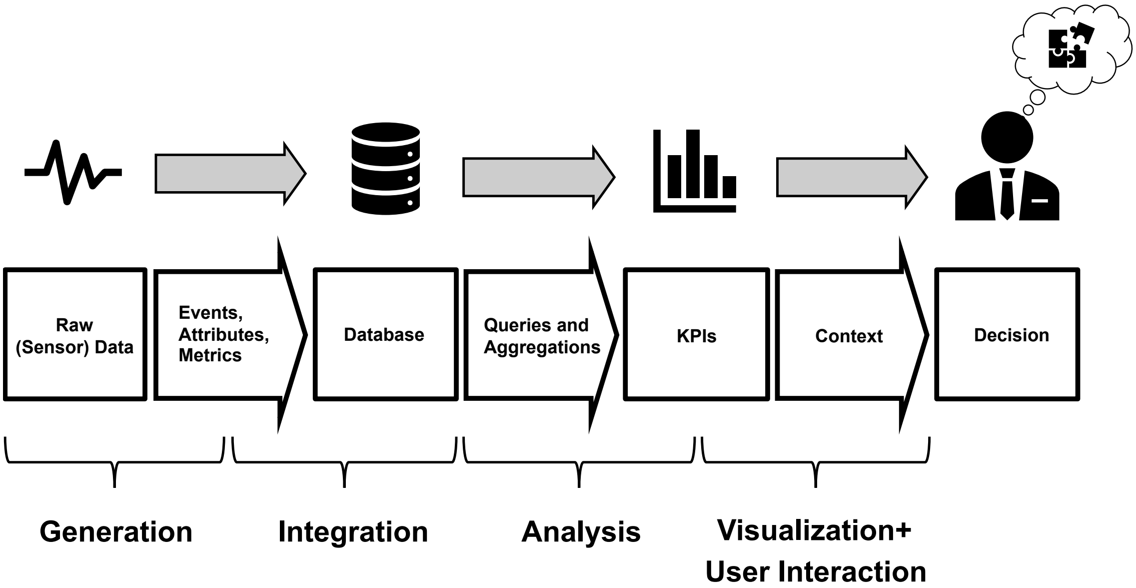


Figure 2.4 Sample data analytics pipeline to support the decision making process of coaches and analysts.

2.2.1.1 Raw Data

Raw data in sports typically occur in the form of physiological data, Inertial Measurement Unit (IMU) data, or spatio-temporal tracking data. While the internal load can be measured through physiological data, the external load is measured through motion capturing via IMU data or spatio-temporal tracking data. We formally describe a single raw data item in sports as an 8-tuple (see Definition 2.4). Raw data types are defined in Definition 2.5.

Definition 2.4 Raw data item.

A *raw data item* r is a tuple $r = \langle rdid, sid, cid, aid, tid, ts, j, \langle a_{1_j}, ..., a_{n_j} \rangle \rangle$ where $rdid$ represents the unique identifier of the raw data item, sid the unique identifier of the corresponding sensor, cid the unique competition or match identifier, aid the unique agent identifier, tid the unique team identifier, ts the timestamp, j the raw data type, and a_{1_j} to a_{n_j} raw data type-specific attributes.

Definition 2.5 Raw data type.

Typically, a raw data item r is of the type physiological data item pd, IMU data item imd, or spatio-temporal tracking data item td, among other options:

$$r.j \in \{pd, imd, td, ...\}$$

Each raw data item r is unique independent from the type or the corresponding sensor. To ensure this, the sensor identifier $r.sid$ can be part of the raw data item identifier $r.rdid$. An agent in the context of sports can be, for instance, an athlete or a player, the ball in basketball or football, or the gearbox in a Formula 1 car. In the following we present the three types of raw data in more detail.

Physiological Data. Physiological data are recorded signals from the human body. This includes, e.g., heart rate, heart rate variability, blood pressure, breath rate, or even brain activity. Heart rate sensors like products from Polar[1] can measure the heart rate and its variability via electrocardiography (ECG) or via optical measurement technologies. Breath rate can also be measured with sensors. Brain activity can be measured via electroencephalography (EEG) technology. Physiological data are mainly used to monitor the internal load of athletes during training sessions or competitions. Additionally, physiological data allow to draw conclusions on athletes' recovery potential and (mental) stress profile. This information can be crucial for the prevention of injuries.

IMU Data. An IMU is a technical device which consists of different sensors: an accelerometer, a gyroscope, and often a magnetometer [vdKR18]. Such devices are widely applied in sports to capture movements of athletes like, for instance, the acceleration, the orientation of the body (or parts of a body), or the orientation of sports equipment like, e.g., rackets [CBF+18]. The devices have to be fitted on the athlete (or the body part) or on the equipment of the athlete for the measurement. Xsens[2] is an example of a company providing such measurement units. Data can then be used for injury prevention through the analysis of, for instance, joint angles during certain movements which an athlete often performs in the respective sport. Another use case is the training and improvement of technical skills of athletes, e.g., by optimizing the angle of the racket during the service in tennis. Acceleration and deceleration data are used to analyze the external load of athletes.

Spatio-Temporal Tracking Data. Spatio-temporal tracking data, known also as positional data, contain information on the position (e.g., of players or playing equipment) in terms of (x, y, z) coordinates over time. Today's technology allows the capturing of positional data of athletes with a reasonable accuracy [BTL+18]. In the following we introduce different technologies to generate this type of raw

[1] https://www.polar.com/us-en/products/heart-rate-sensors
[2] https://www.xsens.com/

data. One option is based on computer vision technology. Cameras and corresponding software are used to track the position of the athletes. ChyronHego's TRACAB Optical Player Tracking[3] system is one example where this technology is applied. The major advantage is that athletes do not have to wear a sensor, which for some sports is anyway not allowed in competitions. One problem coming up with that technology is that in outdoor sports the weather conditions are not always ideal due to fog, rain, or snow which makes it difficult for most software to accurately and validly track positions of athletes or playing equipment like a football. Another problem can occur when athletes are overlapping in the video for a short moment and software wrongly swap the identities of athletes. Sensors can cope with these problems. Two sensor-types need to be further differentiated: radar-based sensors and Global Positioning System (GPS)-based sensors. For radar-based technologies like it is applied by Local Position Measurement[4] or KINEXON[5] there is a need for a certain set-up which consists of a base-station and further reference stations which are installed around the area of interest, e.g., a basketball court. GPS-based technology like in Catapult[6] or Field-Wiz[7] sensors does not need an additional set-up. After the calibration process radar-based systems provide certain advantages. The technology is very accurate and can be applied indoors in contrast to purely GPS-based technology. Both, radar-based and GPS-based technology have the advantage of retaining correct identities of the athletes as each wears an own device. However, this in turn is a disadvantage for both types compared to video-based tracking technology. As already mentioned, the wearing of sensors is not always allowed in a competitive setting or simply not accepted by athletes. Additionally, the ball tracking with sensors is still a problem. It is either not allowed by the rules of a certain sports league, or not supported by the players as a sensor which is fixed inside the ball might have a major influence on the ball's physics. Therefore, the ball tracking is mostly based on computer vision technology. Spatio-temporal tracking data are useful for various reasons. Among other options, these data can be used to analyze the external load of athletes in terms of, e.g., running distances, time spent in different speed zones, or the number of intense sprints. Additionally, accurately tracked positions of player positions also allow the analysis of tactics and patterns of play in sports contests [FLD+11]. Researchers and

[3] `https://chyronhego.com/products/sports-tracking/tracab-optical-tracking/`
[4] `https://inmotio.eu/technology/`
[5] `https://kinexon.com/sports-technology`
[6] `https://www.catapultsports.com/products`
[7] `https://asi.swiss/products/fieldwiz/`

analysts assume an enormous potential hidden in positional data. This potential seems to be not fully utilized so far [GMB+21].

2.2.1.2 Event Data

Another type of sports data are so called event data. In contrast to raw data, event data contain information on important moments in a certain match or competition and thus are sport-specific. Event data cover everything which happens in terms of actions, including temporal information, as well as optional spatial information and event-specific attributes. To be able to formally describe events in sports, we first introduce an event data item as an 8-tuple (see Definition 2.6).

Definition 2.6 Event data item.

An *event data item* e is a tuple $e = \langle edid, cid, \langle aid_1, ..., aid_n \rangle, \langle tid_1, ..., tid_n \rangle, ts, k, \langle (x_1, y_1, z_1), ..., (x_n, y_n, z_n) \rangle, \langle a_{1_k}, ..., a_{n_k} \rangle \rangle$ where $edid$ represents the unique identifier of the event data item, cid the unique competition or match identifier, aid_1 to aid_n the unique agent identifiers, tid_1 to tid_n the unique team identifiers, ts the timestamp, k the event type, (x_1,y_1,z_1) to (x_n,y_n,z_n) optional spatial coordinates, and a_{1_k} to a_{n_k} event type-specific attributes.

Analogous to a raw data item r (see Definition 2.4), an event data item e stores unique agent identifiers aid (e.g., athletes, players, ball) with the difference that event data items can contain also more than just one agent identifier. The same holds for the unique team identifiers tid.

Two types of event data now need to be differentiated: events happening at a single point in time (e.g., a stroke in tennis), so called atomic events, and events covering a short time period (e.g., a dribbling in football), so called non-atomic events. To formally describe both, an atomic event and a non-atomic event as a 3-tuple we introduce Definitions 2.7 and 2.8, respectively.

Definition 2.7 Atomic event.

An *atomic event* e_a is a tuple $e_a = \langle aeid, edid, k \rangle$ where $aeid$ represents the unique event identifier, $edid$ the unique event data item identifier, and k the event type.

Definition 2.8 Non-atomic event.

A *non-atomic event* e_{na} is a tuple $e_{na} = \langle neid, \langle edid_1, ..., edid_n \rangle, k \rangle$ where $neid$ represents the unique event identifier, $edid_1$ to $edid_n$ the unique event data item identifiers, and k the event type.

Atomic events contain one single event data item identifier $edid$ (see Definition 2.7) which also provides the timestamp of the atomic event via the timestamp $e.ts$ of the respective event data item. In contrast to that, non-atomic events contain a set of event data item identifiers (see Definition 2.8). The duration of a non-atomic event can be calculated via subtracting the timestamp of the last event data item $e_{last}.ts$ from the timestamp of the first event data item $e_{first}.ts$ which represent the end and the beginning of the non-atomic event, respectively.

An atomic event in tennis can be, for instance, a stroke. The attributes of this atomic event are stored in the corresponding event data item and could be, e.g., the velocity of the ball or the area where the ball hits the ground of the tennis court. A non-atomic event in football can be, for instance, a dribbling. Here, the first event data item represents the start and the last data item the end of the dribbling, respectively. The attributes of the respective data items of this non-atomic event could contain information, e.g., about the current length and direction of the dribbling on the football pitch. The access to the event data items is possible through the unique event data item identifiers $edid$ (see Definitions 2.6 to 2.8). Furthermore, it is possible to represent a non-atomic event (e.g., a dribbling or a one-two pass) also as a (new) single event data item. This is useful to simpler detect more complex events (e.g., a counter attack) from the set of event data items like described later in this section (see Definition 2.11). This new event data item then contains combined information from the corresponding event data items (e.g., the timestamp from the first data item as starting point or cumulated attribute values from all data items). To make this clearer we present Example 2.3.

Example 2.3 Representations of a one-two pass.

A non-atomic event *one-two pass* can look like the following:

$$e_{na} = \langle neid_1, \langle edid_1, edid_2 \rangle, oneTwoPass \rangle$$

The respective event data items e_1 and e_2 both are of type *successfulPass* with a sample attribute *packing* (i.e., the number of bypassed defenders) and can look like the following:

$$e_1 = \langle edid_1, cid_5, \langle aid_5, aid_7 \rangle, \langle tid_1 \rangle, ts_{15}, successfulPass, \langle (x_1, y_1, z_1), (x_2, y_2, z_2) \rangle, \langle packing : 2 \rangle \rangle$$

$$e_2 = \langle edid_2, cid_5, \langle aid_7, aid_5 \rangle, \langle tid_1 \rangle, ts_{18}, successfulPass, \langle (x_2, y_2, z_2), (x_3, y_3, z_3) \rangle, \langle packing : 3 \rangle \rangle$$

A *one-two pass* can be represented as a new event data item e_3 which combines information from e_1 and e_2:

$$e_3 = \langle edid_3, cid_5, \langle aid_5, aid_7 \rangle, \langle tid_1 \rangle, ts_{15}, oneTwoPass, \langle (x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3) \rangle, \langle packing : 5, duration : 3 \rangle \rangle$$

From a sports practical view, the two types of event data, atomic events and non-atomic events, are often handled equally during the analysis process. With Definition 2.9 we formally describe all events.

Definition 2.9 Event.

The set of atomic event identifiers $AEID$ and the set of non-atomic event identifiers $NEID$ are disjoint. The event identifier set EID represents the union of all atomic and non-atomic event identifiers:

$$EID = AEID \cup NEID$$

There are different ways of gathering event data. One option is to tag events manually. That is basically what Charles Reep did in the 1950's with the notational analysis (see Section 1.1). Today, companies like Opta[8] have a large number of employees tagging matches and competitions from different sports. The resulting event data are sold as a product. Even though the tagging is computer-aided it remains mostly a manual and time-intensive task. Another opportunity for event data generation is automatic event-detection. Systems like, e.g., StreamTeam [PRS+18; Pro20; PSS+20] can detect event data from spatio-temporal raw tracking data. Of course, events can also be detected from other raw data like, for instance, from IMU data (e.g., a stroke is detected when the ball is hit with the tennis racket). Additionally, also events might lead to the detection of a specific event, e.g., a one-two pass in football can be derived out

[8] https://www.optasports.com/

of two successful passes with a territorial gain after the second pass. In Definition 2.10 we formally describe the set of identifiers which serves, amongst others, as a basis for the transformation from raw and event data to a single event data item.

Definition 2.10 Identifier set.

The set of all raw data item identifiers $RDID^*$ and the set of all event data item identifiers $EDID^*$ are disjoint. The set of all identifiers ID^* represents the union of raw and event data item identifiers:

$$ID^* = RDID^* \cup EDID^*$$

With this identifier set we can now formally describe the transformation from raw and event data to a single event data item (see Definition 2.11). To detect a certain (sport-specific) event data item a function based on definitions from domain experts (e.g., coaches) is required.

Definition 2.11 Transformation from raw and event data to an event data item.

A single event data item e can be derived out of a single raw data item r, a set of raw data items R, or a set of event data items E which all are accessible via their unique identifiers:

$$f : \mathcal{P}(ID^*) \to E, ID \mapsto e$$

Event data, either manually tagged or extracted from raw data or other events like depicted in Definition 2.11, are mainly used for quantitative analyses (see Definition 2.2) because they can easily be aggregated. This leads to a better overview on what happened during a match or competition despite some context information will be lost. After or during a handball match it might be interesting for analysts and coaches to get information about the total number of successful passes completed by a certain player to better evaluate the performance. It might also be interesting to get information on maximum, minimum, or average values of certain event attributes. When coming back to tennis as an example, the average ball speed of the services a player made together with the success rate can be interesting when analyzing the player's performance. An additional use case for event data is the combination with video footage via certain software solutions (see Section 6.2.2).

2.2.1.3 Video Data

Turning now to the last important type of data for sports analytics: the video. Depending on the goal coaches or analysts currently pursue, different video types are focused. When coaches are interested in the tactical performance of the team, it is more suitable to watch a video covering all players on the field. Such videos are called “scouting feed” or “tactics cam”. SportecSolutions[9] is a company providing such videos for German Bundesliga, Bundesliga 2, Relegation, and Supercup. In contrast to that, a video recorded from behind the goal would be preferred by an assistant coach, who supervises the goalkeeper. This perspective is better suitable to evaluate the performance of the player. To generate videos with different angles and of different content, cameras with varying calibrations (i.e., angle, height, or position) need to be installed. The video is used for qualitative analyses (see Definition 2.1) of athletes performances. It is, as already mentioned in Section 1.2, the most important source of information because the whole context is maintained compared to, e.g., pure event data where context information might get lost.

Because raw data and event data items have a timestamp ts (see Definitions 2.4 and 2.6) it is possible to link these data to the video. The only requirement is an initial video timestamp to synchronize the video with raw and event data. This allows the analyst or the coach to directly watch and evaluate, e.g., a player’s events with the whole context in the respective video scenes.

2.2.2 Data Integration

Having presented the most important data for sports analytics we now introduce the next step of the sports data analytics pipeline, the data integration. Data integration is important to make the generated data accessible for later data analysis. In this section we focus on the three “V’s” of Big Data: (1) volume, (2) variety, and (3) velocity. We present the challenges of data integration in sports practice along these three “V’s”, the potential profits of a solid and fast data integration, as well as an outlook for future developments of sports practice for ensuring good data integration.

2.2.2.1 Volume

Formula 1 is the example which shows best that Big Data already arrived in the field of sports. Between 150 and 300 sensors are fitted to a Formula 1 car

[9] https://www.sportec-solutions.de/en/index.html

to measure various parameters like, e.g., tyre and brake temperatures, air flow, pedal positions, or engine performance. This leads to a transmission of 300 GB of data per Grand Prix weekend. Together with other data which is generated by the team the amount of data can increase up to 50 TB per week [MAP20]. In football matches the positions of all 22 players and the ball are tracked with a frequency of 25 Hz over 90 minutes. Therefore, data from one season, e.g., of German Bundesliga can have the size of about 400 GB [RM16]. Goes et al. [GKM+19] talk about 3,100,000 data points which are produced during one regular football match. These examples show the enormous amounts of data which are nowadays gathered in sports and which are representative for the first "V" of Big Data: volume. These data contain a lot of information and coaches and analysts could profit a lot from this potential if data would be accessible for their analyses. However, such amounts of data also mean a challenge for sports practice to handle and to integrate the data into an adequate IT-infrastructure.

2.2.2.2 Variety

The different types of data which we introduced in Section 2.2.1 lead to a large heterogeneity. This in turn represents the second of the three Big Data "V's": variety. Data not only come from different sources (e.g., raw sensor data, commercial event data, recorded videos, etc.) but also have different formats. The export file of sensors might be a traditional Microsoft Excel file. On the other side, commercial event data feeds and spatio-temporal tracking data feeds mostly come as Extensible Markup Language (XML) files like the ones from Opta [Opt20] or TRACAB [TOF+20]. Other technologies in turn might produce data in JavaScript Object Notation (JSON) format or even in other data formats. The main problem which results from the heterogeneity of the data reveals when it comes to the integration of the data. In most sports clubs there is no adequate infrastructure or at least not the necessary know-how to support a proper data integration into a centralized database which then would contain all data from the different sources. Instead, each available dataset is handled separately. This provokes that each department performs own analyses: the medical staff analyzes the stress profiles and sleep data of athletes, the strength and conditioning coaches regard the training load parameters of heart rate sensors and tracking data, whereas the head coach and the performance analyst take a look at the event data and KPIs of the last match to analyze athletes' tactical performances. Blobel and Lames [BL20] also observe this development:

> "In professional sports clubs, the need for a central information sys-

> tem has grown in recent years due to the increasing number of different information technology (IT) systems in each department that generate huge amounts of decentralized data."

This de-centrality is a massive drawback, because the potential of overarching analyses is not exploited. Interesting questions and correlations cannot be answered and analyzed adequately in the data analysis step like, e.g., how current sleep quality and physical stress influence the performance of an athlete. The results of such analyses would allow to proactively detect physical overload. This in turn can support decisions to let athletes recover more or at least to reduce the load in the training session which prevents injuries and maximizes performance in competitions or matches.

2.2.2.3 Velocity

The third "V" of Big Data represents the velocity which is also present in the context of sports data. As already introduced in Chapter 1 sports is very dynamic and with this also the data generation. A lot of actions and movements happen in parallel in a very small amount of time. This results in data being generated at a high velocity. As decision making, for instance, by football coaches must happen very fast to optimize the performance of the team and to prevent, e.g., disadvantages by wrong tactical measures, data integration must happen just as quickly as the data are generated. If this succeeds, coaches can profit from a live analysis because then, for instance, adaptations of tactics or formation changes can be executed based on the data-driven analysis results in real-time. However, fast data integration requires appropriate technology and know-how to adequately process and integrate the fast generated data. Stream processing is the key technology with which that problem can be tackled. This technology allows to process and integrate data into a database (nearly) in real-time. This in turn then enables that data can be provided and accessed live in the data analysis step.

2.2.2.4 Outlook

We have seen many challenges of the data integration step that come with the three "Vs" of Big Data in the context of sports data. To solve the described problems and cope with the challenges of each "V" sports practice should think about different options. First, it would be useful to collaborate with experts from computer science to get the expertise to handle and process large datasets on the one hand and to build an infrastructure allowing the data integration of various data on the other. Additionally, investing in and applying stream processing technology should be considered as this will allow a (near) real-time analysis of athlete and/or team performance.

To summarize this section, we can say that a proper and fast data integration into one database is desirable for sports practice. Data models, such as the ones we introduced in Definitions 2.4 and 2.6 can help integrating data and facilitate the access for the later analysis processes. Raw data and their attributes, for instance, can then easily be accessed via the unique raw data item identifiers $rdid$ and/or the corresponding unique sensor identifiers sid (see Definition 2.4). The same holds for event data and their corresponding attributes. These can also be accessed with their unique event data item identifiers $edid$ (see Definition 2.6). A proper data integration will enormously increase the potential of DASs as this can then access a database containing all data from different sources. This in turn allows to answer questions with larger contexts in the following data analysis step.

2.2.3 Data Analysis

With the data analysis we get one step closer to the data-based decision support for coaches and analysts. Coaches and analysts should be able to access data and to formulate queries to find answers to specific questions. This will provide the required information to support a good decision making. However, we first need to clarify that there are various contexts where decision making takes place in sports practice. Coaches and analysts have to make decisions mainly in the following seven contexts: (1) pre-match or competition analysis, (2) live-match or competition analysis, (3) post-match or competition analysis, (4) training monitoring, (5) squad or athlete selection, (6) athlete development, and (7) scouting and talent detection. In this section we briefly describe the seven different contexts of decision making in sports with short examples from football (see Example 2.4). Furthermore, we explain why the analysis processes in these contexts are different and require data in different granularity.

Example 2.4 Seven decision making contexts in football.

In the *pre-match analysis* context coaches and analysts prepare the own team for an upcoming match. This means that the opponent as well as the own team needs to be analyzed through the identification of strengths and weaknesses.

In order to be successful analysts and coaches analyze, for instance, the playing style of an opponent with the goal to anticipate what will happen in future matches [RPW+17]. The resulting information is used to develop appropriate tactics by making use of own strengths and exploiting opponents' weaknesses. In general, data from the last three to five matches are considered for the analysis process in this context. *Live-match analysis* happens during a match. Here, the coach might need to adapt the tactics and to substitute players if the match plan does not work out well. For decision making the coach only considers the last past minutes of the running match. *Post-match analysis* is important to evaluate the players' and team's performance of the past match. Here, the coach again takes a closer look at the match plan and compares whether or not the team and the players succeeded in their tasks. This in turn might have consequences, e.g., for subsequent training sessions or for the line-up in the next match. In the post-match analysis context only the past match is taken into consideration. In the *training monitoring* context only the current training session is considered. Here, decisions mainly concern the internal load of players. The latter should be handled in a way to not risk injuries of players on the one side, but also to keep the intensity level high enough for a good training effect on the other side. For the next decision making context, the *squad selection*, it is important to not only take the performance of the past match and training session into consideration. Coaches need to take a look at long-term performance trends of players, in training sessions as well as in matches. This can influence decisions about the squad for the whole season (on a club level) or for the next competition like a world championship (on a national team level). The following two contexts focus even longer time periods. In the *athlete development* context, coaches consider the athletes' technical, tactical, physical, and psychological progresses of at least the last six months to decide if a player can take the next step and, for instance, can get promoted from the youth team to the first team. In the *scouting and talent detection* context time periods of more than a year are considered. The decision to sign a talented player can be an enormous advantage from a sports but also from an economical perspective.

In Example 2.4 we see that each context focuses different time periods. This also influences the data which are required for the individual analysis processes. Different data need to be provided: raw data, event data, and video data. Additionally, data in different granularities are required, namely in the form of aggregated statistics and performance indicators which might even be KPIs.

Statistics are mostly based on event data. This is because event data can easily be aggregated and represented in statistics as previously described in Section 2.2.1.2. Statistics can also be based, for instance, on raw data like spatio-temporal tracking data (e.g., running distances). In general, a single statistic in sports is an aggregation of a set of raw data, event data, or their corresponding (type-specific) attributes which are accessible via their unique data item identifiers (see Definitions 2.4 and 2.6). The aggregation usually considers a certain (user-defined) time interval. To formally describe a single statistic we present Definition 2.12.

Definition 2.12 Statistic.

F is the set of aggregation functions which calculate, among other options, average avg, maximum max, minimum min, or total values tot:

$$F = \{f_{avg}(), f_{max}(), f_{min}(), f_{tot}(), ...\}$$

A single *statistic* $s \in \mathbb{R}$ can be derived out of a set of raw data items R or a set of event data items E which are accessible via their unique identifiers:

$$f : \mathcal{P}(ID^*) \rightarrow \mathbb{R}, ID \mapsto s \text{ with } f \in F$$

Performance indicators can be equal to single statistics (e.g., total number of successful passes) but often combine several statistics in more complex models (e.g., expected goals). Well chosen performance indicators can help coaches and analysts to identify good and bad player or team performances [DBT+07]. With clear definitions of such indicators by domain experts like coaches, their calculation for the analysis based on statistics is possible. To formally describe this we present Definition 2.13.

Definition 2.13 Indicator.

A certain (sport-specific) *indicator* $i \in \mathbb{R}$ can either be derived out of a single statistic s or a set of statistics S:

$$f : \mathcal{P}(S) \rightarrow \mathbb{R}, S \mapsto i$$

Whether an indicator represents a KPI or not is dependent on its relevance for evaluating performances of athletes or teams. This relevance is again defined by coaches or other domain experts.

Returning briefly to Example 2.4 this example highlights the need to provide coaches and analysts with the option to access raw data like, e.g., heart rate or speed data. This is important for the monitoring of athletes' internal and external loads during training sessions or competitions (e.g., via physiological data, IMU data, or spatio-temporal tracking data). Event data and video data are also useful for the analysis process in various decision making contexts like the post-match analysis. Additionally, coaches and analysts need information in the form of aggregated statistics and should be able to query them. This is useful especially in contexts which respect longer time periods like, e.g., talent detection or athlete and squad selection, respectively. It is more interesting to take a look at long-term performance which is better represented in aggregated statistics compared to the performance in only one match or competition. Performance indicators and KPIs, which are a special form of statistics (see Definition 2.13), should also be provided during the data analysis step. These will allow coaches and analysts to draw better conclusions on athletes' performances in both, contexts focusing short and contexts focusing long time periods.

As a summary, we can state that DASs should provide the users with the options to take a look at different time periods. Additionally, different data and data in different granularity should be provided as well. Together, this will allow for a better decision support in different decision making contexts.

2.2.4 Data Visualization and User Interaction

The last step of the sports data analytics pipeline is the data visualization and user interaction. Data visualization is a crucial part when it comes to the decision making processes of coaches and analysts because the resulting data of the previous data analysis step are now set into context. Depending on the analysis option (qualitative analysis, quantitative analysis, visualization, see Section 2.1.2) each type of sports data needs different ways of presentations to make the most out of the data. This is what we present in the first part of this section (see Section 2.2.4.1). Afterwards, we introduce a model for the UX design of DASs which ensures an intuitive and time-efficient analysis process with the respective systems (see Section 2.2.4.2).

2.2.4.1 Options for Presenting Data

As already introduced in Section 2.2.3 there are different contexts where decision have to be made by coaches and analysts in sports practice. In each of these contexts different types of data and data in different granularities are interesting for the decision maker. This in turn requires various types of presentations and visualizations to contextualize information which is essential to make the most out of the data.

Raw data will not help coaches if they are, for instance, provided as raw output files of the respective sensors. It is more interesting and suitable for the analysis if the raw values are put into graphs or into a live visualization of the current key value like, for instance, the heart rate. With the latter the coach can directly observe, e.g., the current level of physical exertion of an athlete. Continuous metrics like the athlete's running speed might be better visualized as line graphs instead of a sensor's output table. For spatio-temporal tracking data it will be more supportive for coaches to get visualized trajectories of athletes instead of a table filled with the raw (x, y, z) coordinates. Such a table would be just too abstract for the spatial imagination of coaches and analysts.

The presentation of event data looks somewhat different. Event data also have a temporal and often a spatial component (see Definition 2.6). For the analysis of single events it might be interesting for coaches and analysts *where* and *when* these events happened. This requires visualizations showing exactly where (on the court, field, pitch, etc.) this event occurred. However, event data might also be interesting in the form of aggregated statistics. In this case it will be more helpful when they are represented in charts or graphs (bar charts, line graphs, etc.) as such visualizations allow the detection of trends at first glance. An additional way of presenting statistics especially for the quantitative analysis (see Definition 2.2) are tables. If they are clearly presented this can also be a suitable way of providing information for the decision making process. As performance indicators and KPIs are a special form of aggregated statistics (see Definition 2.13), the previously mentioned visualization options are suitable for them as well.

The visualisation of video data is straightforward. The only requirement is a media player which is able to play the corresponding video files. The video is mainly used for the qualitative analysis (see Definition 2.1) of coaches and analysts. Additionally, there are options via software tools enhancing the video content, e.g., by using overlay features to highlight certain athletes or spaces (see Section 6.2.2). Furthermore, the synchronization of the video with raw and event data, as already mentioned in Section 2.2.1.3, provides an interesting opportunity for the joint presentation of important information from different sources.

In summary it has been shown that different data types and data of varying granularities require different visualization options. These options should be provided by DASs as data visualization is the key step to contextualize information and thus has the biggest influence on the decision making process of coaches and analysts in sports practice.

2.2.4.2 User-Centered Design of a Data Analytics System

To allow for a time-efficient analysis process it is important to query, analyze, and visualize data in an intuitive and flexible way. For that, the development and design of DASs needs to be well-planned, and with that also the UX. UX is a multifaceted construct [DS15]. Garrett [Gar12] therefore presents a model consisting of five elements which have to be considered for a user-centered design: (1) strategy, (2) scope, (3) skeleton, (4) structure, and (5) surface (see Figure 2.5). Originally this model was developed particularly for web design but it can also be used for the design of other interactive applications like, for instance, a DAS. Each of the five elements, or layers, is dependent on the layer(s) below. Starting from the lowest layer, the questions we have to deal with get more concrete with each layer [Gar12]. Following this approach will ensure to develop products with a good UX because the user is always particularly focused during the whole planning phase. The individual layers are described more detailed in the following.

Strategy. The first layer is called strategy. Here, we need to define the fundamentals like the general product objectives and the user needs. We need to answer questions like "Why are DASs for sports needed?", "What is the benefit of a DAS for the user?", or "In which context should a DAS be applied?". A lot of research is required to really understand the interests of target groups and their intentions of usage. All further decisions we make are dependent on the strategy.

Scope. Now that we know what the product objectives are and what the users need we can continue with the second layer, the scope. Here, we plan how we can reach the previously defined goals. For that, it is important to translate the product objectives and the user needs into specific requirements [Gar12]. This means that functional specifications need to be defined based on the strategic decisions from the previous layer. Moreover, content requirements need to be assessed and specified.

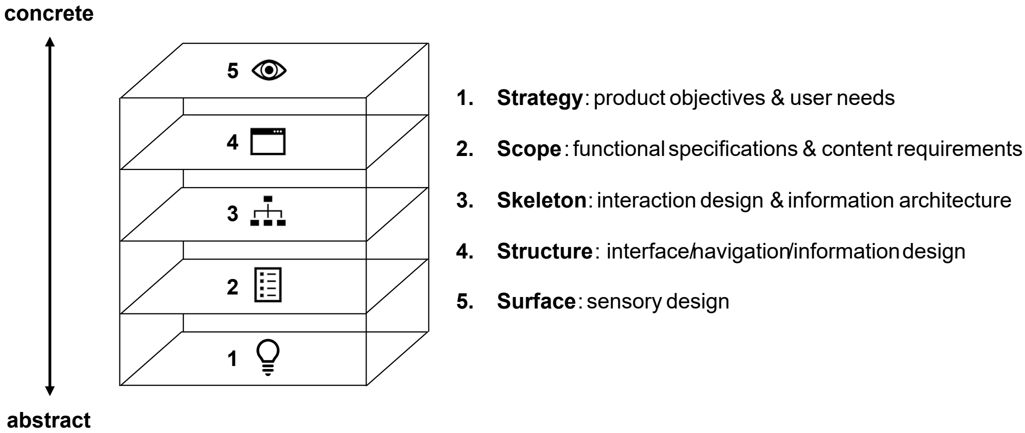


Figure 2.5 Five-layer model for user-centered design according to Garrett [Gar12].

Skeleton. The development and design of a DAS becomes more concrete with the third layer, the skeleton. The previous layers resulted in a precise picture of the product's content. Therefore, we can now continue with planning the interaction design. For this, some research on potential use cases is required to understand how users will behave and interact with the system. This in turn will ensure that the DAS can be developed in a way to provide the user with everything to fulfill his/her tasks. Additionally, the information architecture is focused in this layer. It is important to structure the content-related elements in a way which is understandable for all users [Gar12].

Structure. The requirements originating from the strategy layer get more and more concrete. In the fourth layer, the structure, we need to clarify and define the specific requirements for the designs of interface, navigation, and information. Here, the previously collected data on use cases serve as a basis. The elements of the interface should be chosen and arranged in a way to allow an easy interaction for the users. The navigation design focuses on the elements with which the user can move through the information architecture. Additionally, information need to be presented in an understandable way for an easy communication [Gar12].

Surface. The last layer is called surface. Here, it is all about the User Interface (UI) and the concrete details of visualization and sensory design of all contents and functionalities. The DAS should appeal to the senses and fulfil all goals formulated in the previous four layers [Gar12].

Part II

Concepts and Implementation

The purpose of abstraction is not to be vague, but to create a new semantic level in which one can be absolutely precise.

— Edsger W. Dijkstra

3 Conception and Design of Data Analytics Systems for Game Sports

Thus far, we have seen in Chapters 1 and 2 that performance analysis and decision making in sports, especially in game sports, is very complex. Additionally, we illustrated in Section 1.2 the challenges which performance analysts are exposed to in their daily work. In sum, this highlights the need for data-driven analytics for decision making in game sports. DASs would be very helpful for analysts and coaches in sports practice if they are developed appropriately. To ensure the latter we now introduce concepts which allow the development of DASs for optimal decision support in game sports. Additionally, we present a specific approach for designing the UX of these DASs. This in turn allows for a flexible and efficient workflow and thus further facilitates the work of analysts and coaches.

In the following, we present a way how to bridge the semantic gap of sports analytics. This ensures that DASs will provide information which really is relevant for the decision making process of coaches and analysts in sports practice (see Section 3.1). Furthermore, we present an approach to describe the complexity levels of different game sports. This highlights the requirement for DASs to combine different analysis options (see Section 3.2). Moreover, we show that DASs need to support different query types to allow for a holistic information retrieval. Additionally, we propose how to model data in order to allow for a performant retrieval process and further optimize the analysis process in game sports (see Section 3.3). Finally, we present a generic model for designing the UX of DASs which will enable analysts and coaches to work flexible and (time-) efficient by intuitively accessing, analyzing, and visualizing data with a DAS (see Section 3.4).

3.1 Bridging the Semantic Gap of Sports Analytics

In Section 2.1.3 we presented the semantic gap of sports analytics in detail. In a nutshell, this gap means that the resulting information of the analysis process is not always optimal and sufficient for a good decision support because data often do not represent the concepts of coaches. For the development of powerful DASs it is thus crucial to bridge or at least to narrow the semantic gap of sports analytics even if this is a very challenging task. However, this will then allow to develop DASs which provide coaches and analysts with really relevant information for their decision making processes in game sports because the coaches' concepts are finally represented in the resulting information from the analyses. To make exactly this possible, we now introduce an approach how to bridge the semantic gap of sports analytics.

In this section, we take a closer look at the coaches' concepts which are applied to implement the mental models. These concepts are fundamental for the analysis and the decision making process. Despite the decision itself is not part of the data analytics pipeline we introduced in Section 2.2, it nevertheless influences the whole rest of the pipeline. We try to visualize this in Figure 3.1 which represents a data analytics pipeline and which we extended by the visualization of the semantic gap of sports analytics.

Our approach to bridge the semantic gap of sports analytics consists of three major steps (see Figure 3.2) and starts from the coaches' side and not from the data side. Parts of the approach have already been presented by Seidenschwarz et al. [SRP+20a]. We introduce the approach using football as an example. First, we present a way how to extract concepts from coaches through conducting semi-structured interviews and applying methods from qualitative content analysis (see Section 3.1.1). Second, we assign the resulting concepts to different data types in a performance modeling step (see Section 3.1.2). Third, we translate the performance model into a data model which is needed for a subsequent implementation (see Section 3.1.3). Despite this approach is shown in the context of football, we want to mention here that this is a generic approach which can be applied to other game sports without major changes.

3.1.1 Performance Factors Assessment

To bridge the semantic gap of sports analytics several steps need to be taken. In this section we present the first step, the performance factors assessment.

We have already seen that coaches have their own mental model of the sport

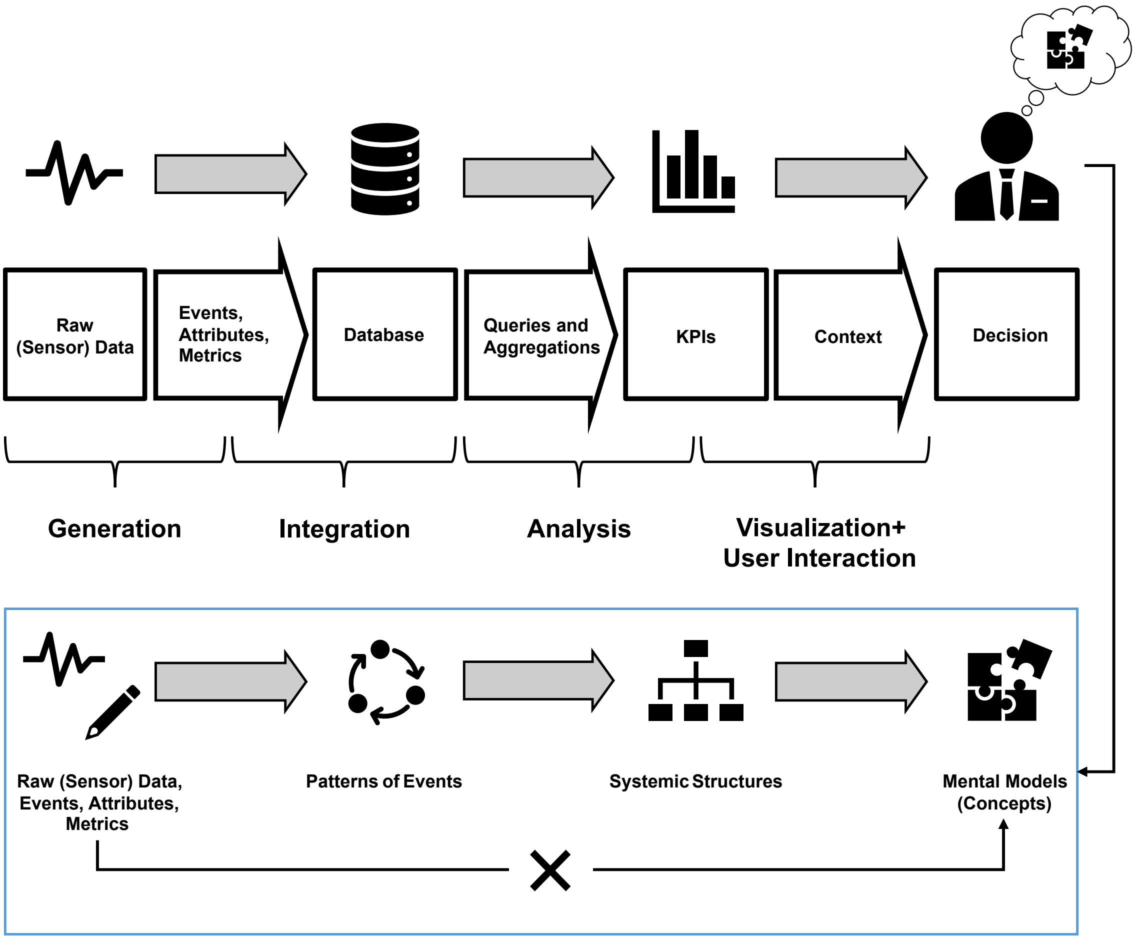


Figure 3.1 **Data analytics pipeline extended by the semantic gap of sports analytics (blue box) which shows that concepts of coaches are often not represented in the data used for the analysis proces.**

in which they are engaged and that their teams apply specific concepts to implement it (see Section 2.1.3). If we want to provide coaches with DASs which really can support their decision making processes, we must extract and understand these concepts. We furthermore need to model the concepts into systemic structures and patterns of events (see Figure 3.1). This is essential to generate, for example, raw data, event data, statistics, or indicators, which represent the coaches' concepts and only then allow DASs to provide really useful analyses. In the following, we show a way how to get access to domain knowledge of coaches and how concepts can be extracted. We present our approach in the context of the invasion game football to demonstrate a specific use case. We even further limit the context by focusing only on the match plan which represents a collection of tactical decisions that have to be taken before a match [SRP+20a].

We conducted semi-structured interviews with six UEFA Pro License coaches where the questions primarily concerned the match plan. This means that most

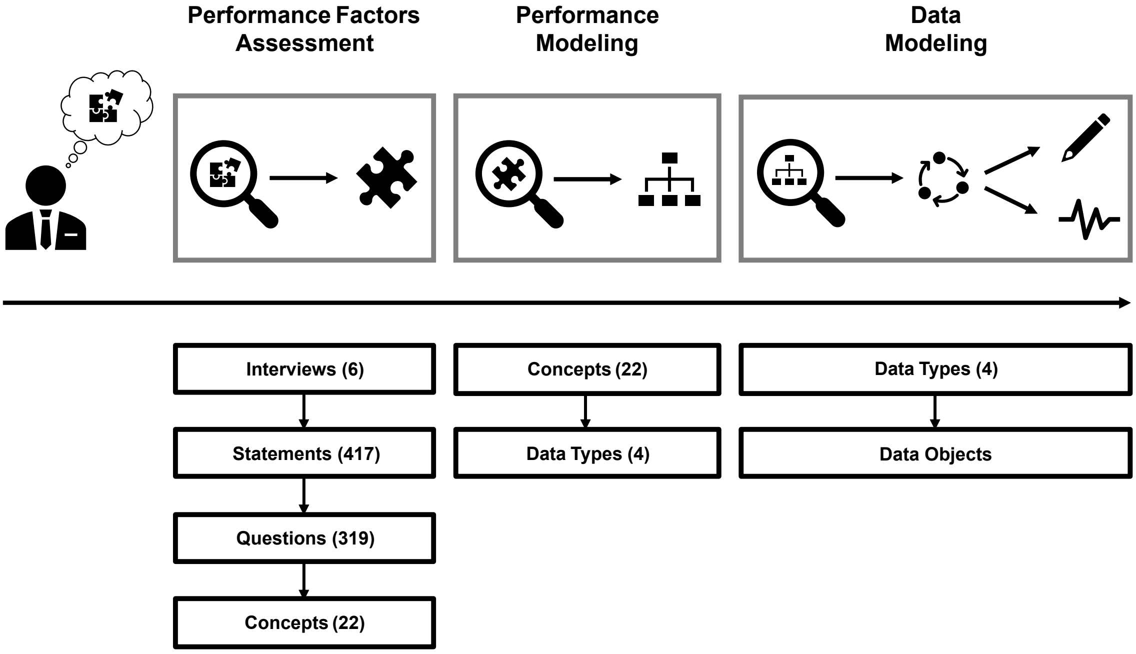


Figure 3.2 Three steps to bridge the semantic gap of sports analytics.

of the questions covered the following three decision making contexts of coaches in particular: (1) pre-match analysis, (2) live-match analysis, and (3) post-match analysis (see Section 2.2.3). This is the case, because the match plan is always taken as a reference whether a certain team or player performance can be achieved (pre-match), is achieved (live-match), or was achieved (post-match). The interview guideline which we have created consists of 29 questions and is attached in Appendix A.

Each question can be categorized into one of nine different topics. All topics together with one sample question each and the resulting number of coaches' statements and questions on that topic are depicted in Table 3.1. The interviews were recorded using an audio recording device. Afterwards, the audio files were transcribed into text files. To analyze these, all detailed examples and descriptions of certain situations which coaches have explained are deleted and only general statements on concepts, patterns of play, and arising questions (n=417) are considered in a first step [SRP+20a]. Then we made use of Kuckartz' approach of content-related structured qualitative analysis [Kuc18]. This approach is used to categorize the coaches' statements and questions into main categories and subcategories. Afterwards, only statements (n=260) and questions (n=59) of coaches that can be assigned to one of the nine topics from Table 3.1 are considered for the following steps. To bring all data on the same level, we created questions out of each of the remaining 260 statements. This leads to a total number of 319 questions which all reveal the individual information needs of coaches

Table 3.1 The nine topics of the interview guideline with sample questions and the number of resulting statements and questions of coaches on each topic (#) according to Seidenschwarz et al. [SRP+20a].

Topic	Example	#
Principles	What are possible reasons for changing the system?	97
Player Profiles	What are possible weaknesses of a player?	63
Offensive Organization	Where and how can the offense be started?	42
Opponent Information	How is an opponent's key player characterized?	41
Set Plays	How should the defense act in a corner kick situation?	26
Defensive Organization	How can the defense be organized?	24
Transition DEF-OFF	What options exist during transition DEF-OFF?	12
Build-up Play	What options exist during build-up play?	7
Transition OFF-DEF	What options exist during transition OFF-DEF?	7
Total		*319*

when preparing or adjusting the match plan [SRP+20a]. In a final step we extracted the underlying concept of each question. The concept is the semantic element in the question for which we have not yet an unambiguous definition. To better illustrate these single steps we present Example 3.1.

Example 3.1 The procedure of concept extraction.
Three steps to extract concepts from coaches' statements according to Seidenschwarz et al. [SRP+20a].

1. **Statement.** The defending team decides where to start a pressure situation.
2. **Question.** Where does the defending team start a pressure situation?
3. **Concept.** Pressure.

Having completed this procedure for all 319 questions we were able to extract 22 different football-specific concepts. These concepts together with a short description and the corresponding number of questions from which these concepts were extracted are listed in Table 3.2.

3.1.2 Performance Modeling

After the successful extraction of 22 football-specific concepts of coaches, the next step to bridge the semantic gap of sports analytics is the assignment of these concepts to different data types. To realize that for the 22 concepts, we need to introduce four different data types: (1) events, (2) phases, (3) continuous

Table 3.2 The 22 resulting concepts of coaches with short description and the total number of questions (#) according to Seidenschwarz et al. [SRP+20a].

Concept	Short description	#
Individuality	Player profile: mental, physical, tactical, technical	50
Line-up	Formation of a team and system of play	42
Attacking Play	Team speeds up, attempt to create a scoring opportunity	34
Pressure	Pressure of the defensive team on the offensive team	33
Principle	General idea how to play, line-up and movements	29
Running Trajectory	Trajectory per player from acceleration to deceleration	18
Coordination	Degree of team organization: compactness, synchronicity	14
Set Play Variation	Different options in offensive set plays	14
Defensive Play	Defensive schema after loss of ball possession	11
Cooperation	Profile of functional units: duos, trios, etc.	9
Zonal Marking	Defending player covers a specific space on the pitch	9
Open Space	Open spaces on the pitch for the offensive team	8
Build-up Play	Before attacking play, intention to gain territory	7
Manipulating Space	Creating, defending, and using space on the pitch	7
Man-to-Man Marking	Assignment of a player to a direct opponent	7
Gain Possession	Player gains possession of the ball	6
Change of Speed	Transition from build-up play to attacking play	5
Duel	Two players involved, intention to gain/defend possession	5
Passing Option	Number of players with high likelihood to receive a pass	4
Lose Possession	Player loses possession of the ball	3
Fall Back	Defending players quickly getting behind the ball	2
Orientation	Orientation of a player relating to the opponent's goal	2
Total		*319*

states, and (4) profiles. A schematic representation is shown in Figure 3.3.

Events, phases, and continuous states contain information about the (current) dynamics of a match. In contrast to that, profiles contain information which has been aggregated over a certain number of past matches [SRP+20a]. In the following we present these four data types in more detail. Afterwards, we present a concrete example which shows how these four data types are related and how some of the extracted concepts are represented therein.

Events. We already introduced atomic and non-atomic events in Section 2.2.1.2. In a nutshell, an atomic event is a distinct, basic, and inseparable event that occurs at a specific time [SRP+20a]. In contrast, non-atomic events cover a short time period. A formal description of both event types is given in Definitions 2.7 and 2.8. Both event types can be used to segment a match into phases.

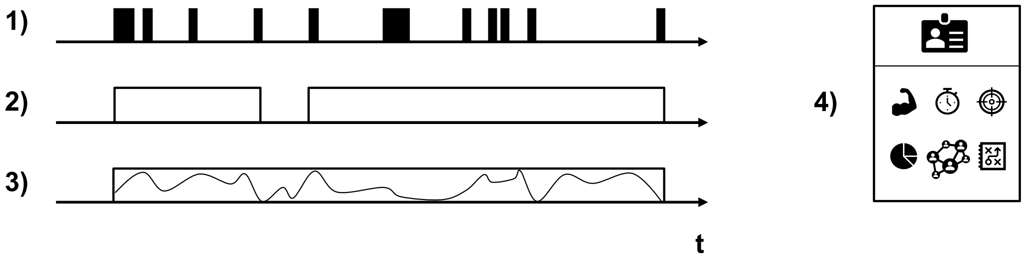


Figure 3.3 Schematic representation of the four different data types based on Seidenschwarz et al. [SRP+20a]: (1) events, (2) phases, (3) continuous states, and (4) profiles.

Phases. A phase segments the dynamics of a match. It has a beginning and an end. In a phase, information is aggregated (i.e., as statistics and indicators) over the time period delimited by the phase (e.g., the number of successful passes during a build-up play phase) [SRP+20a]. To formally describe a phase as a 9-tuple we introduce Definition 3.1.

Definition 3.1 Phase.

A *phase* p is a tuple $p = \langle pid, cid, ts_{start}, ts_{end}, j, \langle aid_1, ..., aid_n\rangle, \langle tid_1, ..., tid_n\rangle, \langle s_1, ..., s_n\rangle, \langle i_1, ..., i_n\rangle\rangle$ where pid represents the unique phase identifier, cid the unique competition or match identifier, ts_{start} the start timestamp, ts_{end} the end timestamp, j the phase type, aid_1 to aid_n the unique agent identifiers, tid_1 to tid_n the unique team identifiers, and s_1 to s_n as well as i_1 to i_n the aggregated statistics and indicators during the phase.

Continuous States. A continuous state can be defined during a phase and can further characterize the latter. Otherwise, it can also be regarded separately. It consists of a continuous (raw) data stream which denotes the evolution of a set of parameters over time. These parameters can be discrete (e.g., the number of players behind the ball) or continuous (e.g., the average distance between players). Continuous states can be defined for individual players (e.g., *Running Trajectory*), for groups of players (e.g., *Coordination*), or for the entire team (e.g., *Pressure*) [SRP+20a]. With Definition 3.2 we formally describe a continuous state as a 7-tuple.

Definition 3.2 Continuous state.

A *continuous state* cs is a tuple $cs = \langle csid, cid, k, \langle aid_1, ..., aid_n\rangle, \langle tid_1, ..., tid_n\rangle, \langle r_1.ts, ..., r_n.ts\rangle, \langle r_1.a, ..., r_n.a\rangle\rangle$ where $csid$ represents the unique continuous state identifier, cid the unique competition or match identifier, k the continuous state type, aid_1 to aid_n the unique agent identifiers, tid_1 to tid_n the unique team identifiers, and $r_1.ts$ to $r_n.ts$ and $r_1.a$ to $r_n.a$ the streams of timestamps and attributes of the corresponding raw data items.

Profiles. A profile characterizes individual players, groups of players, or even whole teams through aggregated information from previous matches (e.g., in the form of statistics and indicators). Individual player profiles may contain information on the physical fitness and summarized information on typical running trajectories. Profiles of groups of players (e.g., the defensive line) may contain measures of *Cooperation*. Team profiles might contain typical *Line-Ups* and *Principles* [SRP+20a]. To formally describe a profile as a 7-tuple we introduce Definition 3.3.

Definition 3.3 Profile.

A *profile pr* is a tuple $pr = \langle prid, l, \langle cid_1, ..., cid_n \rangle, \langle aid_1, ..., aid_n \rangle, \langle tid_1, ..., tid_n \rangle, \langle s_1, ..., s_n \rangle, \langle i_1, ..., i_n \rangle \rangle$ where $prid$ represents the unique profile identifier, l the profile type, cid_1 to cid_n the unique competition or match identifiers, aid_1 to aid_n the unique agent identifiers, tid_1 to tid_n the unique team identifiers, and s_1 to s_n as well as i_1 to i_n the aggregated statistics and indicators.

An overview on how the 22 concepts are assigned to and distributed between the four different data types is depicted in Table 3.3. Example 3.2 and Figure 3.4 are used to illustrate how the different concepts and data types are related. Of course many other examples can be found which confirm that our approach works until this point. However, to make an implementation of these results into a DAS possible one final step is still pending, the data modeling.

Table 3.3 Distribution of the 22 concepts between the different data types.

Data Type	Concepts
Events	Change of Speed, Duel, Gain Possession, Lose Possession
Phases	Attacking Play, Build-up Play, Defensive Play, Set Play Variation
Continuous States	Coordination, Fall Back, Man-to-Man Marking, Manipulating Space, Open Space, Orientation, Passing Option, Pressure, Running Trajectory, Zonal Marking
Profiles	Cooperation, Individuality, Line-ups, Principles

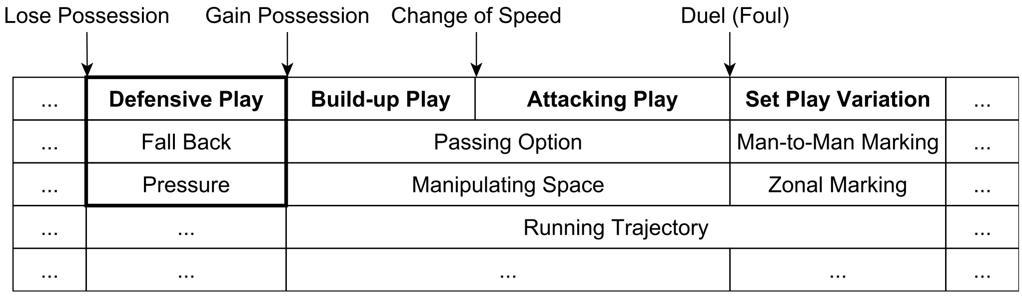


Figure 3.4 Defensive play stack.
Relation of events, phases, and continuous states with highlighted *Defensive Play* stack (black box) according to Seidenschwarz et al. [SRP+20a].

Example 3.2 Defensive play stack.
The example shows how events, phases, and continuous states are related.

We present the *Defensive Play* stack in Figure 3.4. We can see that the two atomic events *Lose Possession* and *Gain Possession* are used to segment the match into phases. After ball possession is lost, a team's defending phase starts and lasts until the possession of the ball is regained. During this *Defensive Play* phase, two continuous states are particularly interesting for coaches. First, it is important that the team exerts *Pressure* on the opponent's player in possession. If the regaining of possession does not succeed, it is important for the defending team to quickly get as many players as possible behind the ball (*Fall Back*).

3.1.3 Data Modeling

The last step which is required to bridge the semantic gap of sports analytics is the data modeling. The result from Section 3.1.2, the performance model, needs to be translated into a data model. This step is essential for a subsequent implementation into a DAS. However, this then will finally enable coaches and analysts to perform really useful analyses with these systems. High-level tactical analyses are also an important requirement in the daily work of performance analysts as we have seen in Section 1.2. These will then also be possible with DASs because all the concepts of a coach are finally represented in the data.

A data model consists of data objects with each having mandatory and optional attributes. In the following, we take a look at each of the four data types which we introduced in the previous section and explain which mandatory and optional attributes they contain.

Events. An atomic event e_a points to exactly one event data item (see Definition 2.7), a non-atomic event e_{na} to several event data items (see Definition 2.8). Therefore, the following attributes are mandatory: event data item identifier(s), match identifier, player and team identifiers, timestamp, type, and (x,y,z) coordinates. Optionally, one can add a set of event type-specific attributes (see Definition 2.6). Example 3.3 highlights the attributes for the atomic event *Duel*.

Example 3.3 Mandatory and optional attributes of the atomic event duel.

When we take a *Duel* as example for an atomic event we need to know when (i.e., match and timestamp) and where (i.e., location) the duel happened. Additionally, it is important to know the identity of the players and their teams who were involved in this duel. Optionally, we might get information on the outcome of the duel like, e.g, if the duel was won or lost, if it resulted in a foul and further into a yellow or a red card.

Phases. The mandatory attributes of a phase p are the phase identifier, the match identifier, the two timestamps which delimit the phase, the type, as well as the identifiers of the involved players and teams. Often, phases are delimited by events. In these cases the two timestamps ts_{start} and ts_{end} are identical to the ones of the corresponding (first) event data items. However, phases can also be delimited by continuous states, e.g., if the value of a metric exceeds a certain threshold. Optionally, phases can contain attributes consisting of aggregated information (i.e., statistics or indicators) over that phase (see Definition 3.1). To illustrate that we take a look at the *Attacking Play* phase (see Example 3.4).

Example 3.4 Mandatory and optional attributes of the attacking play phase.

Two atomic events delimit the *Attacking Play* phase: *Change of Speed* and, in our example, a *Duel* resulting in a foul (see Figure 3.4). Consequently, the timestamps of the phase are equal to the timestamps of the two event data items. Additionally, it is important to know the identity of the players and their teams who were involved in the phase. Optionally, we might get information on the total number of successful passes or the fastest sprints during that phase.

Continuous States. The mandatory attributes of a continuous state cs are the respective continuous state identifier, the match identifier, the type, as well as the streams of timestamps and raw data item attributes. Optional attributes of continuous states can be used to further describe the metric [SRP+20a]. This can be the identifiers of the involved players or teams (see Definition 3.2). To better explain this we take the continuous state *Running Trajectory* as an example (see Example 3.5).

Example 3.5 Mandatory and optional attributes of the continuous state running trajectory.

The *Running Trajectory* consists of a spatio-temporal tracking data stream with the mandatory attributes timestamp and (x, y, z) coordinates. Optionally, the player and team reference can be an additional attribute.

Profiles. A profile *pr* has the following mandatory attributes: a profile identifier, a type, and references to a player or to a set of players. Optional attributes can represent aggregated information in the form of statistics or indicators on matches and can be added to the profile [SRP+20a]. Other optional attributes are the references to matches and teams (see Definition 3.3). In Example 3.6 we show potential attributes for the profile *Cooperation*.

Example 3.6 Mandatory and optional attributes of the profile cooperation.

The profile *Cooperation* consists of the player reference and of one or several cooperation measures. Such a measure can be the involvement of the player in the passing of the team. It can be interesting, for instance, with which team mates the player passes most often.

Having described the data types and their mandatory and optional attributes we finally come back to the *Defensive Play* stack which we introduced in Example 3.2 and Figure 3.4. This example shows that we can represent all its concepts by five data objects: two atomic events (*Lose Possession* and *Gain Possession*) which delimit the *Defensive Play* phase and two continuous states that add evolving metrics to describe the phase, namely *Pressure* and *Fall Back* [SRP+20a].

We successfully showed that we can bridge the semantic gap of sports analytics by extracting concepts from domain experts like coaches, modeling these concepts into a performance model, and subsequently translating that model into a data model which finally allows an implementation into a DAS. By adjusting the interview guideline, our approach can be transferred to other (game) sports as well.

3.2 Combining Different Analysis Options

In the previous section we have shown how to bridge the semantic gap of sports analytics. However, there are still other points which have to be considered for the development of potent DASs for game sports. In Section 1.2 and Section 2.1.2 we have seen that analysts and coaches need different options for their analyses: qualitative analysis, quantitative analysis, and visualization. In this section we will now highlight why a combination of these three options is especially important for (tactical) analyses in more complex game sports like invasion games, compared to simpler game sports. For that, we present an approach to describe the complexity levels of the different subcategories of game sports (see Section 2.1.1) in Section 3.2.1. We furthermore highlight what these results imply in general for the development of DASs to provide a good decision support in game sports (see Section 3.2.2).

3.2.1 The Complexity of Game Sports

Based on examples we present how the subcategories of game sports differ in complexity. For our approach we use the following three points to describe the complexity levels: (1) game state, (2) segmentation, and (3) interaction. Before starting with the examples we first want to introduce this systematics.

At any moment a match is in a certain state which we call the game state. This can be, for instance, the current score, the player or team in attacking mode, or the spatial arrangement of players or playing equipment, like balls or pieces. It can also be a combination of different factors. The game state can be modelled arbitrarily complex. In general, the more difficult it is to describe the game state, the more complex a sport is. We will see in the course of this section that there are game sports where it is easier to describe the game state but also others where this task is more difficult. We want to mention at this point that there is no strict rule how the game state has to be described. The game state thus can be modelled very individually.

The second point is segmentation which means how and how easy a single match can be divided into different segments. The segmentation of a match into phases based on certain criteria is a common task in sports analysis [GH17]. There are, e.g., regular segmentations like breaks defined by the rules of the game, but also artificial segmentations of the match defined, e.g., by coaches. Single events as well as the game state can also serve to segment the match. A segmentation is required from an analysis perspective as it is essential to analyze performances of the same or at least of comparable segments, e.g., by aggregating events or attributes over certain time periods. This also means that the more segmented a sport already is due to the rules of the game, the easier it gets to analyze it. On the other side, it gets more complex to analyze sports with a low level of segmentation where it is necessary to implement artificial segmentations.

The third element of our systematics is interaction, which we already introduced briefly in Section 2.1.1. There exist several types of interaction, e.g., interactions between athletes (inter-athlete), between or within small groups of athletes (inter-group and intra-group), or between or within whole teams (inter-team and intra-team). Game sports with no or less interaction also have less complex tactical patterns. On the other side, complex tactical patterns get more numerous in game sports with lots of interaction because more athletes are involved which are all mutually influencing each other. From our understanding these are the crucial points which allow to draw conclusions about the complexity of a specific game sport.

To get more concrete, we now start with the description of the complexity levels of four different sports. Again, we want to mention that the way we described the game states, segmentations, and interactions represents only one option. In the following we also explain some of the rules of the corresponding sports. These rules are only excerpts of the official rules of the game and are not intended to be complete but serve only to illustrate the point of segmentation. The first sport we take a closer look at is billiards (see Example 3.7) which is an example for a target game (see Table 2.2).

Example 3.7 Complexity of the target game billiards.

Game State. The game state consists of the player whose turn it is, the spatial arrangement of the balls, the number of balls in the pockets and on the table, and the overall game score.

Segmentation. The segmentation is defined by the rules of the game (e.g., "eight-ball"). One player starts the first game. It is this player's turn and he/she continues shooting each time he/she is scoring (pocketing). The player's turn ends either if the game is won with pocketing the decisive ball (black ball #8) or after a failed shot or a foul. Then it is the opponent player's turn. After one game is decided the next game starts with the other player.

Interaction. The interaction lies in the adaptation of the player to the new game state originated from the last move the opponent made.

To describe the complexity level of striking and fielding games, we use baseball as an example (see Example 3.8)

Example 3.8 Complexity of the striking and fielding game baseball.

Game State. The game state consists of the team roles (offense/defense) which are assigned at a moment. Additionally, the current pitcher, batter, and catcher have to be considered. The distribution of the players across the different bases (offensive team) and the spatial arrangement of the players on the field (defensive team) also have to be taken into account. The game score is an additional factor describing the game state.

Segmentation. The segmentation is defined by the rules of the game. One team is assigned the offensive team and can score points (i.e., to make runs), the other team is the defensive team. After each half-inning the team roles change and the offensive team becomes the defensive team and vice versa. The pitcher throws the ball towards the catcher. The opponent's batter tries to hit the ball. If he succeeds, the batter runs as far as possible from base to base and ideally reaches the homebase. Otherwise, he stops at a base before the ball gets back to that base. After the ball is back at a base the next turn starts and the pitcher will start again throwing the ball.

Interaction. Inter-athlete interactions mainly consist of the pitcher-batter relation. Intra-group and intra-team interactions are mainly apparent for the defensive team when throwing/catching the ball until it is back at a base.

Tennis is used in the following to describe the complexity level of net and wall games (see Example 3.9).

Example 3.9 Complexity of the net and wall game tennis.

Game State. The game state consists of the current game score (games and sets) and the serving player.

Segmentation. The segmentation is defined by the rules of the game. A tennis match consists of games and sets. Both players can score points during a game. A game starts with one player serving and is won by the first player to score at least four points in total and at least two points more than the opponent player. A set consists of several games with the serving player alternating between games. In general, a player wins a set by winning at least six games and at least two games more than the opponent player. A tennis match consists of a sequence of sets. The winner is determined through a best of three or a best of five sets system.

Interaction. Inter-athlete interactions in tennis are highly dynamic as each player has to react to the opponent's shots.

The last subcategory of game sports are invasion games. To point out the complexity level we use football as an example (see Example 3.10).

Example 3.10 Complexity of the invasion game football.

Game State. The game state consists of the current game score, the team in possession of the ball, as well as the spatial arrangement of the players and the ball on the pitch.

Segmentation. In football the only segmentation defined by the rules of the game is the partition into first and second half. But during each half there is no discrete segmentation possible due to the continuous fluidity of the game. So there is a higher need for artificial segmentations. Tactical periodization is one example for that and segments the game into the following four phases: (1) offensive organization, (2) transition from attack to defense, (3) defensive organization, and (4) transition from defense to attack [DM12].

Interaction. There are intra-team interactions like, e.g., pass sequences between team mates. There also exist intra-group interactions between different team parts like, e.g., the defensive and midfield lines. Inter-team, inter-group, and inter-player interactions are present during the whole match, because it is always needed to adapt to the opponents behaviour on an individual level, on a group level, and also on a team level.

Example 3.10 shows that the modeling of the game state in invasion games is particularly complex. The description of the game state like depicted above is not satisfying because in football the score is mostly too low to allow for a valid

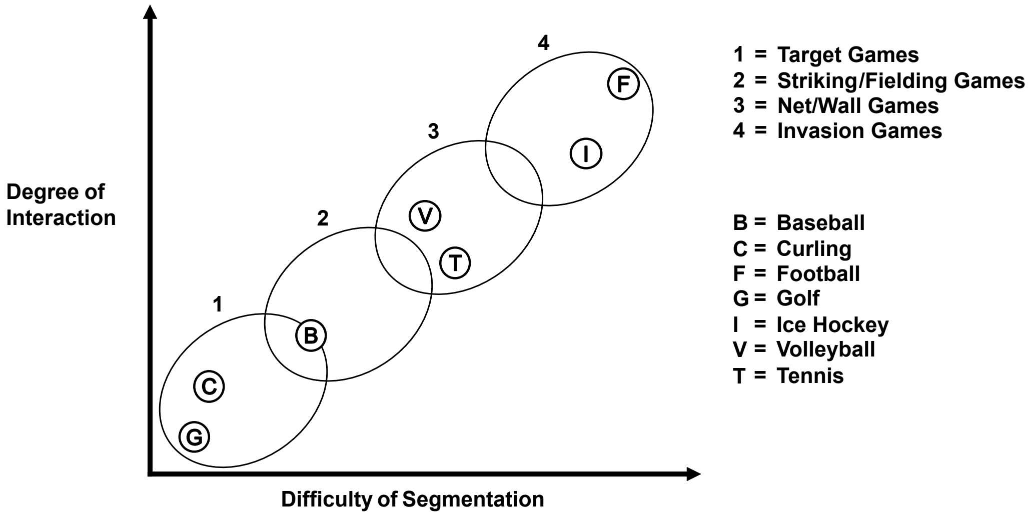


Figure 3.5 Segmentation-interaction matrix showing the four different types of game sports (bubbles) with some specific sports as examples.

representation of the current state of the game. There is often the scenario in which the better team fails to score despite lots of attempts whereas on the other side the inferior opponent succeeds in scoring a goal out of only one chance.

In summary, we can state that there are major differences concerning the complexity levels of the four subcategories of game sports. Of course not all sports of one category can be equated. It is more likely that there exist complexity "bubbles" for each type of game sports. We try to visualize the complexity levels of different game sports on a segmentation-interaction matrix (see Figure 3.5). This matrix shows the four subcategories of game sports as bubbles and some specific sports as examples representing our opinion of a sport's complexity level. The bubbles are overlapping which means that there might be sports which are more or less complex as their type would let expect.

3.2.2 Consequences for Data Analytics Systems

Having described the different complexity levels of game sports, we now take a closer look at what these results imply for the analysis processes. At this point we want to mention that we focus on the tactical analysis of a sport and not on the technical aspects which are of course also a very important factor. In Section 2.1.2 we presented the need for different options for the (tactical) analysis depending on the complexity and frequency of certain relevant patterns in a sport. A higher complexity level of a sport entails of course also more complex patterns which occur in the course of a match. This can be due to the difficulty in describing the game state because of the continuous fluidity of the sport, the poor segmentation, as well as the dynamic and numerous interactions between athletes, between groups, or also between whole teams which result in anecdotal descriptions of athlete or team performances by the coaches. This higher complexity causes a stronger need for qualitative analyses in addition to quantitative analyses and visualizations (see Definitions 2.1 to 2.3) which are required anyway because of the high number of simple patterns. Complex patterns just can be better analyzed qualitatively by watching the video scenes. On the other side, sports with a lower complexity level not necessarily need qualitative analyses as the (tactical) performance of athletes and teams can be described more easily with statistics in a quantitative analysis like it is the case, for example, in baseball. This implies that DASs which are applied, e.g., in invasion games have to combine all three analysis options to provide an optimal decision support. Despite not all game sports would need that combined provision of the three analysis options it would nevertheless be a big benefit for coaches and analysts in all game sports to have this combination. As previously mentioned in the beginning of this section we focus on the tactical aspects of performance analysis. However, less complex game sports will profit a lot from a qualitative analysis as well, especially for technical performance analyses. Therefore, if a DAS covers the different analysis options, coaches would be completely flexible to conduct their individually preferred analyses in specific contexts and situations.

With Figure 3.6 we illustrate the consequences of the previous results for the development of DASs to provide an optimal decision support in game sports. This figure is an advanced version of the data analytics pipeline which we introduced in Section 2.2 and consists, strictly speaking, of three separated pipelines. These pipelines are necessary for the support of the three analysis options: qualitative analysis, quantitative analysis, and visualization. Here, we want to mention, that these three pipelines cannot be separated strictly in sports practice. It is often the case that the processes get blurred like it is the case, for instance, with video overlays displaying, for instance, dominant regions, passing distances, or other sport-specific concepts (see Section 6.1.1.2).

The upper pipeline represents a typical workflow of a qualitative analysis. Captured video material is used for a subjective video analysis by watching matches or certain scenes of a match. The middle pipeline represents the process of a quantitative analysis. Typically, event data or derived metrics from raw (sensor) data like, for instance, the number of players behind the ball are aggregated as statistics. Additionally, performance indicators are calculated. With subsequent visualizations like, e.g., in the form of tables, bar charts, or line

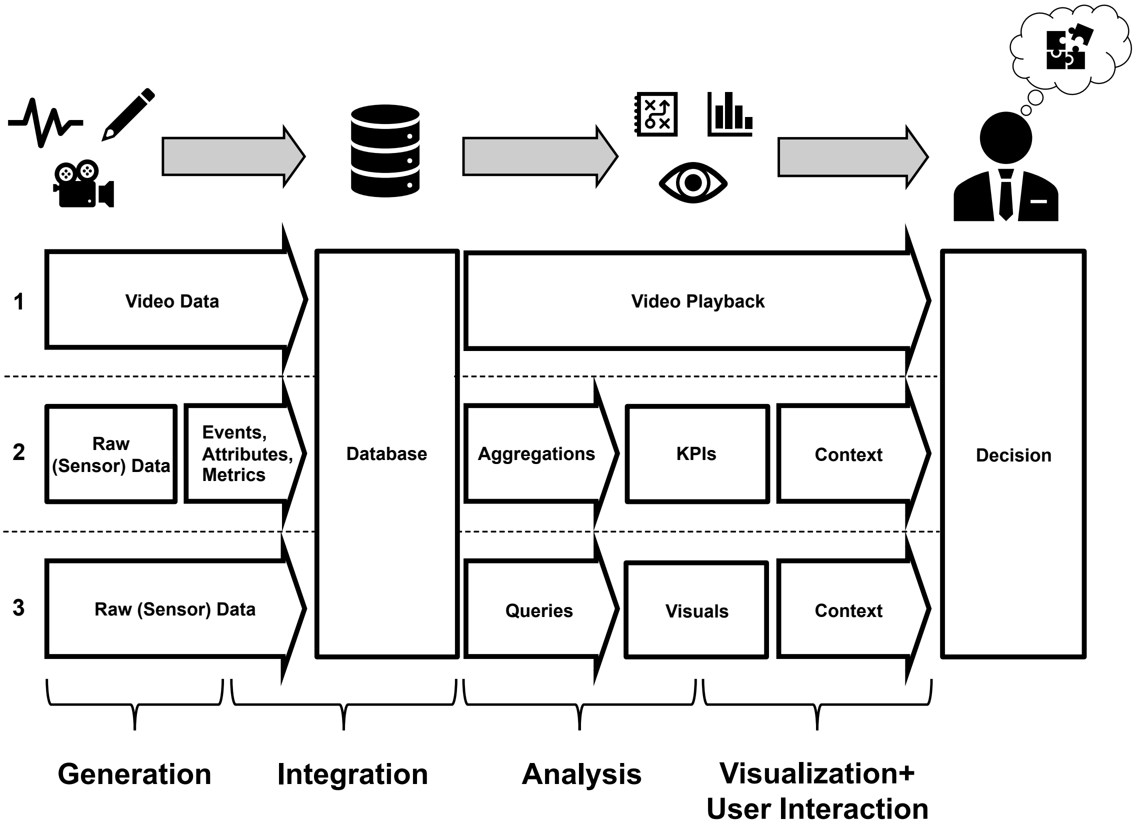


Figure 3.6 Advanced three-layered data analytics pipeline representing the three analysis options for an optimal decision support in game sports: (1) qualitative analysis, (2) quantitative analysis, and (3) visualization.

graphs, these data are contextualized and used to support the decision making process of coaches and analysts. The last pipeline at the bottom represents the analysis process of raw data (e.g., heart rate or speed data) by the coach or analyst. The results are used in a raw format or set into context via corresponding visualization options. The latter can be, for instance, a visualization of the movement trajectories based on the raw (x, y, z) coordinates of the players which are generated, e.g., with GPS sensors. Other visualization options like, e.g., line graphs are especially suitable for continuous states (see Definition 3.2) like, e.g., heart rate or speed data.

3.3 Supporting Different Query Types

In the previous section we explained that DASs ideally should combine different analysis options. However, this still is not sufficient for an optimal decision support of coaches and analysts in game sports. To allow for qualitative and quantitative analyses, as well as visualizations, DASs additionally need to support different query types to retrieve the required information.

We have seen in Section 2.2.1 that there exist diverse types of data in sports: raw data, event data, and video data. Furthermore, data for the analysis ideally should be provided in different granularities like, for instance, as aggregated statistics and as performance indicators (see Section 2.2.3). Additionally, we introduced different data types like phases and continuous states in Section 3.1.2. We also highlighted the high complexity levels of certain game sports in Section 3.2.1. These higher complexity levels lead to more complex (tactical) patterns which can occur in a sport and which of course need to be analyzed. All these points together highlight the necessity that a DAS supports different query types to allow for a holistic retrieval of all these information by coaches and analysts for their qualitative and quantitative analyses and visualizations.

In this section, we thus present query types which need to be supported by a DAS to provide an optimized decision support in game sports (see Section 3.3.1). Additionally, we present how data should be modeled to make the retrieval process as performant as possible (see Section 3.3.2).

3.3.1 Information Retrieval in Game Sports

Game sports analysts and coaches need a large amount of information based on different data and data granularities for their individual decision making processes. To retrieve all these information from a database, sport queries need to be formulated and finally executed. Before the execution of such a sport query, input parameters need to be specified by the corresponding user. This can be a set of objects (e.g., objects from a certain database collection) together with a query specification. Resulting output objects are returned after the execution of the query. In Definition 3.4 we present a simplified formal description of a sport query.

Definition 3.4 Sport query.

A *sport query* is described as a function q_{ID^*} which maps a query specification q_{sp}, applied to a collection represented by the set of all identifiers ID^*, to a set of identifiers ID, with ID being a subset of all identifiers ID^*:

$$q_{ID^*} : Q_{sp} \rightarrow \mathcal{P}(ID^*), q_{sp} \mapsto ID \text{ with } ID \subseteq ID^*$$

As described in Section 2.2.3 the context in which decisions have to be made has a great influence on the time periods which are regarded for the analysis process in game sports. Sometimes even single points in time are interesting to be analyzed for evaluating, for instance, the current level of a player's internal load. Depending on the data (i.e., the output objects) which coaches and analysts are interested in and the observed time (interval), four different query types are required for the information retrieval on players, groups of players, or whole teams: (1) raw data queries, (2) movement queries, (3) event queries, and (4) event pattern queries. Movement queries can be further distinguished between raw movement queries and in-event movement queries (see Table 3.4). In-event movement queries return movement information during certain non-atomic events, whereas event queries return general information on (atomic and non-atomic) events. Event pattern queries also return event data but of patterns where all included events have a timely dependence on each other. The resulting information of all query types allow for qualitative and quantitative analyses, as well as visualizations.

Before we present the four query types in more detail we first need to introduce sport query specifications which can be used to delimit query results. A sport query specification is a 10-tuple q_{sp} which consists of information on locations (LOC, i.e., (x,y,z) coordinates), event types ($E.K$), raw data types ($R.J$), timestamps (TS), sensor identifiers (SID), player identifiers (AID), team identifiers (TID), match identifiers (CID), and type-specific attributes ($E.A$ and $R.A$). A sport query specification can consist of none, one, or several locations, event types, raw data types, sensor identifiers, player identifiers, team identifiers, match identifiers, or type-specific attributes. Furthermore, none, one, or two timestamps (i.e., a time interval) can be specified. A formal description of a sport query specification is presented in Definition 3.5.

Definition 3.5 Sport query specification.

To delimit sport query results a *sport query specification* q_{sp} can be generated as follows:

$$\begin{aligned} q_{sp} = \; & \mathcal{P}(LOC) \times \mathcal{P}(E.K) \times \mathcal{P}(R.J) \times \mathcal{P}(SID) \times \mathcal{P}(AID) \times \mathcal{P}(TID) \times \mathcal{P}(CID) \times \\ & \mathcal{P}(E.A) \times \mathcal{P}(R.A) \times \mathcal{P}'(TS) \\ & \text{with } \mathcal{P}'(TS) \subset \mathcal{P}(TS) \text{ with at most two } TS \text{ elements} \end{aligned}$$

Table 3.4 Query types for holistic information retrieval in game sports. Depending on the output and whether a single timestamp or a time interval (i.e., a set of two timestamps) is observed, different query types are required to retrieve information on a single player p, a set of players P, or whole teams T. All query types have different inputs, query specifications, and outputs. ^{*1}: Only possible for physiological data and IMU data. ^{*2}: Only spatio-temporal tracking data. ^{*3}: Only non-atomic event data. ^{*4}: See Definition 3.10 for more details.

Query Type	Input	Query Specifications	Output	Timestamp	Interval
Raw Data	$RDID^*$	q_{sp} with $q_{sp}.E.K = \emptyset \wedge$ $q_{sp}.E.A = \emptyset$	$RDID$	p,P,T	p,P,T^{*1}
Raw Movement	$RDID^*$	q_{sp} with $q_{sp}.E.K = \emptyset \wedge$ $q_{sp}.E.A = \emptyset$	$RDID^{*2}$	p,P,T	p,P,T
In-Event Movement	$EDID^*$	q_{sp} with $q_{sp}.R.J = \emptyset \wedge$ $q_{sp}.R.A = \emptyset \wedge q_{sp}.SID = \emptyset$	$EDID^{*3}$	p,P,T	p,P,T
Event	$EDID^*$	q_{sp} with $q_{sp}.R.J = \emptyset \wedge$ $q_{sp}.R.A = \emptyset \wedge q_{sp}.SID = \emptyset$	$EDID$	p,P,T	p,P,T
Event Pattern	$EDID^*$	Q_{sp} with $q_{sp}.R.J = \emptyset \wedge$ $q_{sp}.R.A = \emptyset \wedge q_{sp}.SID = \emptyset^{*4}$	$\mathcal{P}(EDID)$	p,P,T	p,P,T

3.3.1.1 Raw Data Query

A raw data query can be used if a coach or an analyst is interested in retrieving either physiological data, IMU data, or spatio-temporal tracking data. Thus, this query type is always executed against the set of raw data item identifiers. With this query type, information on individual players, groups of players, or whole teams either at a single point in time or over a defined time interval (only physiological and IMU data) can be retrieved (see Table 3.4). This also includes the retrieval of continuous states (see Definition 3.2).

A query specification for a raw data query can contain, e.g., the raw data type. Moreover, sensors, players, teams, or matches can be specified via the corresponding unique identifiers (see Definitions 2.4 and 3.5). In some cases it might also be interesting to get raw data in a specified location (e.g., accelerations in the defending zone of the ice hockey rink). Moreover, a timestamp or a time interval, i.e., two timestamps (only physiological and IMU data) can be defined as query specification. Finally, raw data attribute values can be specified. Some of the previously mentioned options of course can be combined in a query specification. The result of this query type always is a set of raw data item identifiers. To formally describe a raw data query we present Definition 3.6. With Example 3.11 we show how a raw data query can look like in practice.

Definition 3.6 Raw data query.

A *raw data query* is described as a function $q_{r_{RDID^*}}$ which maps a query specification q_{sp} (excluding event types and event type-specific attributes), applied to a collection represented by the set of all raw data item identifiers $RDID^*$, to a set of raw data item identifiers $RDID$, with $RDID$ being a subset of all raw data item identifiers $RDID^*$:

$$q_{r_{RDID^*}} : Q_{sp} \to \mathcal{P}(RDID^*), q_{sp} \mapsto RDID$$
$$\text{with } RDID \subseteq RDID^* \wedge q_{sp}.E.K = \emptyset \wedge q_{sp}.E.A = \emptyset$$

Example 3.11 Querying physiological raw data.

To retrieve heart rate data of player $B8$ of the second half (i.e., between minutes 45 and 90) of the football match XYZ, the query specification q_{sp} can look like the following:

$$q_{sp}.R.J = \text{“}heartRate\text{”} \wedge q_{sp}.AID = \text{“}B8\text{”} \wedge q_{sp}.TS = (45, 90) \wedge q_{sp}.CID = \text{“}XYZ\text{”}$$
$$q_{sp} = (\text{“}heartRate\text{”}, \text{“}B8\text{”}, (45, 90), \text{“}XYZ\text{”})$$

After the execution of the raw data query against the set of all raw data item identifiers $RDID^*$ the resulting raw data item identifiers $RDID$ point only to raw data items which fulfil the criteria which are formulated above.

Visualizations of physiological data can help coaches and analysts to evaluate the current internal load of a player at a certain moment of the match. If a time interval is regarded the evolvement of the internal load over time (i.e., a continuous state) is an interesting point to analyze as well. The same holds for IMU data and the analysis of external loads (e.g., players' accelerations and decelerations). If spatio-temporal tracking data at a single point in time (i.e., players' locations on the playing field) are regarded this can help particularly for the evaluation of tactical player performances. The retrieval of tracking data over a time interval is covered by a raw movement query (see Section 3.3.1.2).

3.3.1.2 Movement Query

A movement query can be executed to retrieve either spatio-temporal raw tracking data (raw movement query) or non-atomic event data (in-event movement query) over a certain (user-defined) time interval. Thus, this query type is executed either against the set of raw data item identifiers or the set of event data item identifiers. With a movement query, information on single players, groups of players, or even whole teams can be retrieved (see Table 3.4). Similar to raw data queries, this includes the retrieval of continuous states (see Definition 3.2) like, e.g., running trajectories (see Section 3.1.2).

Different options exist to delimit the results of a movement query. For a raw movement query, the location (i.e., spaces or paths) of movements can be defined. In this case also sensor identifiers or raw data type-specific attributes can be defined in the query specification. For an in-event movement query the location (i.e., spaces or paths) of movements during the non-atomic events can be defined. Here, event types and event type-specific attributes can be defined as additional query specification. For both movement queries the timestamp (i.e., the starting point) or the time interval for which movements should be retrieved can be defined. Query results can further be delimited by defining player, team, or match information via the unique agent, team, or match identifiers (see Definitions 2.4, 2.6 and 3.5). To formally describe a raw movement query we introduce Definition 3.7. A formal description of an in-event movement query is presented in Definition 3.8.

Definition 3.7 Raw movement query.

A *raw movement query* is described as a function $q_{rmov_{RDID^*}}$ which maps a query specification q_{sp} (excluding event types and event type-specific attributes) applied to a collection represented by the set of all raw data item identifiers $RDID^*$ to a set of raw data item identifiers $RDID$ (of type spatio-temporal tracking data), with $RDID$ being a subset of all raw data item identifiers $RDID^*$:

$$q_{rmov_{RDID^*}} : Q_{sp} \rightarrow \mathcal{P}(RDID^*), q_{sp} \mapsto RDID$$
$$\text{with } RDID \subseteq RDID^* \wedge q_{sp}.E.K = \emptyset \wedge q_{sp}.E.A = \emptyset$$

Definition 3.8 In-event movement query.

An *in-event movement query* is described as a function $q_{emov_{EDID^*}}$ which maps a query specification q_{sp} (excluding raw data types, raw data type-specific attributes, and sensor identifiers) applied to a collection represented by the set of all event data item identifiers $EDID^*$ to a set of event data item identifiers $EDID$ (only non-atomic events), with $EDID$ being a subset of all event data item identifiers $EDID^*$:

$$q_{emov_{EDID^*}} : Q_{sp} \rightarrow \mathcal{P}(EDID^*), q_{sp} \mapsto EDID$$
$$\text{with } EDID \subseteq EDID^* \wedge q_{sp}.R.J = \emptyset \wedge q_{sp}.R.A = \emptyset \wedge q_{sp}.SID = \emptyset$$

With Examples 3.12 and 3.13 we show how movement queries can look like in sports practice.

Example 3.12 Querying spatio-temporal tracking data.

To retrieve the spatio-temporal tracking data of players $A4$ and $A5$ between minutes 5 and 10 of the first quarter of the basketball match XYZ, the query specification q_{sp} can look like the following:

$$q_{sp}.R.J = \text{“}TD\text{”} \wedge q_{sp}.AID = (\text{“}A4\text{”}, \text{“}A5\text{”}) \wedge q_{sp}.TS = (5, 10) \wedge q_{sp}.CID = \text{“}XYZ\text{”}$$
$$q_{sp} = (\text{“}TD\text{”}, (\text{“}A4\text{”}, \text{“}A5\text{”}), (5, 10), \text{“}XYZ\text{”})$$

After the execution of the raw movement query against the set of all raw data item identifiers $RDID^*$ the resulting raw data item identifiers $RDID$ point only to raw data items which fulfil the criteria which are formulated above.

Example 3.13 Querying non-atomic event data.

To retrieve all (American football) running plays of player $B3$ which fit the coach's instructions (i.e., a certain running path), the query specification q_{sp} can look like the following:

$$q_{sp}.E.K = \text{“}runningPlay\text{”} \wedge q_{sp}.AID = \text{“}B3\text{”} \wedge q_{sp}.LOC = ((x_1, y_1), ..., (x_{51}, y_{51}))$$
$$q_{sp} = (\text{“}runningPlay\text{”}, \text{“}B3\text{”}, ((x_1, y_1), ..., (x_{51}, y_{51})))$$

After the execution of the in-event movement query against the set of all event data item identifiers $EDID^*$ the resulting event data item identifiers $EDID$ point only to event data items which fulfil the criteria which are formulated above.

Movement queries are particularly interesting for tactical analyses of coaches and analysts. The results of this query type allow for visualizations of trajectories of a single player, a group of players, or even whole teams during the defined time interval. This can be pure movement visualizations (raw movement query) or movement visualizations during sport-specific events (in-event movement query). Two sample scenarios could be when a basketball coach wants to see how players behave in a certain match phase (see Example 3.12) or when an American football coach wants to see all running plays of the running back which matched the tactical instructions (see Example 3.13).

3.3.1.3 Event Query

Another query type is the so called event query which is based on event data and which is used by coaches and analysts to retrieve either a single event or multiple events. Thus, this query type is always executed against the set of event data item identifiers. Event queries are executed to retrieve information on single players, groups of players, or whole teams. Events are interesting to be analyzed at a single point in time, but also over a longer time interval. Here, we want to mention that in most but not in all sports multiple events can happen in parallel at a specific point in time. In Defense of the Ancients 2 (DotA 2), an eSports game which is played between two teams of five players, different fights between players can happen at the same time. In football, (ball-focused) events usually do not happen in parallel (e.g., only one shot can happen at a single point in time). If a certain sport-specific phase (see Definition 3.1) is delimited by two events, event queries can be used to retrieve phases as well. In this case, the start and end timestamps of the phase correspond to the timestamps of the respective events (see Section 3.1.3).

Different options exist for a query specification and can be combined to delimit the results of an event query. First, players, teams, or matches can be defined via the corresponding unique identifiers (see Definitions 2.6 and 3.5). Second, a timestamp or a time interval can be specified. Of course also the event types can be defined. Another option is to specify the location in which events should have occurred on the playing field via defining (x,y,z) coordinates. Finally, event attribute values can be specified to further delimit the query results. The result of this query type always is a set of event data item identifiers. To formally describe an event query we introduce Definition 3.9. With Example 3.14 we show how an event query can look like in sports practice.

Definition 3.9 Event query.

An *event query* is described as a function $q_{e_{EDID^*}}$ which maps a query specification q_{sp} (excluding raw data types, raw data type-specific attributes, and sensor identifiers) applied to a collection represented by the set of all event data item identifiers $EDID^*$ to a set of event data item identifiers $EDID$, with $EDID$ being a subset of all event data item identifiers $EDID^*$:

$$q_{e_{EDID^*}} : Q_{sp} \rightarrow \mathcal{P}(EDID^*), q_{sp} \mapsto EDID$$
$$\text{with } EDID \subseteq EDID^* \land q_{sp}.R.J = \emptyset \land q_{sp}.R.A = \emptyset \land q_{sp}.SID = \emptyset$$

Example 3.14 Querying event data.

To retrieve all goals which team B scored in the ice hockey matches XYZ and UVW, the query specification q_{sp} can look like the following:

$$q_{sp}.E.K = \text{“}goal\text{”} \land q_{sp}.TID = \text{“}B\text{”} \land q_{sp}.CID = (\text{“}XYZ\text{”}, \text{“}UVW\text{”})$$
$$q_{sp} = (\text{“}goal\text{”}, \text{“}B\text{”}, (\text{“}XYZ\text{”}, \text{“}UVW\text{”}))$$

After the execution of the event query against the set of all event data item identifiers $EDID^*$ the resulting event data item identifiers $EDID$ point only to event data items which fulfil the criteria which are formulated above.

Visualizing the results of an event query can give the users a good overview *when* and *what* exactly happened on the playing field and thus can help to evaluate technical and tactical player and team performances.

3.3.1.4 Event Pattern Query

In Sections 3.3.1.1 to 3.3.1.3 we presented different query types. These types can also be combined to retrieve more complex information like patterns of events, but also phases and continuous states. As previously mentioned in Section 3.3.1.3 two event queries can be combined to retrieve certain sport-specific phases (see Definition 3.1) which are delimited by events. However, a phase can, e.g., also start with a certain event and end when the value of a certain metric, i.e., a continuous state (see Definition 3.2) exceeds a defined threshold. In this case an event query and raw data or raw movement queries would need to be combined to retrieve the respective phase. In this section we present another combination in more detail, which we call an event pattern query.

In sports practice not only single events but also whole patterns of events are important to be analyzed by coaches and analysts. Therefore, DASs to support decision making in game sports should also cover event pattern queries which are, as the name already indicates, based on event data and used to search for event patterns. A pattern consists of at least two events and thus is always defined over a certain time interval. Furthermore, a pattern can involve more than just one player. Nevertheless, patterns including the whole team only occur rarely in general.

An event pattern query consists of a totally ordered set of event queries (at least two). Prerequisite of an event pattern is that the results of the second and subsequent event queries have a (user-defined) timely dependence on the respective previous event query results. For each event pattern query, this dependence can be defined either into the future (i.e., events happened after) or into the past (i.e., events happened before).

The resulting number of event data item identifiers from the last event query within the event pattern query is equal to the number of retrieved event patterns. The final result set of an event pattern query is a set of sets of event data item identifiers. For each pattern which is found, one ordered set of event data item identifiers (based on the timestamp order of the events) is contained in the final result set. Each contained set in turn consists of one event data item identifier per event query which was executed within the event pattern query.

To formally describe an event pattern query we introduce Definition 3.10. Example 3.15 shows how an event pattern query can look like in sports practice.

Definition 3.10 Event pattern query.

An *event pattern query* is described as a function $Q_{ep_{EDID^*}}$ which maps a set of query specifications Q_{sp} applied to a collection represented by the set of all event data item identifiers $EDID^*$ to a set of event data item identifiers $EDID_{ep}$, with $EDID_{ep}$ being a subset of all event data item identifiers $EDID^*$:

$$Q_{ep_{EDID^*}} : \mathcal{P}(Q_{sp}) \rightarrow \mathcal{P}(EDID^*), Q_{sp} \mapsto EDID_{ep} \text{ with } EDID_{ep} \subseteq \mathcal{P}(EDID^*)$$

An event pattern query Q_{ep} is a totally ordered set of at least two event queries (as defined in Definition 3.9) where the differences between the resulting events' timestamps of two subsequent event queries lie within a user-defined time range δ. This timely dependence is either defined into the future or into the past and holds for the whole event pattern query:

$$Q_{ep} = (Q_e, \prec) \text{ with } Q_e = \{q_{e,1}, ..., q_{e,n}\},\ n \geq\ 2$$

$$\forall\ q_e \in Q_e : \begin{cases} q_{e,i+1}.TS > q_{e,i}.TS\ \wedge q_{e,i+1}.TS - q_{e,i}.TS \leq \delta \\ q_{e,i+1}.TS < q_{e,i}.TS\ \wedge q_{e,i}.TS - q_{e,i+1}.TS \leq \delta \end{cases} \quad \text{for } i \in \{1, ..., n-1\}$$

The final result of the event pattern query is an ordered set $EDID_{ep}$ of sets of event data item identifiers. The number of sets contained in $EDID_{ep}$ is equal to the number of retrieved patterns m. The number of event data item identifiers within each set $EDID_i$ is equal to the number of executed event queries n within an event pattern query:

$$EDID_{ep} = \{EDID_1, ..., EDID_m\}$$
$$\forall\ EDID_i \in EDID_{ep} : |EDID_i| = n, \text{ for } i \in \{1, ..., m\}$$

Example 3.15 Querying patterns of events.

An event pattern query can be used to analyze how a team A creates goals and shots in a football match XYZ. The query specification q_{sp_1} of the first event query can look like the following:

$$q_{sp_1}.E.K = (\text{“}goal\text{''}, \text{“}shot\text{''}) \wedge q_{sp_1}.TID = \text{“}A\text{''} \wedge q_{sp_1}.CID = \text{“}XYZ\text{''}$$
$$q_{sp_1} = ((\text{“}goal\text{''}, \text{“}shot\text{''}), \text{“}A\text{''}, \text{“}XYZ\text{''})$$

For the second step a time range δ needs to be defined which represents the timely dependence on the previous results $EDID_1$. This can look like the following:

$$EDID_1 = \{edid1, edid3, edid12, edid28\}, \delta_{before} = 5$$
$$q_{sp_2} = (5)$$

After the execution of the second event query against the set of all event data item identifiers $EDID^*$ the resulting event data item identifiers $EDID_2$ point only to event data items which fulfil the criteria which are formulated above. In this case, for each event resulting from the first event query, the last event (if there is any) which happened within a five second time interval before that event is returned. If the result set of the second event query contains, for instance, two elements which fulfil the criteria then we would have found two event patterns:

$$EDID_2 = \{edid2, edid26\}$$
$$EDID_{ep} = \{\{edid2, edid3\}, \{edid26, edid28\}\}$$

It is of enormous value for analysts and coaches to analyze and visualize event patterns as they can, for instance, evaluate if tactical instructions were fulfilled by their players or not. Another use case for event pattern queries is the analysis of opponents preferred strategies by taking a closer look at how the team creates chances or how goals are scored (see Example 3.15) which helps a lot during a pre-match analysis.

3.3.1.5 Holistic Analysis in Game Sports

All of the previously introduced sport query types return important information for game sports analysts and coaches. Based on the query results (i.e., a set of raw data and/or event data item identifiers) qualitative and quantitative analyses, as well as visualizations are now possible.

First, it is possible to qualitatively analyze all relevant scenes in the video because the timestamps of raw data and event data items can be linked to the video (see Section 2.2.1.3). This is very important because the whole context is maintained in the video whereas information is lost through aggregating data like it is the case in quantitative analyses. For coaches and analysts it is of enormous value to qualitatively evaluate player performances during specific events, movements, patterns, or phases, respectively.

Second, through the aggregation of raw data, event data, or their corresponding attributes coaches and analysts are able to evaluate the performances of players and teams quantitatively via statistics and indicators (see Definitions 2.12 and 2.13). For coaches and analysts it is interesting to analyze, for instance, statistics over a user-defined time interval like a certain match phase (e.g., a player's total number of successful passes during build-up play). The same holds for indicators as these are based on statistics (see Definition 2.13). It also might be interesting to analyze, for instance, the value of an indicator (e.g., "dangerousity" see Section 6.1.1.1) at a specific timestamp.

Finally, data visualizations are possible to contextualize the retrieved information which is a huge support for analysts and coaches in sports practice as already described for each query type in Sections 3.3.1.1 to 3.3.1.4.

3.3.2 Data Organisation

Having clarified which query types need to be supported by a DAS for optimal decision support in game sports we now present how to model sports data in a way to ensure a performant access and retrieval process.

We have seen that the main types of sports data (i.e., raw data and event data) are structured differently (see Definitions 2.4 and 2.6). Additionally, there exists only a limited number of raw data types which, in general, are produced at a high frequency leading to a significant amount of raw data. In contrast to that, atomic and non-atomic event data occur in a high variety of sport-specific types which all (can) have different attributes. However, the total amount of event data is by far smaller compared to the amount of raw data. Therefore, it makes sense to store sports data in a document database and to introduce different collections. Each single data item then is stored as a document in a collection of the corresponding data type. The flexible, semistructured, and hierarchical nature of documents [AWS21] fits perfectly the demands of analyzing sports data as this allows, e.g., to query documents with a certain attribute or value.

In addition to raw data and event data, general information on, e.g., players, teams, or matches, so called metadata, should also be stored in a separate collection of the database. Here, also the paths to the respective video files need to be stored as video data themselves will not be stored in the database. We consequently propose the following three data collections: (1) raw data collection, (2) event data collection, and (3) metadata collection. All query types are executed against at least one of these collections.

In the remainder of this section we introduce each of the three collections in more detail.

3.3.2.1 Raw Data Collection

The *raw data* collection contains physiological data, IMU data, as well as spatio-temporal tracking data. Each raw data item is stored as a separate document in the collection and contains the following information: raw data item identifier, sensor identifier, competition or match identifier, agent identifier, team identifier, timestamp, raw data type, and raw data type-specific attributes (see Definition 2.4). If, for instance, spatio-temporal tracking data are captured with a sampling rate of 25 Hz, this results in 25 data items for each second and player which are stored in the raw data collection. This means that all the positions of all players during one or more matches are stored in this collection. In sports which are using, e.g., a ball as playing equipment, the positional data of the ball are stored in this collection as well. A sample document of the raw data collection is shown in Figure 3.7.

Raw data queries are executed against this collection. Additionally, raw movement queries can be executed against this collection to retrieve tracking data which in turn allows the visualization of movements (trajectories).

```
{
      "rdid" : "562219",
      "sensorId" : "HR25",
      "matchId" : "127398",
      "playerId" : "A5",
      "teamId" : "A",
      "ts" : 2102,
      "type" : "heartRate",
      "heartRateDetails" : {
            "beatsPerMinute" : 154,
            "aboveCriticalThreshold" : false
       }
}
```


Figure 3.7 Heart rate data item stored in the raw data collection.


```
{
      "edid" : "236791",
      "matchId" : "127398",
      "playerIds" : ["B3","B5"],
      "teamIds" : ["B"],
      "ts" : 22125,
      "type" : "successfulPass",
      "xyzCoordinates" : [ [-10.93,-10.40,0.00],[9.23,2.18,0.00] ],
      "passDetails" : {
            "packing" : 3,
            "length": 23.76,
            "direction" : "forward"
       }
}
```


Figure 3.8 Successful pass event data item stored in the event data collection.

3.3.2.2 Event Data Collection

The *event data* collection contains all event data items. Each data item is stored as a separate document in the collection and contains the following information: event data item identifier, competition or match identifier, agent identifier(s), team identifier(s), timestamp, event type, spatial coordinates, and event type-specific attributes (see Definition 2.6). When we take ice hockey as an example the collection would contain events like passes, shots, goals, dumps, shifts, or stickhandlings, just to name a few. A sample document of the event data collection is shown in Figure 3.8.

Event queries and event pattern queries are two query types which are executed against the event data collection. Additionally, in-event movement queries can be executed against this collection to retrieve and visualize movements during non-atomic events like, for instance, a dribbling event in football.

```
{
     "matchId" :  "127398",
     "date" :  "2021-02-21T20:15:00Z",
     "venue" :  "Fantasy Club Arena",
     "homeTeamName" :  "Team A",
     "awayTeamName" :  "Team B",
     "homeTeamId" :  "A",
     "awayTeamId" :  "B",
     "homeTeamColor" :  "red",
     "awayTeamColor" :  "blue",
     "homePlayerNames" :  ["Player A1","Player A2","Player A3",
                           "Player A4","Player A5"],
     "awayPlayerNames" :  ["Player B1","Player B2","Player B3",
                           "Player B4","Player B5"],
     "homePlayerIds" :  ["A1","A2","A3","A4","A5"],
     "awayPlayerIds" :  ["B1","B2","B3","B4","B5"],
     "videoPath" :  "TeamAvsTeamB.mp4"
}
```

Figure 3.9 Metadata item stored in the metadata collection.

3.3.2.3 Metadata Collection

The *metadata* collection holds all metadata. Each metadata item is stored as a separate document in the collection and contains general information and details on a match or competition. Each metadata item contains a unique competition or match identifier. Additionally, information about date and venue of the match, names, identifiers, and jersey colors of both teams, as well as player names and identifiers should be contained in the metadata items. Furthermore, the video path(s) to the corresponding match needs to be stored in each metadata item. A sample document of the metadata collection is shown in Figure 3.9.

Each document in one of the other collections contains a property which points to the corresponding metadata item (e.g., `matchId` in Figures 3.7 and 3.8).

3.4 Planning and Designing the User Experience

With the concepts presented so far in Sections 3.1 to 3.3 we highlighted very important points for the development of DASs for decision making in game sports. However, there is one last point missing which is crucial for an optimal decision support: the UX of DASs. Coaches and analysts need to be provided with the option to intuitively query and analyze data to quickly and easily find answers to their specific questions. In this section we thus present an approach how to design a DAS in a way that allows exactly that. With such a system the working efficiency in sports practice will be maximized which is one of the major requirements of performance analysts (see Section 1.2).

In this section we mainly focus on the data visualization and user interaction step of the data analytics pipeline with a special emphasis on how this step can optimize decision support for coaches and analysts through a user-centered planning and designing of the UX of a DAS (see Figure 3.10). Therefore, based on the five-layer model of Garrett [Gar12] which we introduced in Section 2.2.4.2, we present five planning steps and their results which should be considered for the development of a DAS to provide an optimal decision support in game sports.

3.4.1 Strategy

The strategy layer is the first and most important layer because all subsequent layers and decisions are based on it (see Section 2.2.4.2). In this section, we introduce five questions and answers to illustrate the process during this planning step. The first two question concern the product objectives whereas the subsequent questions (3. to 5.) concern the user needs. We summarize the results of this planning step at the end of this section.

1. Why is there a need for DASs to support decision making in game sports? This question has sufficiently been answered in Chapter 1 and Chapter 2 and represents the main motivation of this thesis. To quickly recapitulate some of the key points, game sports are very complex and a lot of factors are influencing the decision making processes of coaches and analysts. The latter ones are also exposed to an enormous time pressure in certain phases of a season and also during matches in which (tactical) decisions have to be made very fast. Additionally, a huge amount of data of different sources and formats are available for the analysis process. A DAS can facilitate this tedious and time-intensive task of coaches and analysts.

2. In which context should a DAS be applied? We introduce our approach in the context of match planning. Therefore, the DAS should also be applied in this context which in turn means that the DAS is important for the support of coaches and analysts in the following three decision making contexts: (1) pre-match analysis, (2) live-match analysis, and (3) post-match analysis (see Section 2.2.3).

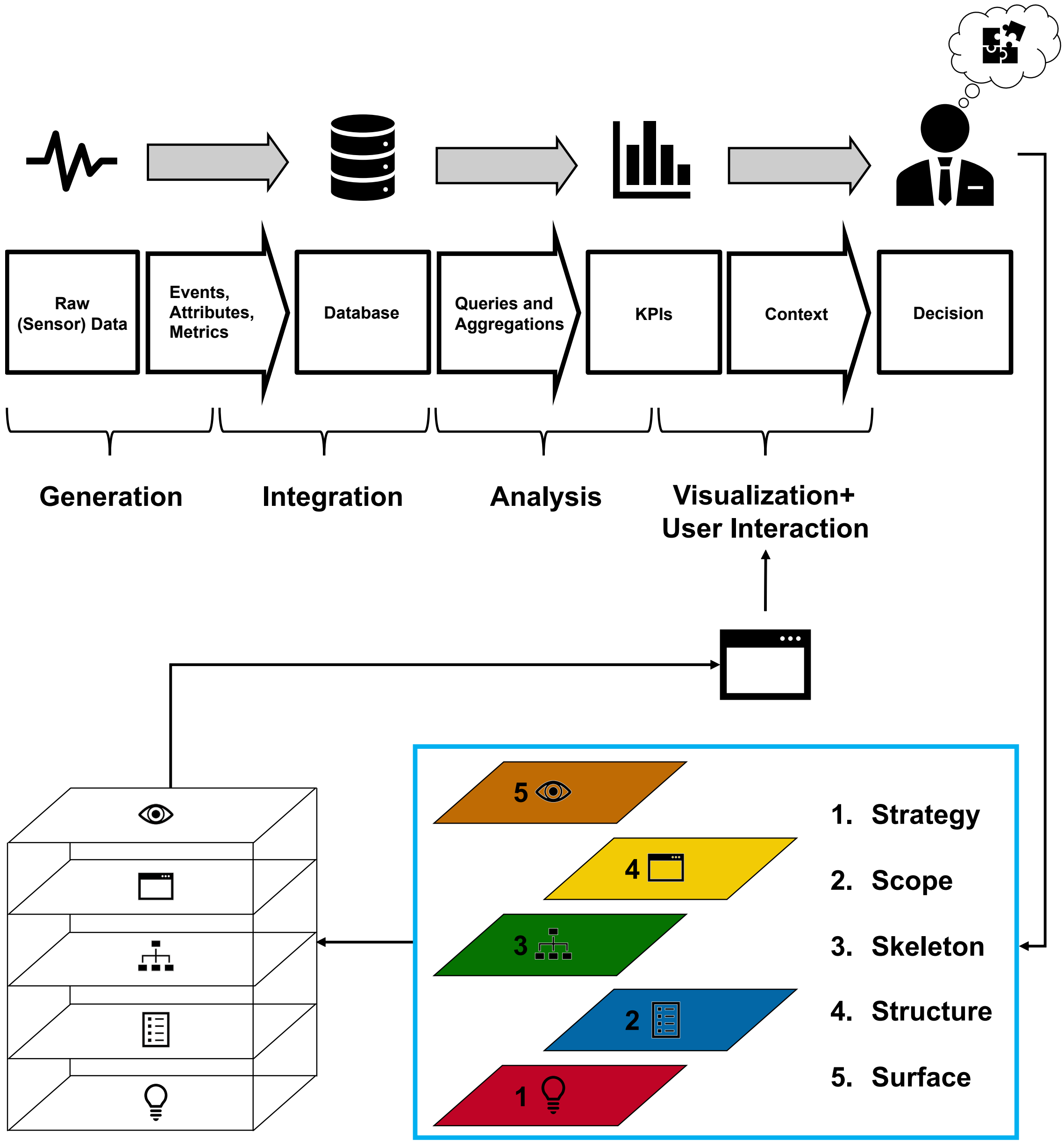


Figure 3.10 User-centered design of the UX of DASs for game sports. The five layers (blue box) according to the model of Garrett [Gar12] influence the data visualization and user interaction step of the data analytics pipeline.

3. Who is the target group? We want to support and facilitate the work of game sports analysts and coaches.

4. What are the interests of the target group? To really understand what the interests of the target group are it is necessary to do research like, for instance, in the form of interviews. In Section 3.1 we introduce results from the interviews we conducted with six UEFA Pro License football coaches. Despite these interviews were conducted in the concrete context of football, some of the results are

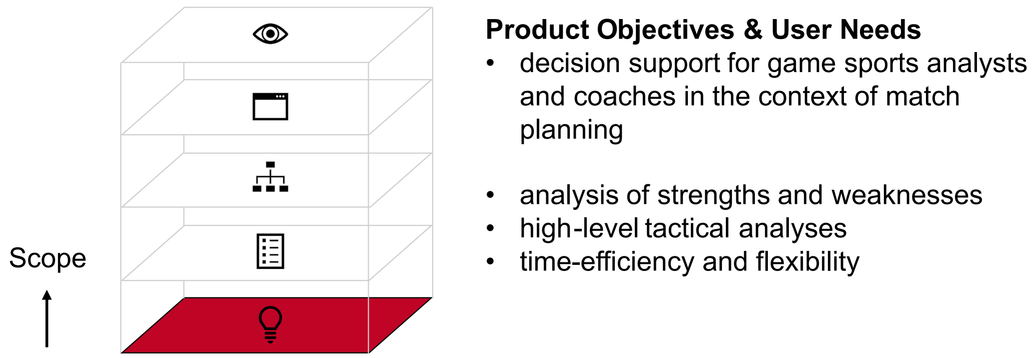


Figure 3.11 Results of the "strategy" planning step for the UX design of a DAS to support decision making in game sports.

nevertheless transferable to most other game sports. Having analyzed the interviews we can state that the following point is particularly important for coaches when they prepare for an upcoming match. They want to analyze the strengths and the weaknesses of the next opponent in order to optimize and adapt the tactics of the own team. This also includes high-level tactical analyses.

5. What is the benefit of a DAS for the user? The coaches and analysts who work with the DAS are able to make their decisions easier and faster. They can do their analyses time-efficiently and flexibly.

Results. The product objective of the DAS is decision support for game sports analysts and coaches in the context of match planning. Therefore, users need options to analyze strengths and weaknesses of opponents including high-level tactical analyses. This should be possible in a time-efficient and flexible way. The results of this planning step are depicted in Figure 3.11.

3.4.2 Scope

Having clarified the question "Why we want to develop a DAS to support decision making in game sports" in the previous section we now focus on the question "What exactly do we want and need to develop" to reach the goals we have defined in the strategy layer (see Section 2.2.4.2). For that we need to answer further questions concerning the functional specifications (1.) as well as the content requirements (2.). Of course the user is again in the centre of the planning. Therefore, we have to think especially about the coaches' and analysts' needs. The results of this planning step are summarized at the end of this section.

1. Which functionalities does the DAS need to provide in order to support decision making in game sports? This question is already answered to a certain extent in Sections 3.2 and 3.3. To analyze strengths and weaknesses of the opponent and of the own team it is essential to perform qualitative and quantitative analyses. Additionally, data visualizations are required which further support the decision making process and help users in making (more) use of the data. Coaches and analysts consequently need to query and analyze raw data, events, and patterns of events, to search for and watch video sequences of important scenes, and also to flexibly generate statistics and indicators on-the-fly. Therefore, the DAS should provide these different functionalities.

2. Which content do coaches and analysts need and should be provided by the DAS? Coaches require information from various sources and of different granularities (see Section 2.2.3). The video is one important part which should be contained in the DAS. Furthermore, the DAS needs to provide access to raw data, event data, as well as to aggregated statistics and performance indicators or even KPIs. Of course, these data should represent the coaches' concepts and mental models. To ensure the latter we presented an approach in Section 3.1. This then allows analysts and coaches to perform high-level tactical analyses as well. Additionally, it is important to analyze the performances of the own and of the opponent's team while taking several past matches into account. This means that historical data need to be provided by the DAS as well.

Results. Functional specifications of the DAS contain the querying, analysis, and visualization of raw data, events, and patterns of events, the video playback of important scenes, and the flexible generation of statistics. The DAS needs to provide content in the form of actual and historical video data, raw data, event data, aggregated statistics, and performance indicators. We highlight the results of the "scope" planning step in Figure 3.12.

3.4.3 Skeleton

With the third layer the development and design of the DAS gets more concrete (see Section 2.2.4.2). In the following, the interaction of the users with the DAS is focused. We need to find out how coaches and analysts will interact with the system and which consequences will arise for the development of the DAS. The so called interaction design is depicted with the first three questions (1. to 3.). Additionally, it is important how to structure the information in a way that

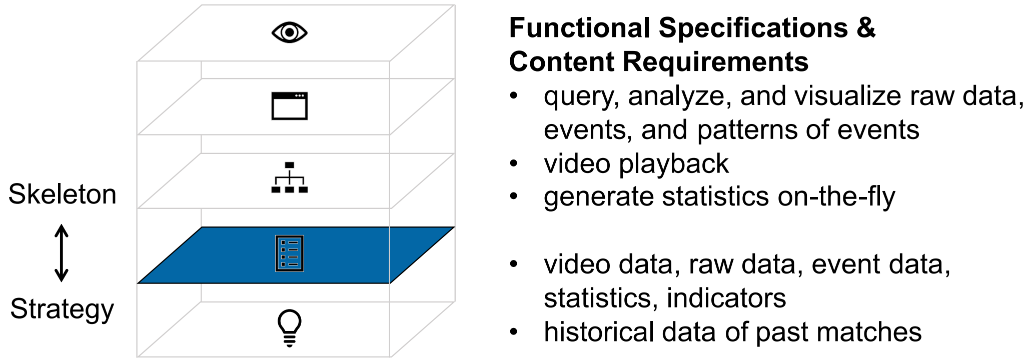


Figure 3.12 Results of the "scope" planning step for the UX design of a DAS to support decision making in game sports.

coaches and analysts will profit from the DAS. The information architecture is covered by the last two questions (4. and 5.). We summarize the results of the "skeleton" planning step at the end of this section.

1. How do coaches and analysts want to search for (patterns of) events? Coaches and analysts mostly know *where* and *what* happened on the playing field. What they often do not know is *when* exactly specific events happened. The easiest way to find events or patterns of events with the DAS is that the system provides dedicated filter options. In game sports, coaches often make use of a tactics board to visualize certain elements of the match plan (e.g., running trajectories) and to demonstrate these to the players. To check whether or not certain tactics have been implemented by the team it would be helpful if coaches and analysts can search for events and patterns of events intuitively with the DAS by means of drawing sketches of these patterns. This would correspond to the use of a tactics board in practice. Therefore, DAS should provide the users with the option to query events and patterns of events not only via filters, but also via sketch-based functionalities. Moreover, when coaches found certain events or patterns of events it would be a great benefit to directly watch the corresponding scene(s) in the video. Consequently, the DAS should also provide an option which makes that possible.

2. How do coaches and analysts use statistics? From the interviews which we conducted with coaches there was a noticeable consensus about the use of statistics. Statistics are mainly used as a supplement to confirm the coach's subjective impressions during a match or from the qualitative video analysis. An additional point was that coaches do not want to use and rely on fixed (static) reports with statistics and indicators, but that they would like to carry out their own quantitative analyses flexibly and to dynamically generate statistics on-the-fly. The DAS should thus be able to provide coaches with this option by allowing coaches to save their queries and subsequently compare them with respect to different parameters.

3. What are the consequences for the interaction design of the DAS? For coaches it is important to have all information visualized in the DAS in a compact manner. Because game sports take always place in time and space the DAS need to cover several interactive components for the information retrieval, but also for the analysis and visualization of the results. One component is a timeline, which allows to directly see *when* certain events or patterns happened. A canvas-element representing the playing field is another important component because coaches need to know *where* exactly an event occurred. The field can thus be used to visualize results. Furthermore, the canvas-element should also be usable for the previously mentioned and highly required intuitive sketch-based information retrieval. A video player to visualize the video (scenes) should be another component of the DAS. The options to filter for events, save queries, and generate statistics on-the-fly should be part of the DAS as well as the option to perform high-level tactical analyses. All these components need to be covered by the DAS.

4. How can information be presented in a way that coaches and analysts easily understand it? All descriptions and labels which will be implemented in the UI of the DAS must correspond to the coaches' jargon to not (again) enlarge the semantic gap of sports analytics (see Section 2.1.3). In Section 3.1 we have seen that four different data types are required to represent the different concepts of coaches. Furthermore, each type of sports data need to be visualized differently as already described in Section 2.2.4. This means that the DAS needs to cover various visualization options for the different data types to contextualize the information. This is required to make data understandable and thus optimizes the decision support for coaches and analysts.

5. How can the DAS support coaches and analysts with historical data? Because coaches and analysts often take several past matches into account for their analysis process it is important for the DAS to be able to access a database with historical data.

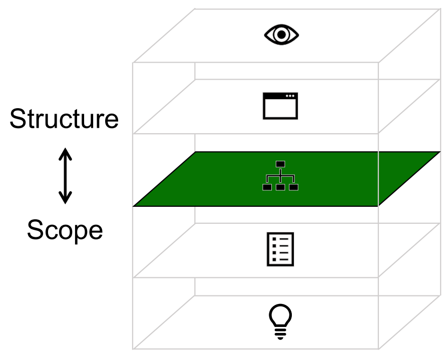


Figure 3.13 Results of the "skeleton" planning step for the UX design of a DAS to support decision making in game sports.

Results. The DAS needs to provide the following components for the interaction: a timeline, a canvas-element representing a playing field, and a video player. Additionally the user should be able to filter, save, and analyze queries with the system. For the information architecture of the DAS three things are important. First, visualizations need to be data type-specific. Second, the labels and descriptions in the UI of the DAS should use a sport-specific jargon. Third, the DAS needs access to a database with historical data. The results of this planning step are presented in Figure 3.13.

3.4.4 Structure

The specific functional requirements have been fixed and the necessary interactive elements are defined. In this step we will now focus on these individual elements of the DAS. Therefore, we cover questions on the interface design (1. and 2.), the navigation design (3.), and the information design (4.) as introduced in Section 2.2.4.2. The results are summarized at the end of this section.

1. How can coaches and analysts (intuitively) interact with the DAS and which functionalities need to be available? The most intuitive way for game sports analysts and coaches to search for events or movements is a sketch-based approach because of the similarity to a physical tactics board. Therefore, the users need to be provided with the option to draw areas or motion paths on the canvas-element (playing field) of the DAS. Because it is also important to search for patterns of events some advanced functionalities to query events happened either before or after a certain event need to be provided. This again should be realized with sketches on the canvas-element. For the sketching of motion paths users can select either straight paths or freehand paths. All these different functionalities can be selected via a dropdown menu. To define the shape of the sketches for event or pattern detection, coaches and analysts can choose between rectangular, circular, or free shapes via buttons. Two buttons to undo the last step of the analysis and to clear the whole analysis need to be provided as well. Furthermore, users are able to select and apply certain filters via dropdown menus, checkboxes, and/or sliders to further delimit the search. These filters include matches, events, teams, players, and time periods. The results are then displayed either on the canvas-element, on the timeline, or even on both components. Clicking on one of the results will allow the users to directly watch the respective video scene. This allows a very time-efficient working. Additionally, it is possible to watch the video without any preceding analysis for a pure (traditional) video analysis. The DAS needs to provide a save button to save the current analysis for a later reuse like, for instance, in team meetings or meetings with individual players. The saved queries are displayed in a scrollable list. Quantitative analyses can be conducted by clicking on one of three buttons which decide on the level of analysis: team, player, or query analysis. One or multiple parameters can subsequently be selected from a list. To analyze and compare one or several teams, players, matches, or specific queries according to the previously selected parameters the user needs to click on a button. To perform high-level tactical analyses various options which represent coaches' concepts can be selected from dropdown menus either on a team level or on a player level.

2. How need the different interactive elements (components) of the DAS to be structured and arranged? It is important to not overwhelm coaches by implementing too many functionalities like it is often the case in commercial products. The DAS thus focuses on the main functionalities which are required for an efficient decision support in game sports. This in turn results in a compact design and manageable number of functionalities and windows. The canvas-element (playing field) and the video player need to be arranged next to each other. This will then facilitate the comparison between the 2D-representation of the results and the "reality" in the video. The dropdowns and buttons for the various sketch-based retrieval methods should be arranged next to the canvas-element. Because the timeline is an additional source of information it should also play a serious role and thus should be arranged below the video player and the canvas-element. The filters for the query specification should be placed visibly next to the playing field because the analysis results are mostly displayed there. Because the quantitative analysis and the high-level tactical analysis are also important features of the DAS the corresponding buttons should be placed visibly for the user in the center of the screen. The save button should be placed next to query results, which means that it will be arranged together with the filter dropdowns. The list of saved queries should also be presented visibly next to the video player and the canvas-element to allow coaches and analysts a fast and time-efficient re-analysis for the meetings. The lists of teams, players, matches, queries, and parameters for quantitative analyses should be arranged next to each other for a clear structure and overview for the user.

3. How should coaches and analysts navigate within the DAS to make optimal use of the functionalities? To navigate within the DAS the timeline needs to support a zooming functionality. This is important because in game sports lots of events happen either at the same time or quickly after another. The zooming will allow to clearly see which event(s) happened at a certain point in time. Hovering over single results will give the user additional information. The video player needs to provide classical functionalities to pause, play, fast-forward, and rewind the video. The quantitative analysis should happen separately in a new window to ensure a good overview and to not confuse users with other information displayed in the DAS.

4. How should information be presented to coaches and analysts? To facilitate the understanding of analysis results, the information need to be presented in an appropriate manner. Atomic events could be represented as single dots on the timeline or as dots and, where it makes sense, as arrows (e.g., successful passes) on the canvas-element. This is easily understandable and coaches and analysts would know *where* and *when* specific events happened. Phases can also be presented on the timeline via bars to highlight the time span covered by each phase. Continuous states require different visualization options like, for instance, in the form of line graphs. It would make sense to also visualize these graphs on the timeline. Profiles are highly individual and could be visualized among other options in the form of heatmaps, tables, or bar charts. This depends on each individual profile and must be handled separately in the sport-specific contexts. A quantitative analysis is best supported using bar charts or line graphs. This is a good way to detect trends at first glance.

Results. The canvas-element of the DAS needs to support a sketch-based information retrieval. All the different interaction elements and components of the DAS need to be arranged in a compact and clear form in the UI. The timeline

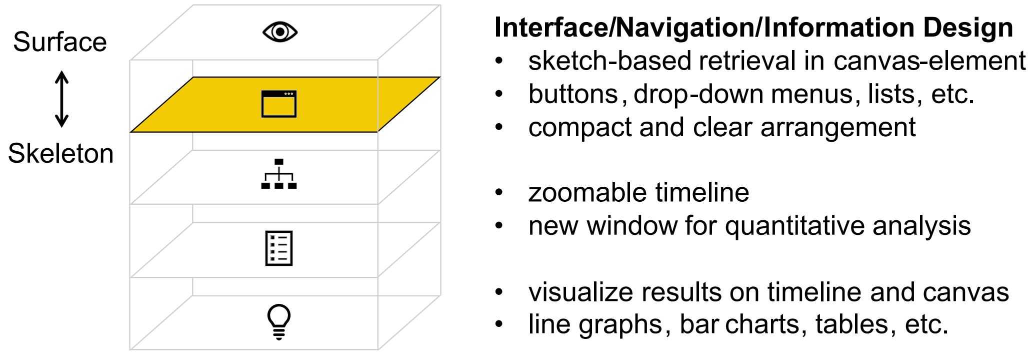


Figure 3.14 Results of the "structure" planning step for the UX design of a DAS to support decision making in game sports.

should be zoomable and a new window for the quantitative analysis would be useful for a good navigation. Results should be visualized on the timeline and on the canvas-element. Line graphs, bar charts, tables, and further visualization options are required for an appropriate presentation of information. The results of this planning step are displayed in Figure 3.14.

3.4.5 Surface

This layer makes the sensory design of the DAS a subject of discussion. For the development and design of the DAS to support decision making in game sports only two senses are of major relevance and should be particularly focused in the last planning step: hearing (1.) and, even more important, seeing (2.). Afterwards, we present the results of the "surface" planning step. To conclude this section we present a DAS for ice hockey as a concrete example of how a DAS for game sports can look like after the consideration of all planning steps (see Example 3.16 and Figure 3.16).

1. Which acoustical information are important and helpful for coaches and analysts when working with the DAS? At first, simple system sounds are helpful to facilitate the interaction with the DAS. This can help, for example, if input parameters are missing during a quantitative analysis. Acoustic alerts can then be applied to make coaches and analysts aware of the missing input parameters. Second, and even more important is the acoustic information contained in the video. The video is the most important source of information (see Section 1.2) because all context information are maintained. However, this context is only then complete if the sound of the video is available as well. Only then the atmosphere (e.g., coach instructions, fan chants, player shouts) of the current situation can be evaluated and also be taken into account when analyzing the performance of the team or player.

2. Which visual information are important and helpful for coaches and analysts when working with the DAS? One of the most important things for coaches to work with technical systems is an adequate representation of elements of sports practice. This concerns functionality but also design. The first step to implement the DAS in such a way is the background image of the canvas-element. This image should be selected to fit the respective sports playing field in reality. This sounds obvious but it is absolutely necessary for the DAS to not get too abstract as coaches want to stay as close to sports practice as possible. The video player and the playing field are the two eye catchers and key components of the DAS. Therefore, these two components need to be prominently designed and presented. This can be realized by applying a grey background color for the DAS as a contrast to highlight the two main components. For the timeline as the third very important component of the DAS some visual details need to be considered for the design as well. The exact time at which certain events happened is very important for coaches. Therefore, the minutes of play need to be appropriately visualized directly on the timeline. Moreover, this is also important when zooming into the timeline. The minutes should then be visualized more precisely and, where and if necessary, also the seconds. The sketches on the canvas-element need to be depicted with a good contrast to the image of the playing field. The same holds for the dots and arrows on the playing field and/or on the timeline. This facilitates the processing of information by coaches and analysts. Visualized movements (trajectories) in the canvas-element in team colors would further support an intuitive understanding. Moreover, the bars in the timeline, which represent the phases should be designed in the corresponding team colors as well to allow a fast and intuitive identification. The same holds for continuous states and profiles which are defined on a team-level. Important buttons for undo-, clear-, and save-functionalities need to be highlighted in traditional colors to underline their importance and to quickly find them in the UI of the DAS. Finally, the quantitative analysis window needs to be well structured with clear lists, and an analysis button to start the analysis process. The results should then be displayed in different colors to facilitate the comparison between players or teams.

Results. Acoustical information in the form of alerts and warnings as well as video sounds should be provided by the DAS. Concerning the visual informa-

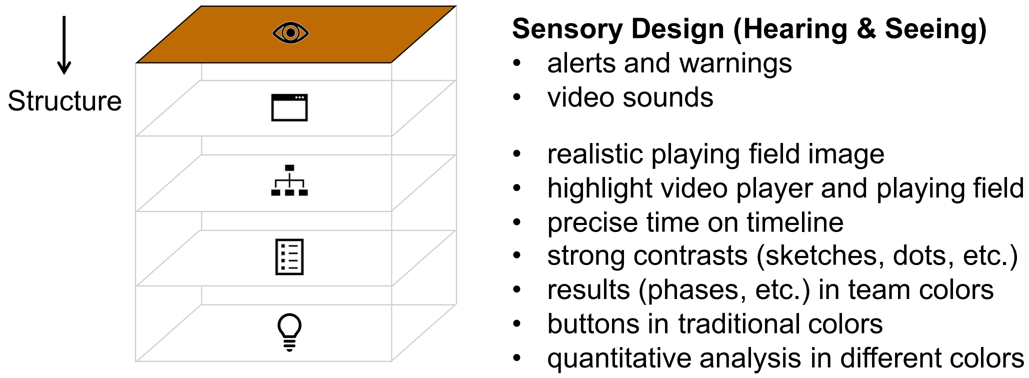


Figure 3.15 Results of the "surface" planning step for the UX design of a DAS to support decision making in game sports.

tion which the DAS should provide the following points are important: the playing field image needs to be realistic, important components (e.g., playing field and video player) need to be prominently presented, the visualization of time on the timeline needs to be precise, strong contrasts need to be chosen, results should be displayed in team colors, important buttons need to be highlighted in colors, and different colors should also be used for quantitative analyses. The results of the "surface" planning step are depicted in Figure 3.15. To show an example of how a DAS for game sports can look like after consideration of all planning steps we present Example 3.16 together with a UI-mockup shown in Figure 3.16

Example 3.16 A DAS to support decision making in ice hockey.

Figure 3.16 represents a mockup of the UI of a DAS for the invasion game ice hockey. We can see how all planning steps have influenced the design of the system and how all components are arranged in a way to optimally support ice hockey analysts and coaches in their analysis process. The video player and the canvas-element (ice hockey rink) are prominently displayed next to each other. The results of the sketch-based query (dots in black rectangle) are visualized directly on the canvas-element. The timeline is visualized at the bottom and also presents the analysis results (dots) clearly. The filter area is placed next to the canvas-element as well as the buttons for the query specification and the undo-, clear-, and save buttons. The quantitative analysis and high-level tactical analysis functionalities can be started via the buttons in the center of the UI. Saved analyses are visualized in a list below the video player.

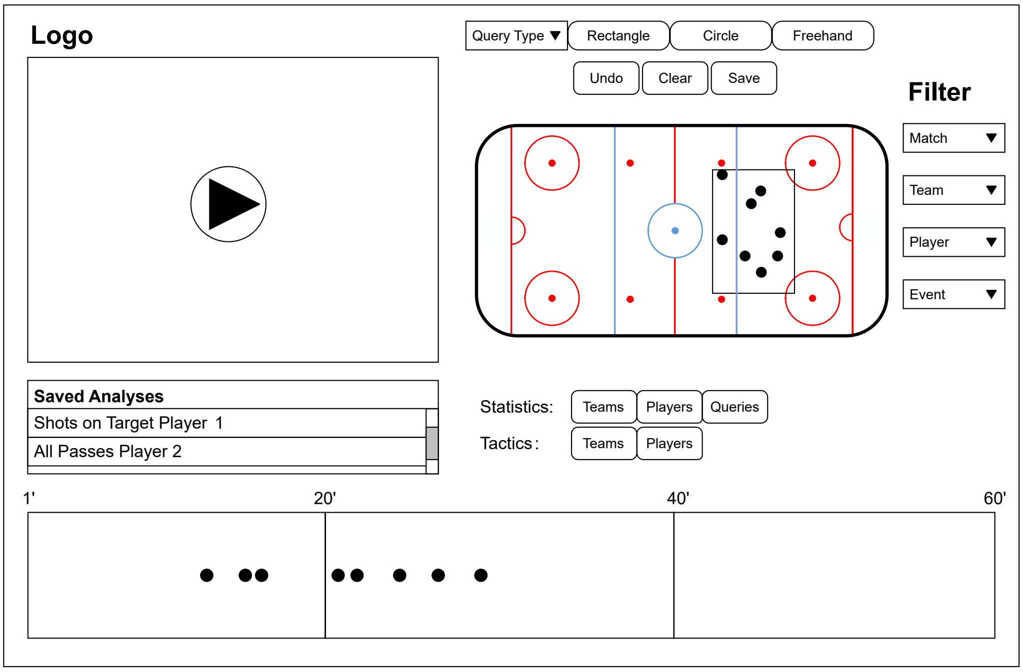


Figure 3.16 UI-mockup of a DAS to support decision making in ice hockey.

4

Die Technologien, die es im Fußball gibt, werden irgendwann Aussagen und Rückschlüsse über die Taktik zulassen müssen und nicht nur über athletische Werte, damit man nach dem Spiel überprüfen kann, was von den Vorgaben umgesetzt werden konnte. Das wird ein großer Entwicklungsschritt.

— Julian Nagelsmann

The SportSense System

Having seen in Chapter 3 how DASs need to be conceptualized and designed in order to support decision making in game sports we now present a specific implementation, the DAS SportSense.

We first introduce the generic architecture of the system (see Section 4.1) for which we considered the concepts presented in Sections 3.2 to 3.4. Then, we show two applications of the SportSense system in game sports or, more precisely, in invasion games: SportSense Football (see Section 4.2) and SportSense Ice Hockey (see Section 4.3). We present these systems together with the sport-specific challenges and requirements and highlight how we implemented the results from Section 3.1. The two applications underline the potential lying in the generic architecture of SportSense and highlight the flexibility to easily adapt the system to other game sports. All implementations which we present in this chapter are published on GitHub under the GNU Affero General Public License v3.0 (see Appendix B). Because of the open source publication of the system some of the libraries required for the **sportsense-web-client** repository are not published on GitHub but replaced with web sources. More details can be found in the repository in question.

4.1 Generic Architecture

The generic architecture of the SportSense system consists of three components: (1) a MongoDB as database instance, (2) a MongoDB REST proxy, and (3) a Web client. With the Web client the user can define the queries for the analysis process, for instance, by drawing sketches on the canvas-element of the Web client UI (see Section 4.1.3.3). The Representational State Transfer (REST) API,

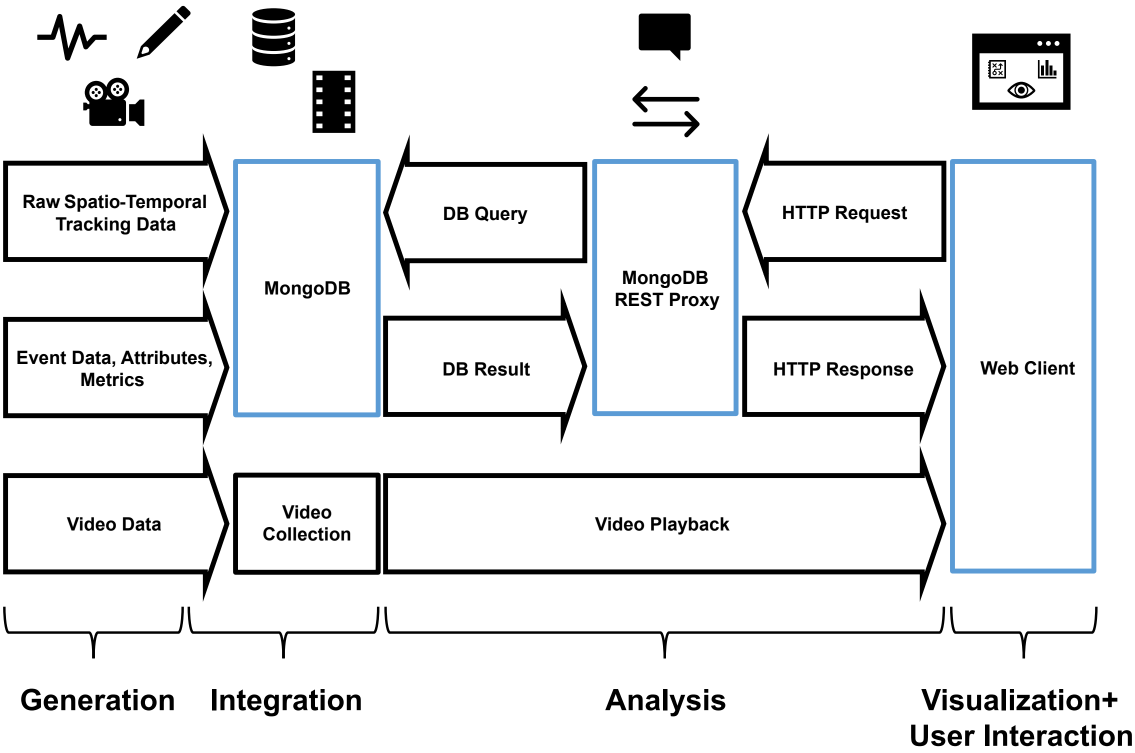


Figure 4.1 Generic architecture of the SportSense system.
SportSense consists of three main components (highlighted in blue): MongoDB, MongoDB REST proxy, and Web client.

the MongoDB REST proxy, translates the Hypertext Transfer Protocol (HTTP) requests into database queries and communicates them to the database. The database results are then translated by the MongoDB REST proxy into HTTP responses which in turn are sent back to the the Web client. There, the results get adequately visualized in the SportSense UI. The generic architecture of SportSense is depicted in Figure 4.1. Each of the three components and their corresponding implementation is described in detail in the following.

4.1.1 MongoDB

The SportSense system is based on a MongoDB instance. MongoDB is a document database (as proposed in Section 3.3.2) and stores data in JSON-like documents [Mon21a]. This means that fields can vary from document to document and data structure can be changed over time [Mon21b]. The decision to take a MongoDB instance for our implementation is based on two reasons: first, MongoDB supports all spatial features SportSense requires and second, it provides good performance and scalability characteristics [Pro20]. Gunawan et al. [GRD19] showed in their experiments that MongoDB has very low average query response times for read, update, and delete processes. For our implementation we used MongoDB version 4.4.1. In the following we present which types of data are needed and stored in the MongoDB to enable a full functionality of the SportSense system (see Section 4.1.1.1). Additionally, we show how data are modeled in different collections within the MongoDB (see Section 4.1.1.2).

4.1.1.1 Data

For a full functionality of SportSense, which allows an optimal decision support for coaches and analysts in game sports, different data are required and thus need to be stored in the MongoDB. First, spatio-temporal raw tracking data in the form of (x, y) or even (x, y, z) coordinates are needed to analyze and visualize the movements of the players and, for instance, of the ball[1]. Second, event data with corresponding attributes are required to depict the actions of a match and to be able to generate statistics on-the-fly based on these data. Optionally, other sport-specific metrics which represent (tactical) performances can be stored in the MongoDB, as well as some predefined (fixed) statistics (e.g., ball possession in football). Of course, analyses which are generated with SportSense and saved, e.g., for later meetings with the team need to be stored in the MongoDB as well. Finally, there is the need to store metadata on match details like, for instance, player and team names, the time and the date of the match, or even the team or jersey colors. Video data as an additional source of information is not stored in the MongoDB, but separately in a video collection (see Figure 4.1). This video collection in turn is accessed directly by the Web client of the SportSense system (see Section 4.1.3.1).

SportSense is optimized to work with data generated and stored in a MongoDB instance by StreamTeam [PRS+18; Pro20]. StreamTeam is a real-time data stream analysis system with applications in team sports. Both systems, StreamTeam and SportSense, originated from the same research project and together (can) form an integrated analysis infrastructure [PRS+18; Pro20]. Nevertheless, SportSense can also be used independently as a standalone system and work with data from other sources like, for instance, from commercial providers like ChyronHego or Opta (see Section 2.2.1). This is possible without any major changes.

[1] Functionalities based on other types of raw data like, e.g., physiological data or IMU data are currently not supported in SportSense as our main focus was the tactical performance which can be better analyzed with spatio-temporal tracking data.

4.1.1.2 Collections

The data which we introduced in the previous section need to be adequately structured and modeled for an optimized and performant access and retrieval process with SportSense (see Section 3.3.2). Therefore, data are stored in six different collections: (1) events, (2) nonatomicEvents, (3) matches, (4) states, (5) statistics, and (6) savedFilters. While the first five collections are resulting from the analysis process with the StreamTeam system, the *savedFilters* collection is created and used by SportSense. Again we want to mention that the analysis of a match with StreamTeam is not essential for working with SportSense. Nevertheless, the database schema remains the same even if data comes from other (commercial) providers. This means, that data from other providers first need to be integrated into the existing collections. This can be done by performing a short transformation step of the data to fit the database schema which SportSense relies on. If this step is successful, users can profit from the full potential of the SportSense system. The schemas of the five collections created by StreamTeam are described in detail in the work of Probst [Pro20]. In the following we briefly introduce all collections and give some examples. Furthermore the full schema of the *savedFilters* collection is introduced together with a sample data item.

Events. The *events* collection corresponds to the event data collection we introduced in Section 3.3.2.2 but contains only event data items of atomic events (e.g., clearances, shots, successful passes). For each atomic event one data item (i.e., a binary JSON document) exists in the collection. The data items contain general information like, for instance, timestamp, spatial coordinates, and references to the players which are involved in this event. Additionally, (optional) attributes are stored which can differ between event types. Therefore, these attributes are stored in a separate property (`additionalInfo`) of the data item which can be filled with an arbitrary JSON object [Pro20].

Non-Atomic Events. The *nonatomicEvents* collection holds all event data items of the non-atomic events and is created by the StreamTeam system. This differentiation between event and non-atomic event collection is required because the non-atomic events which are detected by StreamTeam have a different structure compared to our definition. Each non-atomic event consists of several data items which are stored in the collection. This makes a grouping of data items necessary for a later analysis. To allow that, additional properties (`eventId`, `phase`, and `seqNo`) are stored in the corresponding data items [Pro20]. All event data items of the non-atomic events are stored in the *nonatomicEvents* collection.

Matches. The *matches* collection corresponds to the metadata collection we introduced in Section 3.3.2.3. Here, all metadata of matches are stored. For every match only one data item exists in the collection.

States. The *states* collection corresponds to the raw data collection we introduced in Section 3.3.2.1. It contains, for instance, spatio-temporal tracking data. The collection has the same schema as the *events* collection which means that the data items contain some general properties like, e.g., timestamp, (x, y, z) coordinates, and references to the respective players. Similar to the *events* collection, (optional) additional information can be stored in the `additionalInfo` property of each data item. In the case of tracking data this information can be the current player or ball velocity. Furthermore, other sport-specific metrics are stored in this collection. In the case of football this can be, for instance, the pressing state (see Section 4.2.2.1).

Statistics. As the name already indicates this collection contains (fixed) statistics[2]. If matches are analyzed with StreamTeam the collection even contains the evolvement of corresponding statistics during the matches. This is because StreamTeam analyzes matches in real-time and updates all statistics either periodically in a certain interval or when the statistics change due to a certain event. Each update is stored as a new data item in the *statistics* collection. The schema of the collection is similar to the one of the *events* collection and the *states* collection. If StreamTeam data are not used, fixed statistics like, e.g., ball possession in football can be stored in this collection.

Saved Filters. The *savedFilters* collection is used for the saved analyses a user conducted with the SportSense system (see Section 4.1.3.3). Each query which the user wants to reuse later is stored in this collection. However, SportSense only stores the query parameters in the database, but does not materialize the results we would get by executing the query. This prevents that data get duplicated [SJR+19]. For each query one data item exists in the savedFilters collection. The data items consist of three properties, namely `name`, `eventIds`, and `sport`. The name of the query which is specified by the user is contained in the `name`

[2] For the on-the-fly generation of own statistics data from this collection are not used, but from the events and non-atomic events collections.

```
properties: {
      name: {
            description: "Name of the saved query (formulated by the user)",
            bsonType: "string" },
      eventIds: {
            description: "Array containing all event data item identifiers of
                          the query results",
            bsonType: "array",
            items: { bsonType: "string" } },
      sport: {
            description: "Sport for which the query was formulated",
            bsonType: "string" }
},
required: ["name", "eventIds", "sport"]
```

(a) Schema of the savedFilters collection.

```
{
      "_id" : ObjectId("600ac2ad2a77a05ca26a02a1"),
      "name" : "Shots on Target - 1st Half",
      "eventIds": [
            "5ed76276e6b91f776c9c48d0",
            "5ed76393e6b91f776ca1d5ba",
            "5ed76592e6b91f776caba998"
      ],
      "sport" : "football"
}
```

(b) Sample data item stored in the savedFilters collection.

Figure 4.2 Schema and sample data item of the savedFilters collection of the MongoDB.

property. The `eventIds` property contains all the query results, i.e., the event data item identifiers of the saved query. Finally, the `sport` property contains the name of the sport for which the query was formulated. This is important to allow sport-specific analyses with SportSense. The schema of the savedFilters collection is shown in Figure 4.2(a). A separate property to which match(es) a query belongs to (i.e., `matchIds`) is not necessary as this information is contained indirectly through the individual event data items (see Figure 3.8). We furthermore present a specific example of a data item of the savedFilters collection in Figure 4.2(b).

4.1.2 MongoDB REST Proxy

The MongoDB REST proxy serves as an interface between the Web client and the previously introduced MongoDB (see Figure 4.1). It is used by the Web client to retrieve the required data for the analysis from the MongoDB. Subsequently, the proxy sends the results back to the Web client. The communication between the Web client and the MongoDB is a central element of the whole SportSense system. Each core functionality of SportSense uses the communication channels of the MongoDB REST proxy. Therefore, we present this component of the SportSense system in more detail in this section. First, we introduce implementation details in Section 4.1.2.1. Second, we present the central communication aspects of the MongoDB REST proxy with the database and with the Web client in Section 4.1.2.2.

4.1.2.1 Implementation Details

At first we briefly highlight some general information on the implementation of the MongoDB REST proxy. The programming language used for the development is Java. To receive the requests and to respond to the requests of the Web client the MongoDB REST proxy uses the Jetty[3] HTTP Web server. The reason for choosing the Jetty HTTP Web server is, amongst others, its flexibility, the small size, and that the components are open source [Fou21] which makes it easily embeddable in our SportSense system. For querying the MongoDB the Java MongoDB Driver[4] is used.

4.1.2.2 Communication with MongoDB and Web Client

There are two classes which are of major relevance for the communication of the REST proxy with the MongoDB and the Web client: the `RequestHandler` class and the `RestResponse` class. The `RequestHandler` class handles all the incoming requests from the Web client. Each request of the Web client contains an analysis method (e.g., a certain feature) and needs to be explicitly formulated in the corresponding query (see Section 4.1.3.2). This analysis method must also be specified in the `RequestHandler` class to allow for an appropriate handling of the request. Depending on the analysis method, a corresponding method in the `RestResponse` class is called subsequently. Here, the MongoDB gets queried. The database results are returned to the proxy in the form of a JSON object. This JSON object then gets parsed into a string in the `RestResponse` class and is returned to the `RequestHandler` class. Finally, the string is sent to the Web client by the `RequestHandler` class. There it gets further processed (see Section 4.1.3.2).

[3] https://www.eclipse.org/jetty/
[4] https://docs.mongodb.com/drivers/java

4.1.3 Web Client

The Web client is the component which contains all main features for the analysis and which provides the UI of the SportSense system. It combines qualitative and quantitative analyses and visualizations. Additionally, the Web client supports different query types for an intuitive and holistic information retrieval. For the UX design of the SportSense Web client we made use of the concepts presented in Section 3.4. To access the required data from the MongoDB the Web client needs to communicate with the MongoDB REST proxy. In this section we introduce the Web client component of SportSense in detail. First, we present implementation details in Section 4.1.3.1. Second, we show some insights into the communication with the MongoDB REST proxy in Section 4.1.3.2. Finally, we show generic features for the analysis of game sports with the SportSense system in Section 4.1.3.3.

4.1.3.1 Implementation details

The Web client is written in the programming languages Typescript, HTML, and CSS. When performing analyses with the SportSense system, the Web client does not directly access the MongoDB to retrieve the data, but indirectly via sending the requests to the MongoDB REST proxy (see Section 4.1.2.2). The Web client uses Asynchronous JavaScript and XML (AJAX) to create HTTP GET requests and to send them to the MongoDB REST proxy. AJAX combines different technologies in the AJAX model [Doc21] to allow user interactions with an application to happen asynchronously, i.e., independent of communication with the server [Gar05]. Even if the X in AJAX stands for XML, also JSON is used for packaging information in the AJAX model [Doc21]. AJAX allows an application like SportSense to make quick, incremental updates to the UI without reloading the entire browser page, which in turn makes the application faster and more responsive to user actions [Doc21]. In contrast to the other data, videos of corresponding matches are accessed directly from a local video collection outside the MongoDB (see Figure 4.1).

4.1.3.2 Communication with MongoDB REST Proxy

The most important class for the communication with the MongoDB REST proxy is the `DBConnection` class. This class is used to send all the requests to the MongoDB REST proxy to get the required data. The most important method within the `DBConnection` class is the `sendQuery` method which uses AJAX (see

Section 4.1.3.1) to create HTTP requests (XMLHttpRequest) and to finally call the MongoDB REST proxy. Example 4.1 shows how such a request can look like. Specific analysis methods (e.g., features of SportSense) need to be explicitly specified and passed as a parameter with the `sendQuery` method (e.g., “getAreaEvents” in Example 4.1). This is important to allow the MongoDB REST proxy an appropriate handling of the incoming requests (see Section 4.1.2.2). After receiving the response from the MongoDB REST proxy, the results are forwarded by the `sendQuery` method to corresponding methods for further processing like, e.g., data visualizations.

Example 4.1 Request sent by the Web client to the MongoDB REST proxy.

If a user searches for events in a defined rectangular area, for instance, with an event query (see Section 4.1.3.3) the respective HTTP GET request could look like the following:

http://localhost:2222/getAreaEvents?shape=rectangle&coordinates=
{%22bottomLeftX%22:%2237.81%22,%22bottomLeftY%22:%22-1.87%22,
%22upperRightX%22:%22-16.78%22,%22upperRightY%22:%22-21.70%22}
&eventFilters={}&teamFilters={%22filter%22:A}&playerFilters={}&period
Filters={%22filter%22:F}&timeFilter={}&sportFilter={%27sport%27:football}&
matchFilters={%22match%22:%22279384%22}

The method of the query is defined first and is in this case “getAreaEvents”. Query specifications follow after the “?”. We can see that besides the shape of the rectangle, which is described via the coordinates, other specifications are made by the user like, e.g., a specific match and team.

We illustrate the communication process within the SportSense system by means of some other brief examples. First, we show how the process of saving a query works. Second, we highlight the process of rerunning or deleting queries. Third, we present the communication workflow for the quantitative query analysis feature (see Section 4.1.3.3).

After saving an event query in the UI of SportSense (see Section 4.1.3.3) the Web client sends a request with all event data item identifiers (i.e., the query results) as well as the user-defined query name to the MongoDB REST proxy. The proxy then creates a document and inserts it into the *savedFilters* collection of the MongoDB (see Section 4.1.1.2).

The MongoDB document identifier of the saved query is stored as a parameter in the corresponding list item of the saved query in the UI. This is required to know which query should get deleted or re-executed. Whenever the `Show` or `Delete` button (see Figure 4.3) of a certain query is clicked, the Web client sends a request with the identifier of the saved query to the MongoDB REST proxy [SJR+19] which then queries the database or deletes the document from the collection.

The last example concerns quantitative analyses. After the user has selected one or more queries for the analysis and clicked the `Analyze` button (see Figure 4.6(b)) the Web client sends a request to the MongoDB REST proxy with the query identifiers of all queries that are required for the analysis (i.e., which are selected by the user). After querying the MongoDB, the MongoDB REST proxy returns the data to the Web client where the results finally get visualized.

4.1.3.3 Generic Features

In the following, we show how the Web client's UI is designed. Furthermore, we present how the functional specifications formulated in Section 3.4.2 are implemented in the SportSense system and which generic features are available for the analysis of game sports.

Based on the results of Section 3.4.3 the UI of SportSense consists of three main components: (1) a timeline, (2) a canvas-element (drawing area), and (3) a video player (see Figure 4.3).

The timeline is used to display results of a user's query. For the implementation we made use of the vis.js[5] library. The timeline represents a customizable time period, e.g., the whole match, a period of a match, or just a few minutes of a match. The time period can be shifted by clicking and dragging and can also be extended (zoom in) or reduced (zoom out) by using the mouse wheel [SJR+19] to allow a more detailed look at the results (see Section 3.4.4). Additional information is shown when users hover over individual results. Clicking on one of the results will directly show the corresponding scene in the video player.

The canvas-element (drawing area) is the second important component of the UI. Its background image is a 2D-representation of the sport-specific playing field as recommended in Section 3.4.5. As described as a crucial requirement in Section 3.4.4 the drawing area can be used to define queries in a sketch-based manner, e.g., by drawing regions on it. To enable the sketch-based functionalities

[5] `https://visjs.org/`

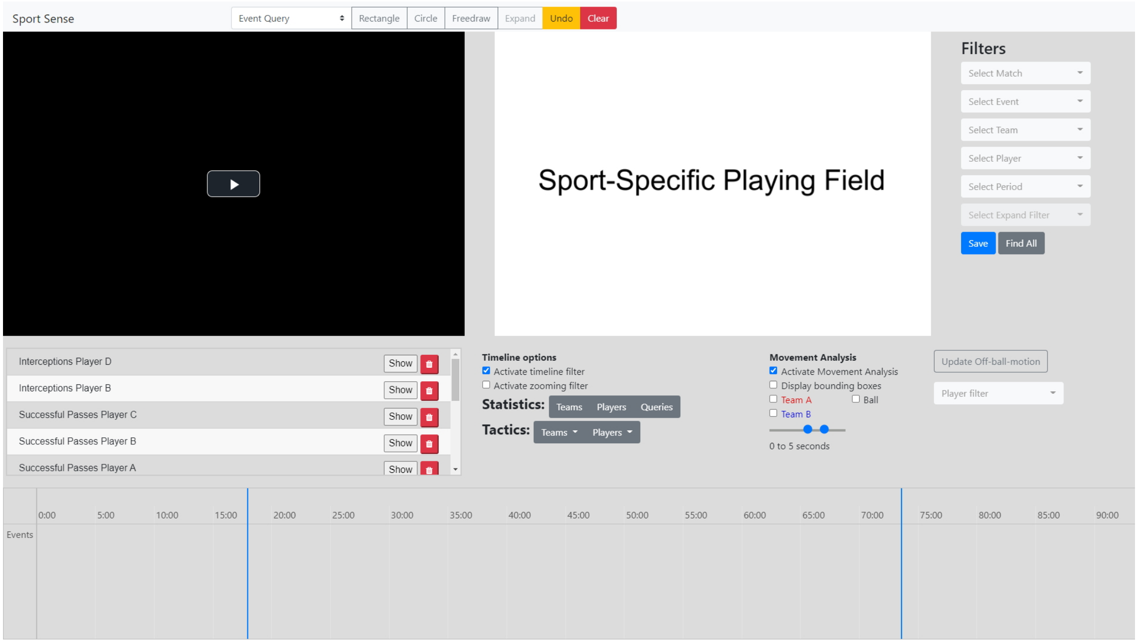


Figure 4.3 Generic UI of the SportSense Web client.
The UI consists of three main components: video player, drawing area (playing field), and timeline. Additionally, there is a filter area to further delimit the search. Saved queries are displayed below the video player. Below the drawing area additional features can be activated via checkboxes. Buttons for quantitative and high-level tactical analyses are located there as well. The UX design is based on the concepts presented in Section 3.4.

we made use of the fabric.js[6] library. Furthermore, the drawing area is also used to display query results in addition to the timeline. This allows to not only see *when* certain events happened but also *where*. Similar to the results which are displayed on the timeline, a click on one of the results on the drawing area leads to the video playback of the respective scene.

The third main component is the video player which supports classic functionalities to pause, play, fast-forward, and rewind the video (see Section 3.4.4). We made use of the video.js[7] library for the implementation of the video player.

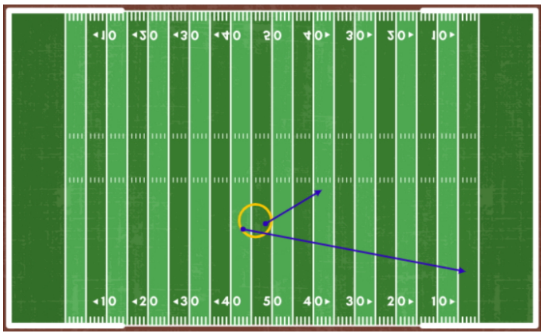
In the following, we depict in detail how users can interact with the system and which additional features are provided by SportSense.

Event Query. SportSense supports event queries (see Definition 3.9) as presented in Section 3.3.1.3. To search for events users can draw sketches on the drawing area [AS13b; PAL+18; SJR+19; SJP+20]. Three different options (buttons) are available to define the shape of the sketch: `Rectangle`, `Circle`, and `Freedraw`. While the first two options are fixed shapes where the user only can

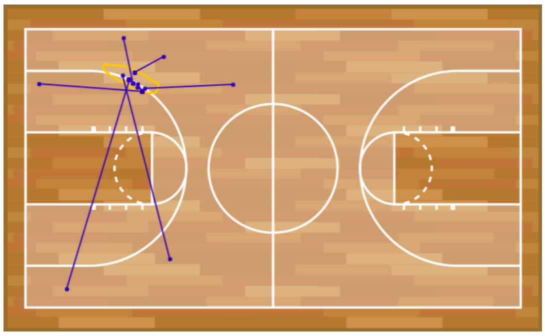
[6] `https://github.com/fabricjs/fabric.js`
[7] `https://videojs.com/`

change the size, the `Freedraw` option allows the user to define a completely individual free-form shape (see Section 3.4.4). Having selected one of the options by clicking the corresponding button, the user can draw a region, in which the events of interest have occurred, on the drawing area. Before the drawing and to further delimit the search, a user can optionally apply one or several filters as query specification (e.g., selecting specific matches, event types, teams, players, or time periods) via dropdown menus in the filter area (see Figure 4.3). Additionally, if the user does not want to delimit the region via a sketch, the `Find All` button can be clicked after the (optional) application of filters. This has the effect that the whole playing field is taken into consideration for the query. The query results are displayed on the drawing area as dots and arrows and on the timeline as dots as recommended in Section 3.4.4. In Figure 4.4(a) the results of a sample event query are displayed on an American football pitch. Here, the user selected a circular shape (orange circle) as query specification.

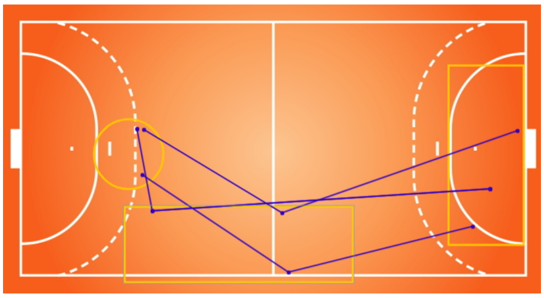
Event Cascades. For coaches and analysts it is important to search not only for single events, but also for patterns of events (see Section 1.2 and Section 3.4.2). Therefore, SportSense supports two kinds of event pattern queries (see Definition 3.10), so called event cascades. If coaches are interested in patterns in which one or several events happened **after** a certain event then the so called *Forward Event Cascade* [AS13b; PAL+18; SJR+19] can be applied. This feature is particularly interesting when searching for known patterns like, for instance, specific plays in American football or set pieces in football which have been practiced during training. The user has to define a time range in which two events have a timely dependence. This is currently implemented with five seconds. The user can define a sketch-based event query (with optional filters) and subsequently define another area (a second event query). Only patterns which fulfil the criteria (i.e., events included in the second shape must be happened within the defined time range after an event included in the first shape) are displayed as results in the drawing area. The workflow of a forward cascade can be repeated any number of times [SJR+19]. A sample forward event cascade is shown in Figure 4.4(c). In this figure, a certain offensive pattern in handball is analyzed. The first rectangle is drawn around the right goal area where the attack starts with a pass from the goalkeeper. The second rectangle is drawn on the left flank of the field. For the offensive pattern, a pass from this position into the attacking center needs to be played. Thus, a third shape in the form of a circle is sketched around the 7m-line. The results (blue connected lines) represent three scenes in



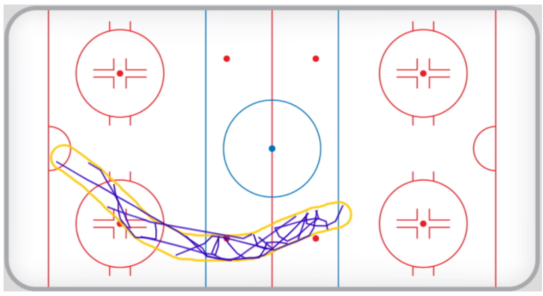
(a) Event query.



(b) Reverse event cascade.



(c) Forward event cascade.



(d) Freehand motion path.

Figure 4.4 Four different sketch-based retrieval methods of SportSense.

the match where this offensive pattern was performed by the team.

The opposite functionality of a forward event cascade is a *Reverse Event Cascade* [AS14; PAL+18; SJR+19]. This functionality is used when users are interested in events happened **before** one or several events. One use case can be if coaches want to know how a team creates chances. To execute a reverse event cascade, the user needs to first define a normal event query. Then, clicking on the `Expand` button (see Figure 4.3) will visualize all selected events together with their preceding event [SJR+19], which again has to lie within a defined time range (currently implemented with five seconds). This workflow can also be repeated any number of times. A sample reverse event cascade is depicted in Figure 4.4(b). Here, the emergence of three-point field goals in basketball is analyzed. In this example a free-form shape around a specific area of the three-point line is drawn with an applied filter to get only the goals originated from that area. Then the `Expand` button is clicked which led to the results as depicted in the figure. Each goal is displayed together with its preceding event (blue lines).

Motion Paths. Not only patterns of events, but also movement patterns, based on spatio-temporal tracking data, are of major interest for game sports analysts and coaches. To give an example, it can be interesting in American football if a certain play was performed by the team as intended or not, i.e., if all players kept their running trajectories as planned. Therefore, SportSense supports raw movement queries (see Definition 3.7) in order to search for certain motion paths (i.e., trajectories) in a sketch-based manner [AS13a; AS13b; PAL+18; SJR+19]. Highlighted as requirement in Section 3.4.4, users can select either the *Straight Motion Path* or the *Freehand Motion Path* option to sketch a straight path or an individual free-form polygon, respectively, on the drawing area of the UI. Additionally, users can define the thickness of the path via a slider. Trajectories which fit the drawn path are displayed as lines in the drawing area. Additionally, users can choose whether specific player trajectories or, where possible, for instance, trajectories of the ball should be displayed. In Figure 4.4(d) we can see an example of a freehand motion path (orange polygon) in ice hockey. All trajectories of the puck which match the drawn path are displayed as clickable blue lines.

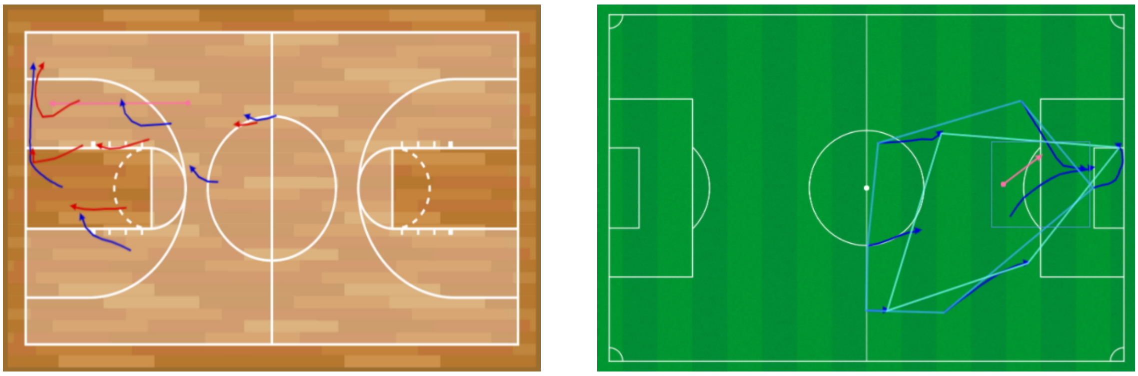
Timeline Filter. In addition to the very coarse time filter (only periods can be selected, e.g., the four quarters in basketball) which can be set for event queries, event cascades, and motion paths we implemented a *Timeline Filter* in SportSense [SJR+19]. This additional filter is needed because, often in game sports, coaches implement specific tactics only for a certain time interval. In football, for instance, an often applied tactical approach is to press very high during the first five to ten minutes of a match. The performance during this first minutes might differ fundamentally from the rest of the match and thus needs to be analyzed separately. Therefore, it makes sense to specify a query more precisely by an individually defined time interval. By activating the filter with a checkbox, users are able to manually adjust draggable (blue) bars on the timeline which set the range of time for the results of the query (see Figure 4.3). The results are dynamically updated each time either the minimum or the maximum slider value changes [SJR+19].

Zooming Filter. In general, a lot of events occur during the course of a match in game sports. If the analysis with the SportSense system is not enough delimited, for instance, by applying certain filters, the result list might get very long. Consequently, the visualization of the results in the drawing area and on the timeline quickly gets unclear and confusing, at the latest if several matches are analyzed at the same time. Thus, we implemented the *Zooming Filter* [SJR+19]. The idea of this filter is to define events with different priorities, i.e., levels of importance. There might be events, which occur very often during a match like, e.g., simple passes. These passes are not important on a match level, but more interesting when analyzing a certain scene of a match, e.g., a specific attacking pattern. On the other side, there are events which are of major importance and which should be visible for the coach in the UI at first glance like, e.g., touchdowns in American football or goals scored in ice hockey. If the filter is activated with a checkbox, the timeline only shows events with a high priority while all other events get greyed out. Zooming into the timeline will then make events visible which have a lower priority. This can be compared to the functionality of Google Maps[8], where first the name of a city is displayed, but zooming deeper into the map will subsequently make also street names visible. Each event can be categorized into a certain level of importance in a configuration file. The number of levels can also be defined individually in the same file. This is extremely useful to individualize SportSense to the personal needs of a coach or analyst [SJR+19].

Save Queries. In sports practice coaches and analysts perform quite a lot of analyses. Some of these analyses are useful to be shown to the staff or the players, for instance, in team meetings to prepare for an upcoming match. For time reasons it would be very useful to save analyses which means that there is no need to do the analyses again during the meeting. Therefore, SportSense supports a feature to save event-based queries for a later reuse (see Section 3.4.4). Event and event cascade queries can be saved and named after clicking the `Save` button. The name of the saved query appears as item in a list below the video player (see Figure 4.3). Clicking on the `Show` button will re-run a query without any further effort. Of course, queries which are no longer needed can be deleted from the MongoDB as well. For that, a `Delete` button in the form of a recycling bin next to the `Show` button can be clicked. Saved queries can also be used for a comparative quantitative analysis on-the-fly [SJR+19] as shown later in this section.

Movement Analysis. An important element of tactical performance analysis considers (off-ball) movements of players in specific situations [SJP+20]. Therefore, SportSense supports another option to execute a raw movement query (see Definition 3.7). After the execution of an event query, one event during which the movements are of particular interest needs to be clicked. Having activated the *Movement Analysis* via a checkbox a user can visualize the trajecto-

[8] `https://www.google.de/maps/`



(a) Player trajectories. (b) Bounding boxes.

Figure 4.5 Movement analysis features of SportSense.

ries of both teams, individual players, or the ball – if available – either through checkboxes (teams, ball) or via a dropdown menu (players) as depicted in Figure 4.3. A slider allows the user to specify the time interval for which the movement trajectories are displayed. Additionally, the bounding boxes, which give a first impression about a team's compactness, can be displayed for each team via checking the respective checkbox. The bounding boxes delineate the convex hulls of all (selected) players of a team on the playing field at the beginning and at the end of the investigated time interval. A sample movement analysis is shown in Figure 4.5(a). Here, the trajectories of the five players per team in basketball can be analyzed. In Figure 4.5(b) we can see two bounding boxes of the blue team's defensive players in football: first, the light blue bounding box, which represents the team's compactness at the beginning of the investigated scene which in turn started with a specific event (pink arrow) and second, the cyan colored bounding box at the end of the analyzed time interval. The evolvement of these boxes gives a coach or an analyst a first impression about the team's defensive performance in a certain scene.

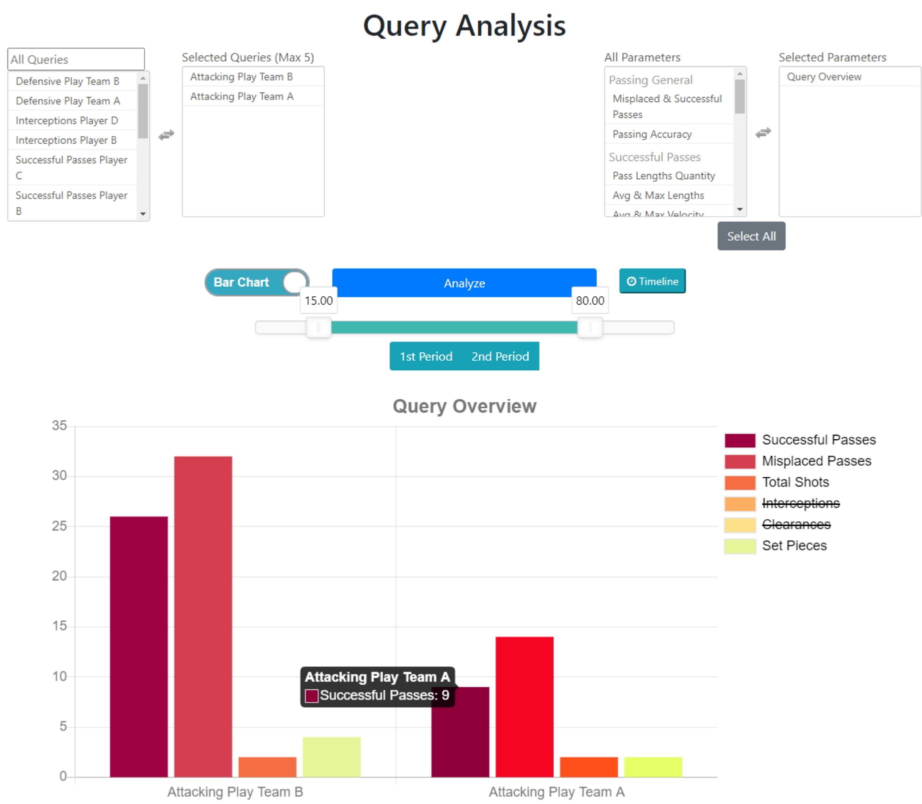
Quantitative Analysis. All previously introduced features serve to visualize data in different ways. But, as we presented in Section 3.2.2, a DAS for optimized decision support in game sports needs to provide quantitative and qualitative analyses as well. In this paragraph we present how SportSense supports quantitative analyses. A quantitative analysis (see Definition 2.2) is important on different levels: on a player level, on a team level, and as we have seen in Section 3.4.3 also on a coaches individual level, i.e., own queries should be comparable quantitatively on-the-fly. All these levels are supported by SportSense. A user can select the level of analysis by clicking on one of three buttons: `Players`, `Teams`, or `Queries` (see Figure 4.3). While the analyses on a player and team level are build in the same way, the query analysis is slightly different.

When the `Players` or `Teams` button is clicked a new tab opens. This is important to ensure a good overview (see Section 3.4.4). Here, the user can select players or teams, respectively, as well as matches and compare them with respect to one or several performance parameters which in turn consist of statistics and indicators. Sample results of a quantitative player analysis are shown in Figure 4.6(a). For the implementation we made use of the multiselect.js[9] library which allows that players, teams, matches, and parameters can be selected from a list and are then represented in another list after the selection. This in turn allows a good overview on the current selections. To execute the analysis a user has to click the `Analyze` button (see Figure 4.6(a)). For the visualization of the results we made use of the jQuery[10] library. Bar charts in the form of colored squares are created for every selected player (or team) and parameter. Here, different colors are used to facilitate the comparison between players or teams (see Section 3.4.5). For each parameter, the player or the team with the highest value gets a bar chart filled to the top with a primary color. The height of the primary color in all other bar charts in a row are calculated with respect to the difference to the highest value of each row [SJR+19] (see Figure 4.6(a)). This allows a simple and clear comparison between the results.

Clicking the `Queries` button also opens a new tab which now allows highly customizable, comparative quantitative analyses as described in Section 3.4.2. Here, the user selects one or several of the previously created (and saved) queries and can compare them on-the-fly with respect to one or several performance parameters (see Figure 4.6(b)). Of course a comparison only makes sense between saved queries which contain the same parameters and/or which return the same types of events [SJR+19]. This type of analysis makes analysts and coaches independent of fixed statistics like they often occur in the form of commercial match reports (see Section 6.2.3). We again used the multiselect.js library for the implementation of this quantitative analysis feature. The results are either visualized as bar charts or as line graphs which can be selected via a toggle next to the `Analyze` button (see Figure 4.6(b)). For the visualization of the results we made use of the chart.js[11] library. This allows to dynamically exclude parameters of the analysis by clicking on the corresponding legend of the graph. The exact value of a parameter is displayed when users hover over the respective bar. In contrast to the relatively static player and team analysis the comparative query

[9] `http://loudev.com`
[10] `https://jquery.com/`
[11] `https://www.chartjs.org/`

(a) Comparing players w.r.t. certain performance parameters.



(b) Comparing queries on-the-fly w.r.t. certain performance parameters.

Figure 4.6 Quantitative analysis features of SportSense.

analysis can be dynamically adapted by changing the range of time via a slider for which we used the noUiSlider.js[12] library to implement it. To activate that slider a small `Timeline` button next to the `Analyze` button needs to be clicked (see Figure 4.6(b)). If the slider values are changed by the user via dragging, the results get dynamically updated together with the corresponding visualizations.

Qualitative Analysis. The last point which is required for a DAS to provide an optimized decision support in game sports is the qualitative analysis (see Definition 2.1). Coaches and analysts need to analyze certain events, patterns of events, and movements with all context information as these are contained, for instance, in video data. Therefore, SportSense also supports qualitative analyses. With the SportSense system the video gets segmented through the data. This is possible because the video can be linked to event and raw data via the timestamps (see Section 2.2.1.3). This is a huge advantage compared to other video analysis tools, where the match mostly has to be annotated manually (see Section 6.2.2). With SportSense users can search for an event and thus do not need to know *when* but only *where* that event happened. After the successful information retrieval, e.g., through an event query (see Definition 3.9), a click on the result will directly show the respective scene in the video player of SportSense. The same works for the results of motion paths and event cascades. This makes the working with SportSense very flexible and time-efficient at the same time [SJR+19].

4.2 SportSense Football

Having introduced the generic architecture of SportSense in the previous section we now present a first specific application: SportSense Football. In this section we show the characteristics of the invasion game football and the resulting adaptations to the generic architecture of SportSense in order to fit the needs of football coaches and analysts (see Section 4.2.1). Additionally, we present implementations of four different features of SportSense Football which are based on concepts of football coaches (see Section 4.2.2). Based on these implementations users are able to perform high-level tactical analyses with SportSense Football, which is an important requirement of performance analysts in sports practice as presented in Section 1.2.

[12] `https://refreshless.com/nouislider/`

4.2.1 Football-Specific Requirements and Adaptations

The generic SportSense architecture needs to be adapted to the characteristics of football to fulfil the requirements of football coaches and analysts. In order to know which changes need to be implemented it first makes sense to take a closer look at football rules. In the following, we briefly present important excerpts of the official rules of the game formulated by the IFAB[13].

Football is a game played between two teams of eleven players (one goalkeeper, ten field players). The match duration is 90 minutes divided into two equal halves of 45 minutes plus additional time for all time lost in that half through, e.g., substitutions or disciplinary sanctions. Football is played on a green playing field (natural, artificial, or hybrid system). To score a goal the whole of the ball needs to pass over the goal line, between the goalposts and under the crossbar. Additionally, there has to be no offence committed by the team which scores the goal. The team which scores the greater number of goals wins the match. If no goal or an equal number of goals is scored, the match is drawn.

Now that we know the key facts about the game of football, we present what the previously mentioned points imply for the three components of the generic SportSense architecture.

MongoDB. No adaptations are required for the MongoDB component. This is because data for football analyses are already stored appropriately in the respective collections to allow for an efficient retrieval process (see Section 4.1.1.2) and can be accessed via different query types (see Section 3.3.1) . As long as no new type of data is available, which differs fundamentally from the already existing ones, the collections presented in Section 4.1.1.2 are sufficient.

MongoDB REST Proxy. For the second component, the MongoDB REST proxy, this looks slightly different. Here, some adaptations need to be implemented. The changes are dependent on the corresponding football-specific analysis features which should be supported by SportSense Football (see Section 4.2.2). These adaptations are essential to adequately query data from the MongoDB and return the results to the Web client of SportSense Football for further processing.

[13] International Football Association Board (IFAB).
The full rules of 2020/2021 can be found here: `https://resources.fifa.com/image/upload/ifab-laws-of-the-game-2020-21.pdf?cloudid=d6g1medsi8jrrd3e4imp`.

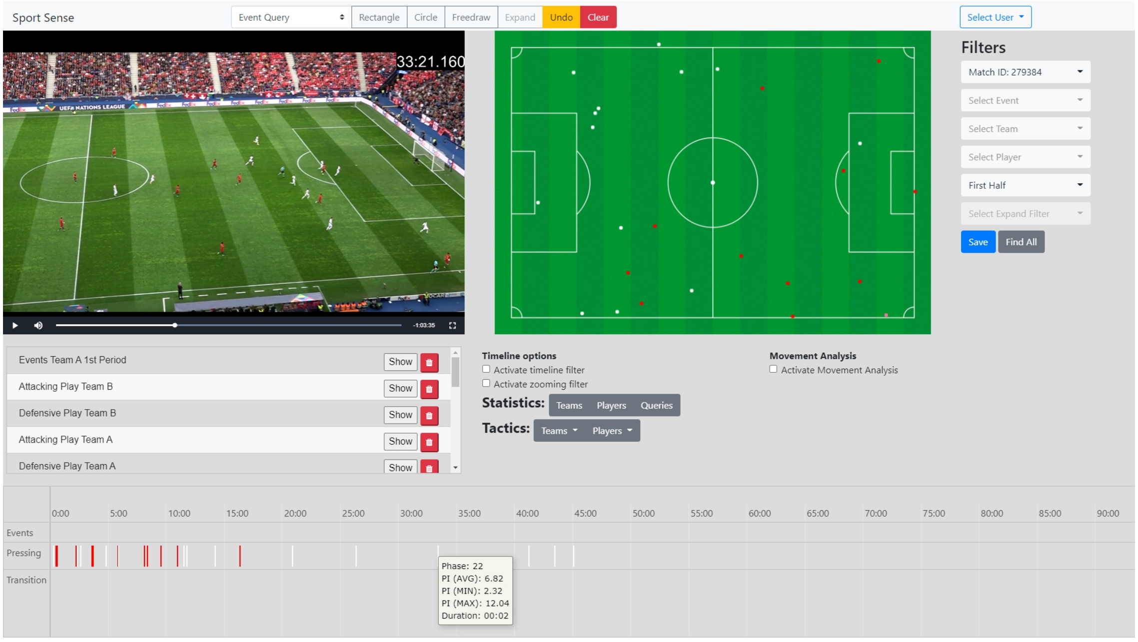


Figure 4.7 UI of the SportSense Football Web client.
The visualized results show the pressing phases of both teams during the first half of a match. Hovering over a result in the timeline shows additional information on that phase. The starting point of each phase is depicted as dot in the respective team color on the drawing area.

Web Client. Major changes which are required for a football-specific version of SportSense concern the Web client component. The characteristics of football imply some adaptations to the UI of the Web client. First, the timeline component needs to represent the 90 minutes of a match plus additional time which occurs regularly in football matches. Second, the background image of the drawing area needs to depict a realistic football pitch (see Section 3.4.5). Third, buttons which are needed to execute football-specific (high-level) tactical analyses need to be implemented. The UI of SportSense Football is depicted in Figure 4.7. Of course the individual football-specific analysis features which are based on coaches' concepts need to be implemented in the Web client component as well. This is shown more detailed in Section 4.2.2.

4.2.2 Implementation of Football-Specific Concepts

In the previous section we introduced the adaptations to the generic architecture which need to be implemented for a football-specific version of SportSense. In this section we now focus on some football-specific data and corresponding features which are supported by the SportSense Football system.

SportSense Football supports all generic features which we presented in Section 4.1.3.3. However, in order to allow users performing high-level tactical analyses with SportSense Football, football-specific concepts which represent the mental models of coaches first need to be implemented. In Section 3.1 we presented a way how to bridge the semantic gap of sports analytics. We were able to extract 22 concepts of football coaches from interviews (see Section 3.1.1). We assigned the resulting concepts to four different data types (events, phases, continuous states, and profiles) in a performance modeling step (see Section 3.1.2). Finally, we translated the performance model into a data model (see Section 3.1.3). With the latter we are now able to implement these concepts into SportSense Football.

In the following, we present four different features of the system which allow for (high-level) tactical football analyses as all these features are either direct implementations of certain concepts of coaches or implementations which are based on these concepts. These features serve as examples how the different data objects which resulted from the data modeling step can generally be implemented in the DAS SportSense. The implementations which we present in the following are based on spatio-temporal tracking data from a commercial provider and event data (atomic and non-atomic event data) generated by the StreamTeam Football [Pro20; PSS+20] system.

4.2.2.1 Pressing Index

The first feature which we want to introduce is the pressing index which is an implementation of the concept *Pressure* (see Table 3.2). This concept is assigned to the continuous state data type (see Definition 3.2) and represents an evolving metric (i.e., the pressing index) over time.

Pressing is very important in modern football and is applied by teams to set the opponent player under pressure with the intention to regain possession of the ball [SRP+20b]. Football coaches and analysts often want to know where, how intense, and for how long a team exerts pressure in certain situations.

To quantify this pressure, the pressing index was introduced and defined in the work of Seidenschwarz et al. [SRP+20b]. StreamTeam-Football [Pro20; PSS+20] calculates the pressing index continuously and stores the data as pressing states in the states collection of the MongoDB (see Section 4.1.1.2). Pressure, as a continuous state data type, consists of a continuous raw data stream (see Section 3.1.3). A single pressing state thus represents the corresponding data object of the concept *Pressure*. The pressing states are an additional, football-

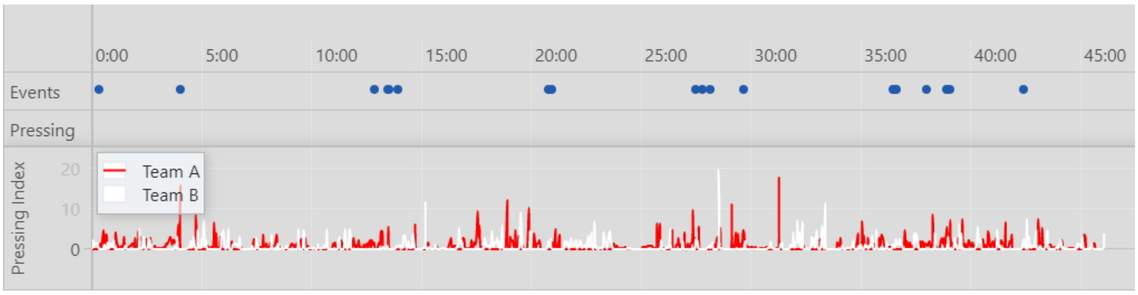


Figure 4.8 Pressing index visualization in the SportSense Football UI. The pressing index of the first half of a match is visualized as one line graph per team. The results of a previous event query are depicted as blue dots and allow users a detailed analysis of the pressing intensity in certain situations via zooming into the timeline.

specific type of raw data and are consistent with the formal description of raw data items (see Definition 2.4).

Users can execute a raw data query (see Definition 3.6) to analyze the pressing index with SportSense Football by selecting a match in the filter area of the UI, clicking afterwards on the team tactics dropdown below the drawing area (see Figure 4.7), and finally selecting the "Pressing Index" option (see Table 4.1 for a list with all options). The query results are visualized at the bottom of the UI below the timeline as two line graphs (see Figure 4.8) where each line represents a team's corresponding pressing index over the course of the selected match. For an intuitive understanding, the lines are visualized in team colors (see Section 3.4.5). Because the timeline is still visible, a coach can execute an additional event query (see Definition 3.9) and is then able to take a detailed look at how intense the pressing was during the resulting events [SRP+20b].

For the visualization of the pressing index we made use of the Graph2d component of the vis.js library. This component allows to display bar charts and line graphs on an interactive timeline. Unfortunately, it is not possible to combine the Graph2d components with the already present timeline component

Table 4.1 Options for tactical analyses with SportSense Football. The two dropdown menus and the corresponding options for a tactical analysis with SportSense Football.

Teams	Players
Pressing Index	Pass Network
Pressing Phases	Speed Analysis
Transition Offensive	
Transition Defensive	

of the Web client in one single instance. To solve this problem we implemented event listeners. These allow a synchronous behaviour of the Graph2d and the timeline instances during dragging and zooming activities of the user [SRP+20b].

4.2.2.2 Pressing Phases

Another analysis feature of SportSense Football are the so called pressing phases. In contrast to the pressing index, pressing phases themselves do not represent an implementation of a concept, but are an application based on a concept (i.e., *Pressure*).

For coaches and analysts this is an important feature as effective pressing phases often lead to critical shifts in the game state [SRP+20b]. SportSense Football facilitates the task of coaches and analysts by segmenting the match into different pressing phases based on the pressing index.

For that, two thresholds need to be defined. The first threshold defines the minimum intensity of pressing. The second threshold defines the minimum duration of a pressing phase. With these two thresholds it is possible to detect and visualize pressing phases with SportSense Football. Furthermore, it is possible to customize the values of the two thresholds. This enables coaches and analysts to work very flexibly with SportSense Football. A pressing phase is, as the name already indicates, represented as a phase data type (see Definition 3.1), which means that it has a beginning and an end. A pressing phase starts if the value of the pressing index exceeds the defined threshold and terminates if the value of the pressing index falls under the defined threshold and the duration of the phase already reached the minimum amount of time.

A user can analyze the pressing phases by selecting a certain match in the filter area of the UI, clicking afterwards on the team tactics dropdown menu, and selecting the "Pressing Phases" option (see Table 4.1). This feature is based on a raw data query (see Definition 3.6) as the pressing states (i.e., football-specific raw data) are used for the detection of pressing phases. The resulting phases are visualized as bars in the respective team colors (see Section 3.4.5) on the timeline component of the UI. Additionally, the starting point of each pressing phase is depicted as dot, again in the corresponding team color, on the drawing area of the UI (see Figure 4.7). The bars and dots are clickable which means that coaches and analysts can perform qualitative analyses and directly watch and analyze the corresponding phases in the video player. By hovering over a specific bar, additional information like, for instance, average, minimum, and maximum values of the pressing index during that phase are highlighted.

For the implementation of phases we made use of the functionalities of the timeline component of the vis.js library. We implemented additional rows within the timeline for the visualization of pressing phases. To display the phases, items (i.e., phases) need to be defined with a start- and endpoint. Additionally, items need to be assigned to a different group because the groups influence the row in which an item will be displayed. As the first group contains all the atomic events (top row of the timeline), we decided to assign pressing phases to the second group. Additional information can be displayed either in text form in the item (i.e., the bar) itself or by hovering over it [SRP+20b].

With this feature we have seen that not only events can delimit phases, but continuous states (see Definition 3.2) can also serve to segment a match by defining certain thresholds of the corresponding metric and the timespan.

4.2.2.3 Transition Phases

The next feature of SportSense Football which we introduce are transition phases. They can occur either as offensive transition phases after gaining ball possession or as defensive transition phases after losing ball possession, respectively. Therefore, we implemented two phases: offensive transition phases and defensive transition phases. These transition phases represent two of the four phases of tactical periodization. The other two are offensive and defensive organization, respectively [DM12]. The transition phases themselves are not implementations of concepts which we extracted from coaches, but they are again based on the implementation of two concepts. An offensive transition phase is based on the atomic event *Gain Possession* and a defensive transition phase is based on the atomic event *Lose Possession* (see Table 3.2).

Transition phases are particularly important to analyze, because these can quickly lead to dangerous situations or even to goal scoring opportunities as can be seen, e.g., in the work of Hobbs et al. [HPS+18]. While the defensive team is temporarily unorganized after the loss of ball possession the offensive team tries to exploit the situation and to quickly gain an advantage by gaining territory. On the other side, it is important for the defending team to get organized as quickly as possible to prevent exactly such a situation [SRP+20b]. Transitions which lead to goal scoring opportunities thus are particularly interesting for coaches [BKL05].

To detect transition phases the SportSense Football system uses event queries (see Definition 3.9) to check for all ball possession changes in the data. Similar to a pressing phase, a transition phase also represents a phase data type

Figure 4.9 Transition phases visualization in the SportSense Football UI. The results show the offensive transition phases of Team B during the first half of a match as bars in the timeline. Additionally, the respective starting points of each phase are depicted as blue dots on the drawing area.

(see Definition 3.1). The start of a transition phase is defined as the moment where the possession changed [SRP+20b]. This means an offensive transition phase of a team starts at the moment ball possession is gained, whereas a defensive transition phase of a team starts at the moment ball possession is lost. The end of a transition phase cannot be defined easily because there exist several philosophies and almost every coach or club might define it differently. For our implementation we thus decided to fix the duration of a transition phase at five seconds [SRP+20b]. This timespan is a good compromise as good teams should be able to get organized after loss of ball possession within that time interval or otherwise should already have gained territory after gaining possession of the ball.

To analyze either offensive or defensive transition phases with SportSense Football users need to specify a match and the team in which they are interested in the filter area of the UI. Afterwards, the team tactics dropdown needs to be clicked and the user can then select either “Transition Offensive” or “Transition Defensive” (see Table 4.1). Transition phases are visualized similarly to pressing phases as bars on the timeline component of the Web client UI (see Figure 4.9). Additionally, the starting point of each transition phase is depicted on the drawing area which allows the detection of potential patterns at first glance if clusters would be visible. This in turn would show where on the pitch a team gains or loses possession of the ball very often and could either indicate a team's strength or weakness [SRP+20b]. Similar to pressing phases, the results (i.e., bars and dots) are clickable and thus can quickly be analyzed qualititatively by watching the corresponding video sequences.

The implementation of transition phases is very similar to pressing phases. Again, the timeline component of the vis.js library needs to be modified. The transition phase items need to be defined with start- and endpoint. Furthermore, the items need to be assigned to a group. Because the first group is reserved for the atomic events and the second group is reserved for the pressing phases, we implemented a third group for the transition phases. This allows to display the transition phases in a third row of the timeline (see Figure 4.9).

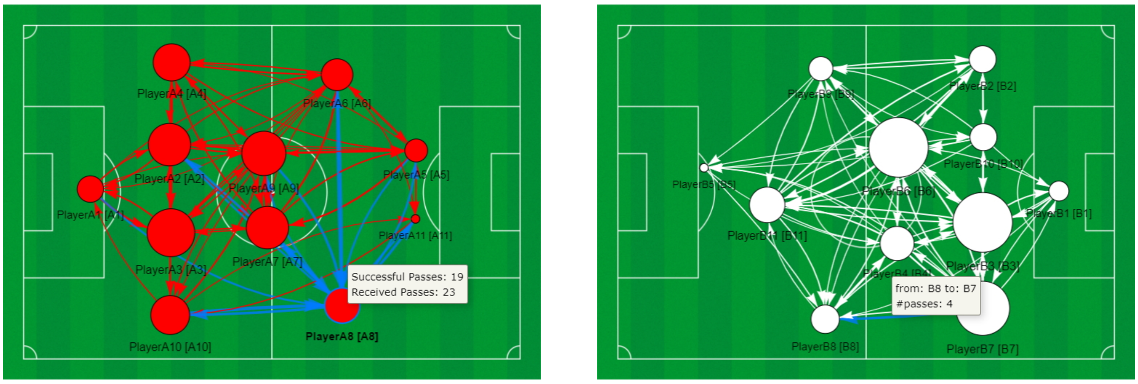
4.2.2.4 Pass Network

The last feature of SportSense Football which we want to highlight is the pass network. This feature is one possible implementation of the concept *Cooperation* (see Table 3.2) which is assigned to the profile data type (see Definition 3.3).

Profiles generally are used to characterize players and teams by aggregating information over several matches [SRP+20a]. Therefore, most profiles are based on statistics and/or indicators. Among other information, it is interesting for coaches and analysts how players cooperate. One way to analyze the cooperation between players is to take a look at how each player is involved in the passing of the whole team. Passing is the most frequent action in football [FBC19]. Additionally, frequent and accurate passing are strongly linked to shots, goals, and points [Col13]. To find key players thus is especially important when coaches and analysts prepare for an upcoming match. The own tactics can then be adapted in a way to disturb, e.g., the build-up play of the opponent's key players [SRP+20b].

Passing is a perfect indicator of team play [Cha13] and pass networks are a good option to visualize the passing of a team. Although this approach is not new and other researchers already introduced (social) network approaches in the field of football before (see Section 6.1.1) it nevertheless has a big potential to facilitate the task of analysts and coaches as it allows the detection of key players at first glance. By iterating through all successful passes in the event data collection (see Section 4.1.1.2), summing up the passes for each player and the corresponding receivers as well as the links between the players, SportSense Football is able to calculate pass networks [SRP+20b].

In order to use the pass network feature, users need to select a match and a



(a) Highlighted node (player). **(b) Highlighted edge (connection).**

Figure 4.10 Pass network visualization in the SportSense Football UI. Clicking on a node or an edge highlights the respective player connection(s) in blue. Hovering over either a node or an edge shows additional information on the involved players.

team in the filter area of the UI. Afterwards the player tactics dropdown menu needs to be clicked and subsequently the "Pass Network" option needs to be selected (see Table 4.1). The results of the event query (see Definition 3.9) are displayed on the drawing area. Two examples are shown in Figure 4.10. Each player is represented as a node. The size of the node is dependent on the total number of successful passes made by that player. The bigger the node, the more passes were played. Edges between two nodes represent the passes between two players. The thickness of the edge represents the number of passes. The thicker an edge, the more passes were played between the corresponding players. Coaches and analysts therefore can quickly find key players or key connections as these are represented as the biggest nodes or edges, respectively [SRP+20b]. Clicking on one node highlights all connections (edges) of that player in blue (see Figure 4.10(a)). By hovering over either nodes or edges, users can see detailed information like, for instance, the exact number of passes as depicted in Figures 4.10(a) and 4.10(b).

To implement the pass network in SportSense Football we made use of another component of the vis.js library, the network component. Nodes can be defined either with a fixed size or with a dynamically adapted size depending on a specified value. The latter is more useful for our implementation and we specified the number of successful passes as the value which influences the size of the nodes. Additionally, the position of the nodes can be defined through (x, y) coordinates. In the case of football it makes sense to position the nodes in a way that they represent the tactical line-up of a team. Nodes can be moved via drag and drop if the positioning does not fit the coach's or analyst's idea. Similar to the size of nodes, the thickness of edges can be dynamically adapted by specifying a value. For our implementation we selected the number of successful passes between the involved players [SRP+20b]. Another important detail which can be defined is the coloring of nodes and edges. We decided to implement them in the team colors to make the visualization as intuitive as possible (see Section 3.4.5). Additionally, information can be shown in text form either inside or below each node and edge, respectively, but also by hovering over one of it [SRP+20b].

4.3 SportSense Ice Hockey

With SportSense Football we presented a first specific application of the DAS SportSense in the previous section. In this section we now introduce a second application: SportSense Ice Hockey. We first show the characteristics of this invasion game together with the resulting adaptations to the generic SportSense architecture which are needed to fit the requirements of ice hockey coaches and analysts (see Section 4.3.1). Second, we present implementations of two features which allow for basic tactical analyses with SportSense Ice Hockey (see Section 4.3.2).

4.3.1 Ice Hockey-Specific Requirements and Adaptations

Performance analysis in elite ice hockey has rapidly grown over the past years [LRM20]. Despite football and ice hockey have some similarities, as both sports are invasion games, the characteristics of ice hockey and football are different. Ice hockey coaches and analysts have thus different requirements compared to their colleagues from football. In the following, we present some important characteristics of the invasion game ice hockey based on the official rules of the game formulated by the IIHF[14].

An ice hockey match is played between two teams of six players: one goaltender and five skaters. A match consists of three periods of 20 minutes plus overtime and, if necessary, a penalty-shot shootout. Ice hockey is played on a white surface of ice, which is called rink. Each team tries to score more goals than the opponent. To score a goal a team needs to put the puck rule-consistent

[14] International Ice Hockey Federation (IIHF).
The full rules of the game (2018-2022) can be found in the official rule book of the IIHF: `https://blob.iihf.com/iihf-media/iihfmvc/media/downloads/rule%20book/iihf_official_rule_book_2018_ih_191114.pdf`.

into the goal net of the opponent. The team which scores more goals at the end of the match is declared the winner. If no winner can be declared after the third period, a so called overtime (5-, 10-, or 20-minute period) needs to be played in which the winner is declared on a sudden-death basis (i.e., the next goal wins). If during the overtime no team can be declared a winner, a penalty-shot shootout needs to be played to finally decide on the winning and losing team.

In the remainder of this section we briefly highlight what these characteristics imply for the three components of the generic SportSense architecture.

MongoDB. Important data for the analysis in ice hockey (i.e., raw data and event data) are, similar to football data, already stored appropriately in the MongoDB component (see Section 4.1.1.2). Data can be accessed executing different query types (see Section 3.3.1). Therefore, no changes to the MongoDB component are required as long as no new data types are available for the analysis, which are fundamentally different to the already existing ones.

MongoDB REST Proxy. Changes concerning the MongoDB REST proxy are dependent on the ice hockey-specific analysis features (see Section 4.3.2). The adaptations are necessary to adequately query data from the MongoDB and to return the query results to the Web client of SportSense Ice Hockey where they get further processed.

Web Client. Similar to SportSense Football, the most changes need to be implemented in the Web client component. Individual components of the UI are influenced by the characteristics of ice hockey. The timeline needs to represent the 60 minutes of a regular match as well as an optional overtime period. The background image of the drawing area needs to represent an ice hockey rink (see Section 3.4.5). Additionally, buttons to execute ice hockey-specific features need to be implemented. The UI of SportSense Ice Hockey with the previously described adaptations is depicted in Figure 4.11. Finally, the ice hockey-specific analysis features themselves have to be implemented in the Web client component. Two examples are presented in the following section (see Section 4.3.2).

4.3.2 Implementation of Ice Hockey-Specific Concepts

Similar to SportSense Football, all generic features which we presented in Section 4.1.3.3 are supported with SportSense Ice Hockey as well. However, in order to allow users performing ice hockey-specific tactical analyses, some

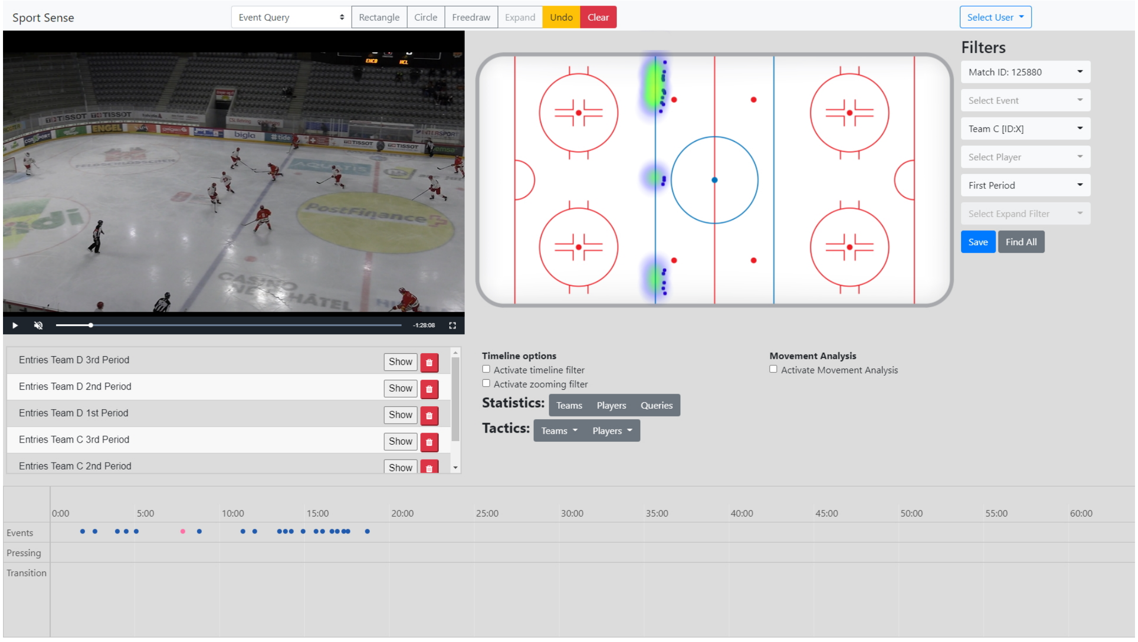


Figure 4.11 UI of the SportSense Ice Hockey Web client.
The visualized results show the entry events of Team C during the first period of a match as heatmaps and blue dots on the drawing area as well as blue dots on the timeline.

additional features need to be implemented. These features ideally should be based on coaches' concepts which represent their mental models of the game. This in turn would require a complete process as depicted in Section 3.1 with domain experts like, e.g., ice hockey coaches. In this case, we cannot profit from interviews and extracted concepts like before in the case of football. Therefore, we present implementations of two simple features based on ice hockey-specific events which allow only for very basic tactical analyses with SportSense Ice Hockey. Nevertheless, these examples show the flexibility of the generic architecture to relatively easy adapt the system and implement new additional sport-specific features. Both implementations which we present in the following are based on manually tagged event data.

4.3.2.1 Entry Analysis

Offensive zone entries represent a fundamental element in the attacking play of ice hockey teams. An entry event is defined at the moment the offensive player in possession of the puck enters the offensive zone on the rink, i.e., he surpasses the blue line. As the zone entry happens at a single point in time, the event can be represented as an atomic event and thus as a single event data item (see Definition 2.6). Consequently, it can be stored in the event collection of the

Table 4.2 Options for tactical analyses with SportSense Ice Hockey.
The two dropdown menus and the corresponding options for a tactical analysis with SportSense Ice Hockey.

Teams	Players
Entries	Pass Network
Shifts	Speed Analysis

MongoDB (see Section 4.1.1.2). The analysis of entry events can help ice hockey coaches and analysts especially when they prepare for an upcoming match. The analysis of the next opponent's entries might reveal certain preferences and can help to develop defensive strategies and to adapt own tactics in order to prevent dangerous situations.

To execute a corresponding event query (see Definition 3.9) and visualize the entry events with SportSense Ice Hockey two different options are available. First, the traditional way to execute event queries with the SportSense system (see Section 4.1.3.3). Here, users need to define the corresponding event (entry) in the filter area first and (optionally) delimit the area on the canvas-element in a sketch-based manner. To further delimit the query results additional filter options (e.g., match, player, team, or time) can be selected. The second option which is supported by SportSense Ice Hockey is to (optionally) select a team and/or a player in the filter area, click on the team tactics dropdown menu, and finally select the "Entries" option (see Table 4.2 for a list with all options). In both cases the MongoDB gets queried and the corresponding entry events from the event collection are returned via the MongoDB REST proxy to the Web client. The results are then displayed in both cases on the drawing area and on the timeline component of the UI. In contrast to the first option, the second option visualizes the results together with an additional heatmap on the drawing area (see Figure 4.11). This allows to detect the preferred entry zones of a team at first glance because clusters of events get highlighted in "warmer" colors (from blue to green, yellow, and red) the bigger they are. In Figure 4.11 it is recognizable that the team conducts most entries into the offensive zone via the right side of the rink.

For the implementation of the heatmap functionality we made use of the heatmap.js[15] library.

[15] `https://www.patrick-wied.at/static/heatmapjs/`

4.3.2.2 Shift Analysis

Player substitutions in hockey are the norm compared to other game sports like baseball, where substitutions are rare and strategic [MF07]. Additionally, in ice hockey, teams can substitute players dynamically during the running match at any time. Compared to other invasion games like football or basketball, this is a very special characteristic. A replacement of a player by another player from the bench is called a shift event. Shift events represent an important tactical element. Coaches can, for instance, replace exhausted players by others which already recovered and therefore are able to perform better. This in turn can either maximize chances of scoring a goal or helps defending, e.g., against a superior offensive block of the opponent's team. Because the timing of shift events is very important and changes can also go wrong it makes sense to analyze and optimize these events. SportSense Ice Hockey supports coaches and analysts in this task.

As a single shift event happens at a single point in time, the event can be represented as an atomic event and thus as a single event data item (see Definition 2.6). Furthermore, it can be stored in the event collection of the MongoDB (see Section 4.1.1.2). In contrast to other atomic events, our (manually) tagged shift events contain no spatial information (i.e., no values for the (x,y,z) coordinates).

To execute an event query (see Definition 3.9) and visualize the shift events with SportSense Ice Hockey two different options are available. First, users need to select the corresponding event (Shift) in the filter area. Optionally, other filters can be selected in the filter area. Because we tagged shift events without coordinates a further limitation of the query results by defining a certain region in a sketch-based manner is not possible. To execute the event query the `Find All` button needs to be clicked. The second option to execute the same query is to (optionally) select a team in the filter area, clicking afterwards the team tactics dropdown menu and select the "Shifts" option (see Table 4.2) which executes the event query. In both cases, the MongoDB gets queried and the corresponding shift events from the event collection are returned via the MongoDB REST proxy to the Web client. The query results are visualized on the timeline component only. Coaches and analysts can then analyze the shift events qualitatively by watching the corresponding video scenes.

Part III

Discussion

5 Evaluation

You miss 100 percent of the shots you don't take.

— Wayne Gretzky

In the previous chapter we introduced the DAS SportSense together with its generic architecture and generic analysis features for game sports. Moreover, we presented two applications which provide promising approaches for decision support of coaches and analysts in football and ice hockey, respectively. SportSense Football is the more sophisticated application and was implemented based on our generic concepts for the development of DAS in game sports. Additionally, its implementation was influenced by sport-specific concepts which we extracted from interviews with football coaches.

In this section, we now evaluate SportSense Football from two different perspectives. First, we conduct user studies to evaluate qualitatively if all generic and football-specific concepts are represented appropriately (see Section 5.1). Second, we evaluate the retrieval performance and the scalability of the system quantitatively (see Section 5.2) to check whether the system is utilizable in "real-world" scenarios of football practice.

5.1 User Studies

To check if the generic and football-specific concepts are appropriately represented in SportSense Football we conducted user studies of the system. Domain experts from football were asked to evaluate SportSense Football with respect to several analysis features and to the UI design. In this section, we first present the hypotheses in Section 5.1.1. Afterwards, we introduce the setup of the evaluation process in Section 5.1.2. The corresponding results are presented in Section 5.1.3. In Section 5.1.4 we finally discuss the results with respect to our hypotheses.

5.1.1 Hypotheses

We developed the DAS SportSense Football with the intention to support football analysts and coaches in their decision making processes in sports practice with a strong focus on the match preparation context. For the implementation we made use of all our concepts which we presented in Chapter 3. Thus, we included a combination of different analysis options (see Section 3.2), the support of four different query types (see Section 3.3) as well as a user-centered design of the UX of the system (see Section 3.4). Additionally, we implemented some football-specific concepts, which we extracted from domain experts (i.e., coaches) to bridge the semantic gap of sports analytics (see Section 3.1). With the user studies we want to check the following hypotheses:

Hypothesis 1. SportSense Football's combination of three different analysis options is highly valuable for analysts and coaches.

Hypothesis 2. The four query types supported by SportSense Football comprehensively address the needs of analysts.

Hypothesis 3. The football-specific concepts *Pressure*, *Lose Possession*, *Gain Possession*, and *Cooperation* are adequately represented in the analysis features of SportSense Football.

Hypothesis 4. The UI components of SportSense Football are adequately represented and clearly arranged.

5.1.2 Setup

We conducted virtual user studies via the video communication software Zoom[1] with twelve domain experts from football: six analysts and six coaches. After a brief introduction to the research project and the exact evaluation procedure by the developer, participants had to answer a questionnaire with 60 different questions, for which we used Microsoft Forms[2]. The questionnaire consists of six parts: (1) general questions, (2) analysis options, (3) query types, (4) the semantic gap of sports analytics, (5) design, and (6) final questions. During part 3 and part 4 the developer demonstrated several analysis features of SportSense Football via sharing the screen with the participants. The type of questions

[1] `https://zoom.us/`

[2] `https://forms.office.com`

varied and included, e.g., yes/no, single choice, and multiple choice questions. Additionally, some questions expected answers in text form. Furthermore, users also had to evaluate certain features or components on a Likert scale from 1 ("very bad") to 5 ("very good"). The whole questionnaire is attached in Appendix C.

5.1.3 Results

In the following, we present the results of the user studies. We do this in six different parts which correspond to the six parts of the questionnaire which participants had to answer one after another. The average time for solving the questionnaire was 49:13 minutes.

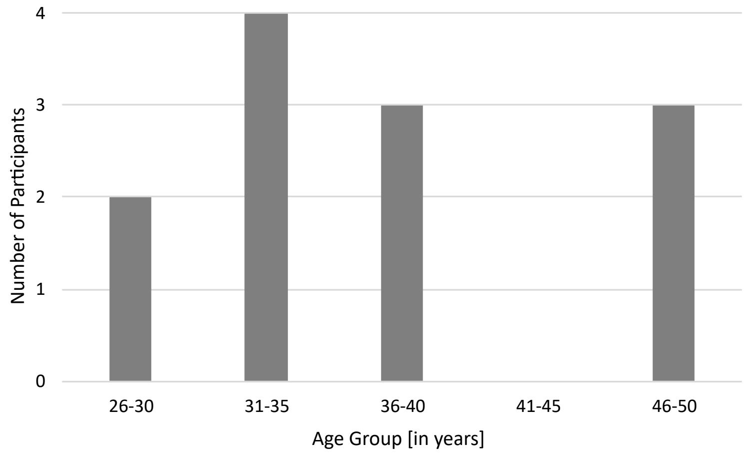
5.1.3.1 Part 1 - General Questions

Part 1 gives a brief overview on the twelve participants (1 female, 11 male). The average age of the participants was 37 years. The age distribution between different groups is shown in Figure 5.1(a). Seven participants had active coaching licenses whereas five had none. The coaching license distribution between different levels is depicted in Figure 5.1(b). Eleven participants were employees at football clubs, whereas one was employed in a football association. Most of the participants were employed in Germany, namely in German Bundesliga (n=6), 2.Bundesliga (n=1), 3.Liga (n=1), as well as in Regionalliga U19 (n=1) and U15 (n=1). The distribution among the different leagues is shown in Figure 5.1(c).

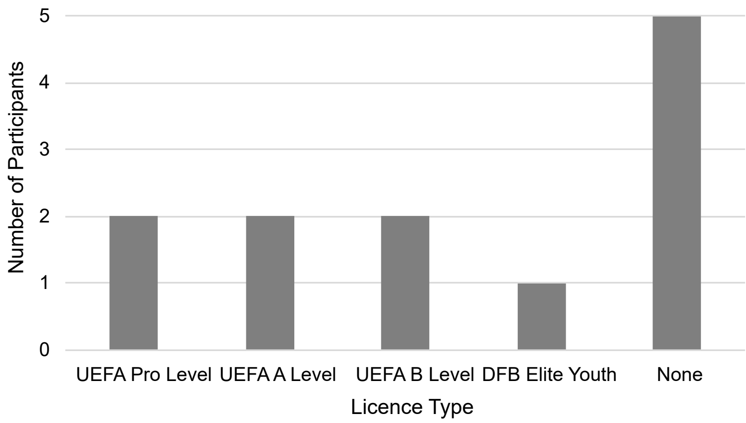
Discussion. The fact that five participants had no active coaching license is easily explainable as, in contrast to coaches, a coaching license is not a mandatory requirement for analysts.

5.1.3.2 Part 2 - Analysis Options

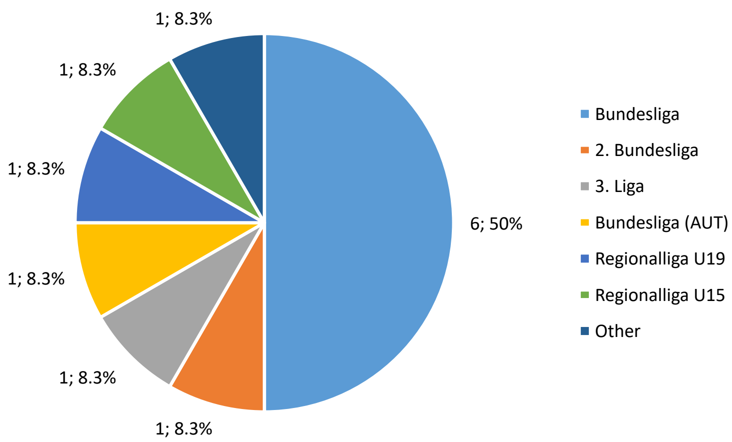
In this section we present the results of the second part of the user studies which mainly focused upon analysis options in general. All of the participants do use software tools for their analyses in order to prepare for an upcoming match. 91.7% (n=11) use at least two software tools. To be more precise, 25% (n=3) use two software tools and 66.7% (n=8) of the participants use even more than two tools. An overview on these statistics is presented in Figure 5.2(a). When we take a closer look at the different analysis options and how they are applied by the participants we can state that video analysis is the most used option



(a) Age distribution.

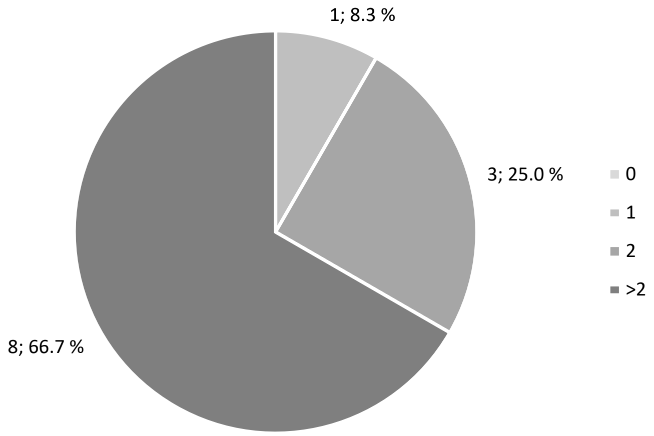


(b) Coaching licenses distribution.

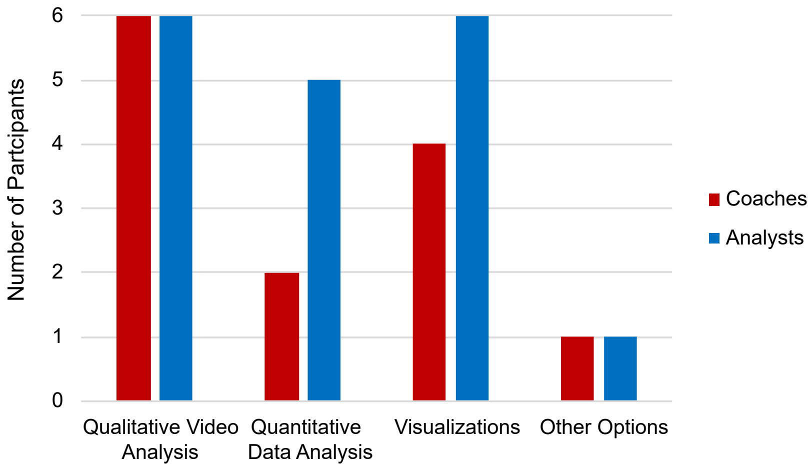


(c) League affiliation.

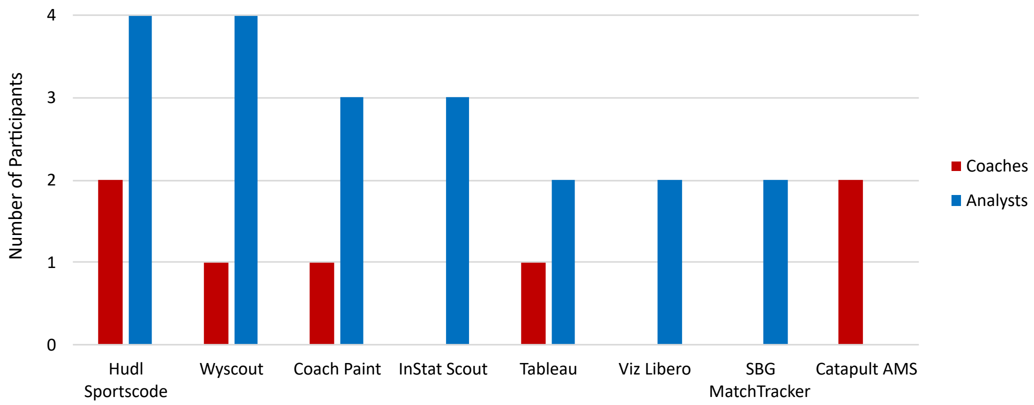
Figure 5.1 General information on participants of the user studies.



(a) Number of software tools used by participants.



(b) Analysis options used by coaches and analysts.



(c) Most used software tools by coaches and analysts.

Figure 5.2 Information on software tools and analysis options of participants.

(100%, n=12) by the participants, followed by visualizations (83.3%, n=10), and data analysis (58.3%, n=7). Figure 5.2(b) shows how coaches and analysts use different analysis options for their analyses. In Figure 5.2(c) we present the participants' most used software tools. Note that only tools are listed, which are mentioned at least twice. A list with all mentioned software tools which are used by the participants in sports practice is depicted in Table D.1. All participants agreed that they would profit from a system which combines different analysis options. The average number of matches which are considered during match preparation is 4.2 (maximum number = 7).

Discussion. The fact that 91.7% of the participants use at least two tools for their analyses is not surprising because the tasks of coaches and analysts are manifold (see Section 1.2). Thus, for each task they often need to use a "specialized" tool, e.g., one tool for the qualitative video analysis, another tool for the quantiative analysis with which they create statistical reports on player strengths and weaknesses, etc. Together with the fact that all participants agreed that they would profit from a system which combines different analysis options, this just underlines the necessity to support coaches and analysts in sports practice with a DAS like SportSense Football.

All participants make use of video analysis for the preparation of an upcoming match whereas only seven participants rely on (additional) data analysis, i.e., via reading reports. This confirms that the video is still the most important source of information as already mentioned in Section 1.2 and is indispensable within a good DAS. Most of the participants use visualizations for their analysis process. This is not surprising as well because visualizations help enormously during the decision making process through the contextualization of information (see Section 2.2.4). In Figure 5.2(b), we can see that only two coaches use data analysis for their match preparation process. It is difficult to say why this is the case. One might assume that coaches, traditional coaches in particular, are not open to statistics and/or indicators which originate from a computer or from other technical devices, but that they only (want to) rely on their own gut instinct. Another probable interpretation is that the statistics and indicators which are part of the data analysis process do not correspond to the coaches' concepts and therefore are not useful for the match preparation. In this case, the problem would be the semantic gap of sports analytics (see Section 2.1.3). To cope with this problem we presented a solution in Section 3.1.

We can see in Figure 5.2(c) and Table D.1 that analysts use more tools than coaches. This could be because coaches do not have enough time to work with many different software tools and concentrate more on work on the pitch. Therefore, this statistics could just show the clear division of work between the analyst and the coach. Another explanation could be that many software tools are just too complex and are not designed intuitively enough for coaches. This intuitive handling is exactly what SportSense Football provides. In the optional textual feedback at the end of the questionnaire one of the coaches (UEFA Pro Level) wrote that SportSense Football is a cool system which he/she would like to use.

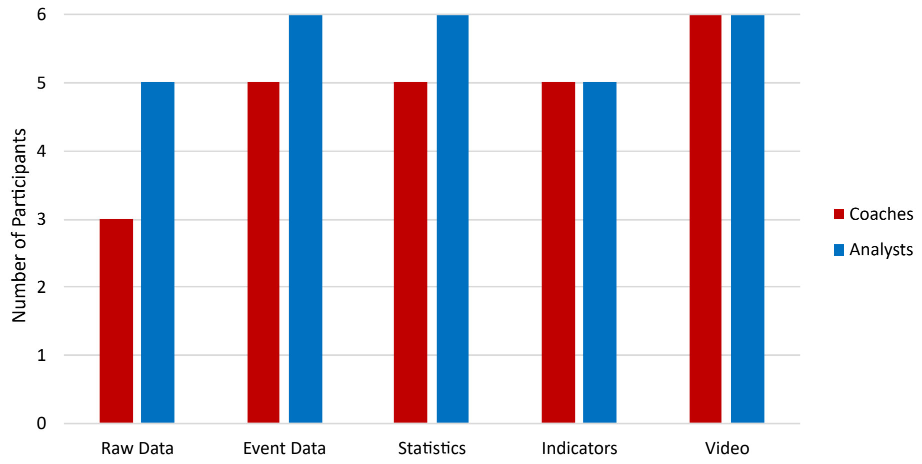
5.1.3.3 Part 3 - Query Types

Part 3 focuses on the four different query types (see Section 3.3.1). With regard to the movement query we only evaluated raw movement queries because in-event movement queries are currently not supported by SportSense Football. Concerning the data which are relevant for the participants' analyses we can state that all participants (100%, n=12) used the video as source of information, followed by event data (91.7%, n=11), statistics (91.7%, n=11), indicators (83.3%, n=10), and raw data (66.7%, n=8). The distribution of relevant data used for participants' analyses is presented in Figure 5.3(a).

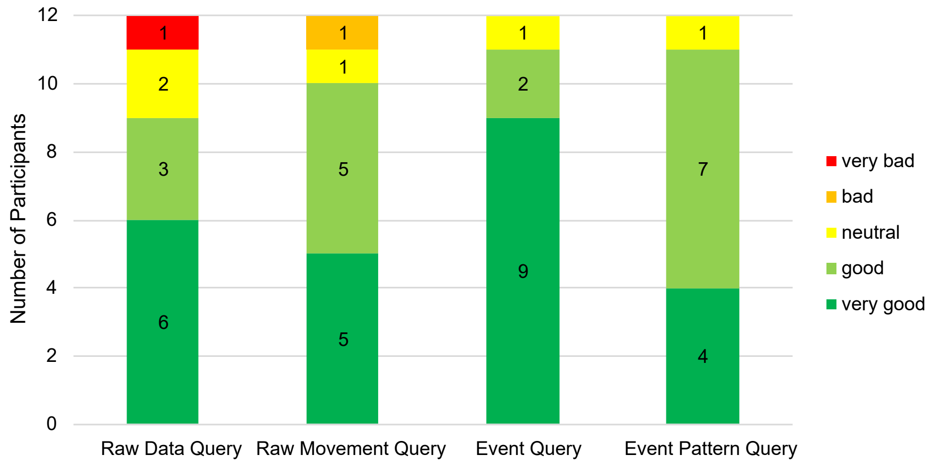
Figure 5.3(b) illustrates the results of the participants' usability rating for the four different query types which we demonstrated during the user studies. The usability of the raw data query was rated by 50% (n=6) of the participants with very good, by 25% (n=3) with good, by 16.7% (n=2) with neutral, and by 8.3% (n=1) with very bad. The usability of the raw movement query was rated by 41.7% (n=5) of the participants with very good, by 41.7% (n=5) with good, by 8.3% (n=1) with neutral, and by 8.3% (n=1) with bad. The usability of the event query was rated by 75% (n=9) of the participants with very good, by 16.7% (n=2) with good, and by 8.3% (n=1) with neutral. The usability of the event pattern query was rated by 33.3% (n=4) of the participants with very good, by 58.3% (n=7) with good, and by 8.3% (n=1) with neutral.

Concerning the questions if participants would use certain query types for their own analyses we can state that 75% (n=9) of the participants would use raw data queries for their own analyses and 25% (n=3) would not. 75% (n=9) would use raw movement queries for their analyses and 25% (n=3) not. All participants would use event queries as well as event pattern queries for their own analyses.

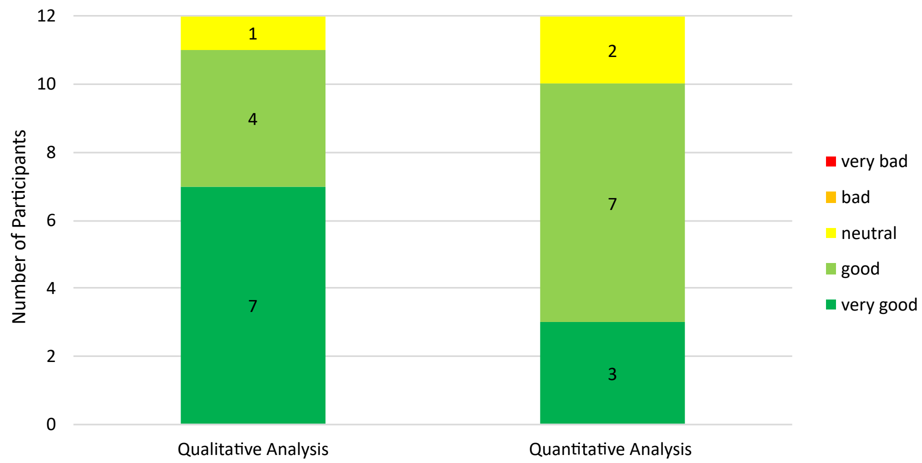
During the third part of the user studies the developer also demonstrated a qualitative analysis and a quantitative analysis with the SportSense Football



(a) Relevant data for participants' analyses.



(b) Results of the usability rating of the four different query types.



(c) Results of the usability rating of qualitative and quantitative analyses.

Figure 5.3 Relevant data for analyses and results of usability ratings.

system to the participants. The usability of the qualitative analysis was rated by 58.3% (n=7) of the participants with very good, by 33.3% (n=4) with good, and by 8.3% (n=1) with neutral. In contrast to that, the usability of the quantitative analysis was rated by 25% (n=3) of the participants with very good, by 58.3% (n=7) with good, and by 16.7% (n=2) with neutral. These results are also depicted in Figure 5.3(c). Again, all participants would use a qualitative analysis as well as a quantitative analysis when they prepare for an upcoming match.

Discussion. Only 66.7% (n=8) of the participants use raw data for their analyses. To be more precise only three of the six coaches use these kind of data (see Figure 5.3(a)). This can be explained with the fact that raw data often are available in football clubs in the form of physiological data which in turn are more used by other staff members like, e.g., the strength and conditioning coach. Other raw data types like spatio-temporal tracking data which also have a great potential for tactical analyses would be more interesting for (head) coaches (see Section 2.2.1.1). Nevertheless, spatio-temporal tracking data are often used to calculate only the external load of players (e.g., running distances) and are much less used for tactical analyses. Consequently, this again might be less interesting for the (head) coach and could explain why coaches use raw data less often for their analyses.

Figure 5.3(b) shows the very positive feedback of SportSense Football's evaluated query types. The best rated query type was the event query which users described very positively in the textual feedback, e.g., as "intuitive" or as a "super fast overview on clear actions". One participant would like to have this functionality extended by predefined zones in the playing field component. We will consider this in our future work. The fact that all participants would use this feature for their own analyses confirms the importance of the event query type. Similar results can be found for the event pattern query. Here, again all participants would use it for their own analyses. Nevertheless, one participant wrote that there is a risk of getting bogged down in details with this feature. One negative vote ("bad") is noticed for the raw movement query. When we take a closer look at the textual feedback it gets clear that for this user the running trajectories are already visible in the video and that he/she does not need to get this information visualized again. Still, most of the participants would use this query type for their own analyses. Another negative vote ("very bad") can be found for the raw data query. Here, the participant in question wrote in the textual feedback that this feature would only be relevant if the results are linked to the video. Other participants confirm this by writing, e.g., that the link to the match actions (i.e., to the video) is missing. This functionality is currently not supported by SportSense Football. However, we will consider this feedback in our future work. Nevertheless, most of the participants would use a raw data query for their own analyses.

A very positive feedback can also be found for both, the qualitative and the quantitative analysis which both were evaluated by the participants. Here, the qualitative analysis was rated slightly better than the quantitative analysis. When we take a look at the textual feedback of the participants we can state, that two participants would like to have a better visualization of the quantitative analysis results. Here, one can argue that the user needs are very individual and that it is difficult to respect all different demands [Gar12]. An option to customize the visualization of the results could cope with that problem. Another comment on the quantitative analysis was that an export functionality should be provided by the SportSense Football system. An export functionality for qualitative analyses would also be essential for one of the participants. We will consider these points in our future work. The fact that all participants would use the qualitative as well as the quantitative analysis functionalities of SportSense Football confirms the successful implementation of these features.

5.1.3.4 Part 4 - Semantic Gap of Sports Analytics

The fourth part of the questionnaire covered questions on football-specific concepts and their representations in the SportSense Football system. Three different features were shown to the participants and should be evaluated with respect to the representation of the corresponding concepts. The results are shown in Figure 5.4. The first feature, representing the concept *Pressure*, was the pressing index and the resultant pressing phases (see Sections 4.2.2.1 and 4.2.2.2). 41.7% (n=5) of the participants rated the representation of the concept *Pressure* with very good, 33.3% (n=4) with good, and 25% (n=3) with neutral. The second feature, representing the concepts *Lose Possession* and *Gain Possession*, were the transition phases (see Section 4.2.2.3). 33.3% (n=4) of the participants rated the representation of the concepts *Lose Possession* and *Gain Possession* with very good, 33.3% (n=4) with good, and 33.3% (n=4) with neutral. The third feature, representing the concept *Cooperation*, was the pass network (see Section 4.2.2.4). 8.3% (n=1) of the participants rated the representation of the concept *Cooperation* with very good, 50% (n=6) with good, and 33.3% (n=4) with neutral, and 8.3% (n=1) with bad.

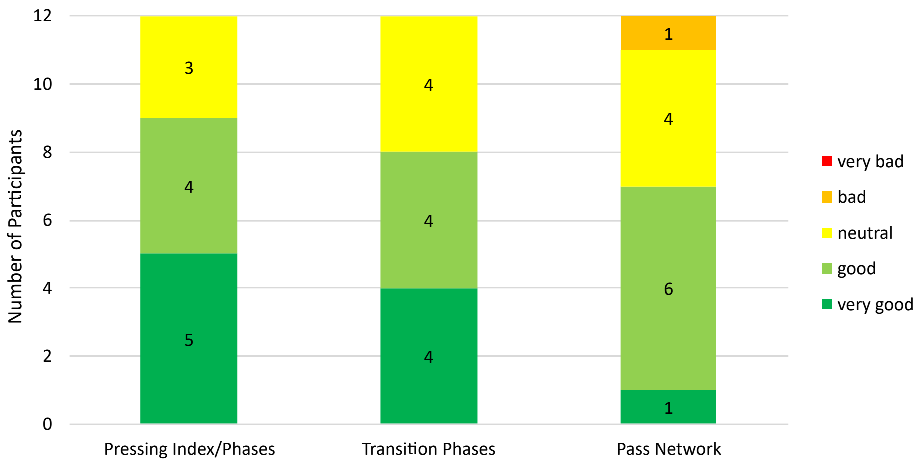


Figure 5.4 Results of the representation rating of different concepts. The concept *Pressure* is represented in the pressing index/phases feature of SportSense Football. The concepts *Lose Possession* and *Gain Possession* are represented in the transition phases feature. The *Cooperation* concept is represented in the pass network feature.

When we take a look at the answers to the questions if participants would use these features for their own analyses we can see that 83.3% (n=10) would use the pressing index/pressing phases feature and 16.7% (n=2) would not use it. 91.7% (n=11) of the participants would use the transition phases feature and 8.3% (n=1) would not use it. The pass network feature would use 83.3% (n=10) of the participants whereas 16.7% (n=2) would not use it.

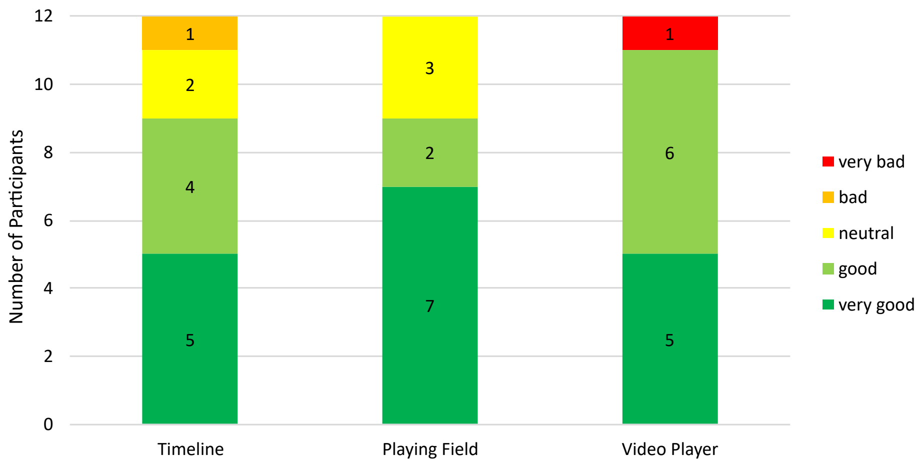
Discussion. All three football-specific features which represent different concepts were rated positively by the majority of the participants. The concept *Pressure* is adequately represented through the pressing index and pressing phases feature of SportSense Football (see Figure 5.4) and most of the participants would use it for their own analyses. However, the general feedback given by several participants in text form was that the pressing index implementation needs to fit the club's or the coach's playing philosophy. In other words, a customization of the pressing definition would be required to make this feature more valuable in practice. Additionally, a more detailed differentiation of pressing on the pitch (e.g., high pressure, midfield pressing, low pressure) would be helpful to get more detailed insights into the characteristics of a team's defensive playing style. We will consider all these points in our future work.

The concepts *Lose Possession* and *Gain Possession* are represented adequately in the transition phases feature of SportSense Football (see Figure 5.4). Almost all participants would use this feature for their own analyses. However, four participants only rated the feature with "neutral". In the textual feedback we can see that participants wish a support for more in-depth analyses like, e.g., the analysis of transitions with immediately following actions (i.e., events), or the specification of the position of a possession gain/loss. This is currently not supported but we will consider this point in our future work to make the transition phases feature even more valuable for coaches and analysts.

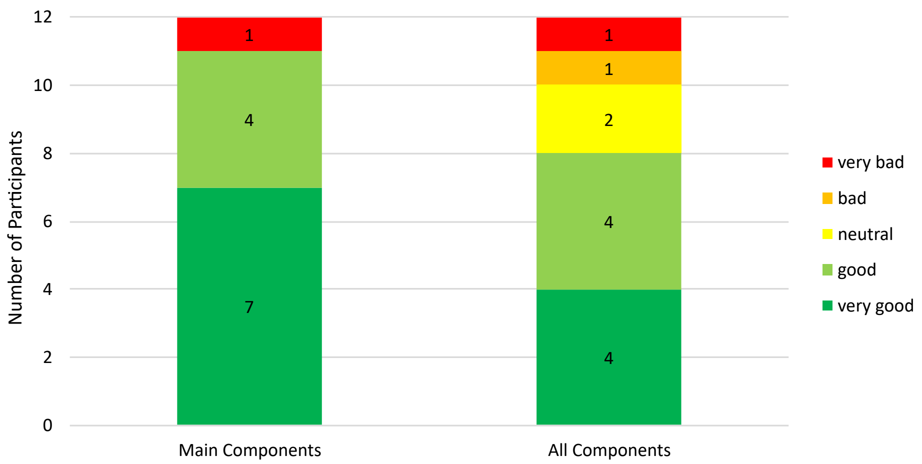
SportSense Football supports the representation of the concept *Cooperation* via the pass network feature. This is the only football-specific feature which has one negative vote ("bad"). Additionally, four "neutral" ratings can be found (see Figure 5.4). The problem of most participants with the feature is that context is missing. The loss of context is always existent when working with aggregated data (see Section 2.1.2). However, context is very important especially in football. So it might be the case that two defenders are heavily involved in the passing of their team but that most passes were played horizontally with very low risk and high success rate. Therefore, this feature might lead to misinterpretations and the coach or the analysts would get a wrong impression about a player's passing performance. To become more powerful, the pass network feature needs to be adapted in a way which considers more context. One option would be to only include passes which are played vertically. Another alternative provided by a participant is to consider the match phases during which passes are played (e.g., passes played during offensive organisation vs. passes played during offensive transitions). For another participant the link to a qualitative analysis is missing. This can be implemented, for instance, with a video collection of a player's passes which is generated when users click on the corresponding node. Despite the pass network feature is implemented with some weaknesses, most of the participants would use it for their own analyses.

5.1.3.5 Part 5 - Design

Part 5 covers questions on the design of the SportSense Football system and how its different components are represented. The implementation of these components was based on the concepts presented in Sections 3.4.4 and 3.4.5. The first questions concerned the representation of the three main components. The results are depicted in Figure 5.5(a). 41.7% (n=5) of the participants rated the representation of the timeline component as very good, 33.3% (n=4) as good, 16.7% (n=2) as neutral, and 8.3% (n=1) as bad. The representation of the playing field component was rated by 58.3% (n=7) of the participants as very good, by 16.7% (n=2) as good, and by 25% (n=3) as neutral. The representation of the



(a) Results of the representation rating of main components.



(b) Results of the clarity rating of components.

Figure 5.5 Results of components' representation and clarity ratings.

video player component was rated by 41.7% (n=5) of the participants as very good, by 50% (n=6) as good, and by 8.3% (n=1) as very bad. All participants agreed that they would use the timeline component for their own analyses. 91.7% (n=11) of the participants would use the playing field component and 8.3% (n=1) would not use it. The results are similar for the video player component. 91.7% (n=11) of the participants would use this component and 8.3% (n=1) would not.

Part 5 also evaluated the clarity of the arrangement of SportSense Football's components. The results are shown in Figure 5.5(b). If only the three main components are considered 58.3% (n=7) of the participants rated the clarity as very good, 33.3% (n=4) as good, and 8.3% (n=1) as very bad. If all components are considered, the clarity was rated by 33.3% (n=4) of the participants as very good, by 33.3% (n=4) as good, by 16.7% (n=2) as neutral, by 8.3% (n=1) as bad, and by 8.3% (n=1) as very bad.

Discussion. The representations of all three main components are rated positively by the majority of participants (see Figure 5.5(a)). When we take a look at the timeline component only one single negative vote ("bad") sticks out. Unfortunately, the participant in question did not leave a comment in text form to explain the vote. However, all participants would use the timeline component for their own analyses which might indicate that the negative voting resulted from a different design preference.

The playing field component came off very well with seven participants voting the representation as "very good" and no participant voting negatively (see Figure 5.5(a)). As most participants would use this component for their own analyses the implementation of the playing field component can be regarded as successful. When we take a look at some textual feedbacks of participants two suggestions for improvements are formulated. First, the playing direction should be defined. Such a definition reduces misunderstandings and wrong analysis results especially when more matches are analyzed. Second, jersey numbers of the players would be helpful if they get visualized on the playing field component. We will consider these points in our future work.

An interesting distribution of votes can be observed for the video player component. Here, one negative vote ("very bad") can be found (see Figure 5.5(a)) and this participant would not use the component for own analyses. All other participants rated the representation of the video player component positively and would use it for own analyses. When we take a look at the textual feedback of the participant with the negative vote we can see that he/she expects highest standards from a tool like SportSense Football with respect to broadcast. One interpretation of this statement can be that only one video source was provided during the user studies. However, in practice often several video sources are required (e.g., tactics cam, TV broadcasting view, etc.). In our future work, we plan to extend SportSense Football with the option to dynamically switch between different video sources. Another interpretation of the participant's statement is that the fast-forward and rewind functionality of the video player is too coarse and should be more fine. We will consider this point in our future work as well to make the UX even more pleasant.

The clarity of the arrangement of the three main components was rated mainly positive by the participants (see Figure 5.5(b)). Again, one negative vote ("very bad") can be observed. From the corresponding textual feedback one can conclude that a user dashboard is required which allows a flexible and individual arrangement to the user's needs. For all other participants the arrangement of SportSense Football's main components is harmonious.

This looks a bit different for the clarity of the arrangement of all components. Here two negative votes ("very bad","bad") stick out (see Figure 5.5(b)). According to one participant clarity gets lost when more filters are selected in the filter area of the SportSense Football UI. Here, it can help to have a small info box which displays all currently defined query specifications. We will consider this for our future work. Nevertheless, two thirds of all participants rated the clarity of the arrangement as currently implemented in SportSense Football positively. Again, one can summarize that user needs are individual and that it is difficult to respect the different demands [Gar12].

5.1.3.6 Part 6 - Final Questions

In the last part of the questionnaire participants were asked about their overall impression about the SportSense Football system. The results are depicted in Figure 5.6. 41.7% (n=5) of the participants rated their impression about the system as very good, 50% (n=6) as good, and 8.3% (n=1) as neutral.

Participants also had the chance to name additional features which they would like to have supported by SportSense Football. The results are listed as bullet points in the following:

– measure and visualize distances between players

– division of the playing field into different zones (e.g., 3 zones, 5 lanes) and visualize, e.g., possession gains in each zone

– implement playing direction

– integrate jersey numbers in 2D-representation

– link pressing index and other raw data visualizations (e.g., speed) to the video

– make the system more customizable

– consider context of events

– select principles

Figure 5.6 Results of the overall impression rating of SportSense Football.

5.1.4 Discussion

Having presented the results of the user studies in the previous section we now discuss these results with respect to the hypotheses we introduced in Section 5.1.1.

Hypothesis 1. The majority of participants of our user studies uses at least two software tools for the match preparation process. All participants confirmed that they would profit from a system which combines different analysis options. SportSense Football provides exactly such a combination of different analysis options and therefore allows a very efficient working. This also means that users can save a lot of time. This in turn is highly valuable as many coaches and analysts have to work under enormous time pressure (see Section 1.2). Therefore, we can confirm our first hypothesis.

Hypothesis 2. The positive results presented in Section 5.1.3.3 confirm also our second hypothesis. The majority of the participants would use each of the different query types for their own analyses. All sports data which participants use for their analyses can be retrieved with these query types. Thus, the query types are very valuable in practice as they can be used for a flexible and holistic retrieval and analysis of relevant sports data.

Hypothesis 3. The results presented in Section 5.1.3.4 show that the football-specific concepts which we already implemented are adequately represented in the analysis features of SportSense Football. Therefore, these results confirm also our third hypothesis. With the implemented features users are able to conduct high-level tactical analyses which is particularly important in the daily work of analysts (see Section 1.2). These results also confirm that our approach to bridge the semantic gap of sports analytics as presented in Section 3.1 works well.

Hypothesis 4. The mainly positive evaluation results on the representation of SportSense Football's main components, as well as on the arrangement of the components confirm the successful planning and designing steps for a good UX and show that the approach as presented in Section 3.4 works well. Having regard to that design and user needs are always individual we nevertheless can confirm our fourth hypothesis based on the results presented in Section 5.1.3.5.

Summary. The participants' overall rating of SportSense Football was positively (see Figure 5.6). Together with the other results of the user studies, the conception, design, and implementation of the system can be regarded as a success. With SportSense Football users have a DAS which provides a good decision support in the context of match preparation. The positive results of the user studies are even more valuable as participants were very critical and (partly) worked in top-level clubs with highest standards.

5.2 Performance Evaluation

Moving on now to evaluate the retrieval performance and the scalablity of the SportSense Football system. With this performance evaluation we want to check whether SportSense Football is utilizable in football practice. For that, we first check the retrieval performance of the system for a single match. Second, we check if the system scales for seven matches. In our user studies this was the maximum number of matches which are analyzed for the preparation of an upcoming match (see Section 5.1.3.2). Consequently, we do evaluate the performance of SportSense Football for a "real-world" scenario of football practice.

In the following, we first present the hypotheses of the performance evaluation in Section 5.2.1. Afterwards, we show how the data for the evaluation process are generated in Section 5.2.2. The setup and the results of the retrieval performance measurement are presented in Section 5.2.3. Furthermore, the setup and the results of the scalability measurement are presented in Section 5.2.4. In

Section 5.2.5 we finally discuss the results of the performance evaluation with respect to the formulated hypotheses.

5.2.1 Hypotheses

As we already stated in Section 5.1.1 we developed SportSense Football with the intention to support football analysts and coaches in their decision making processes in sports practice. Hereby, the focus lies on the match preparation context. A good support, of course, also requires a pleasant user interaction with the SportSense Football system. Among other things, this means that the response time, i.e., the amount of time from when a request is sent until the results are returned, should be very short. This is important to keep the user focused and in the flow. For the classification of our evaluation results we use the following four categories of system response times introduced by Doherty and Sorenson [DS15] in which users still remain attentive: (1) instantaneous (< 300 ms), (2) immediate (300 ms - 1 s), (3) transient (1 s - 5 s), and (4) attention span (5 s - 10 s). With the performance evaluation we want to check the following two hypotheses:

Hypothesis 1. The retrieval performance of SportSense Football for one match is very good, i.e., query results are returned immediately (i.e., in less than 1 s).

Hypothesis 2. SportSense Football scales for seven matches, i.e., the system scales for a "real-world" scenario of football practice.

5.2.2 Data Generation

For the performance evaluation of SportSense Football we generated data with the StreamTeam Football system which is deployed on a cluster consisting of six homogeneous machines. All machines are equipped with an Intel Core i7-4770 CPU, 32GB RAM, and two SSDs and run Ubuntu 16.04.32 as the operating system [Pro20].

On the basis of a commercial provider's spatio-temporal tracking dataset from a high professional football match, which is captured with 25 Hz, the StreamTeam Football system automatically generates event data, states, and statistics and stores them in a MongoDB instance. A screenshot of the UI of the StreamTeam Football Web client is depicted in Figure 5.7. So far, StreamTeam

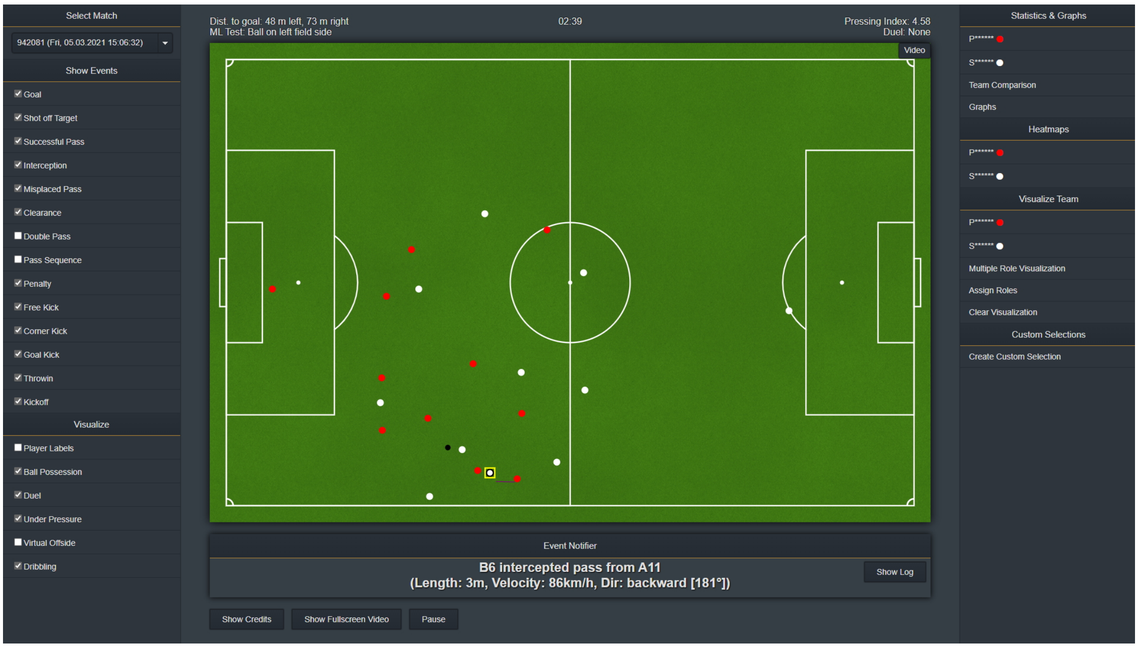


Figure 5.7 UI of the StreamTeam Football Web client.
Relevant match actions like events, which are detected by the system, are visualized in real-time. Furthermore, users can decide which information should be visualized by activating/deactivating diverse checkboxes.

Football analyzes the first half of a match. Thus, to get (dummy) data for the second half, we duplicated all database items from the first half and added 45 minutes to each item's timestamp. With that we successfully simulated one full match.

To now mimmick a "real-world" scenario of football practice, where more match data are available and used for the analysis process, we repeated the same procedure as described above ten times with different match identifiers. After each round we created a new MongoDB dump. This leads to a total of ten database dumps, of which the first dump contains data of only one match, the second dump contains data of two matches, etc.

One machine of the StreamTeam Football cluster setup also runs the MongoDB REST proxy. We used scripts to automate the performance evaluation, i.e., the retrieval performance and scalability measurements on that machine (see Sections 5.2.3.1 and 5.2.4.1). The evaluation scripts are published on GitHub under the GNU Affero General Public License v3.0 (see Appendix B).

5.2.3 Retrieval Performance Measurement

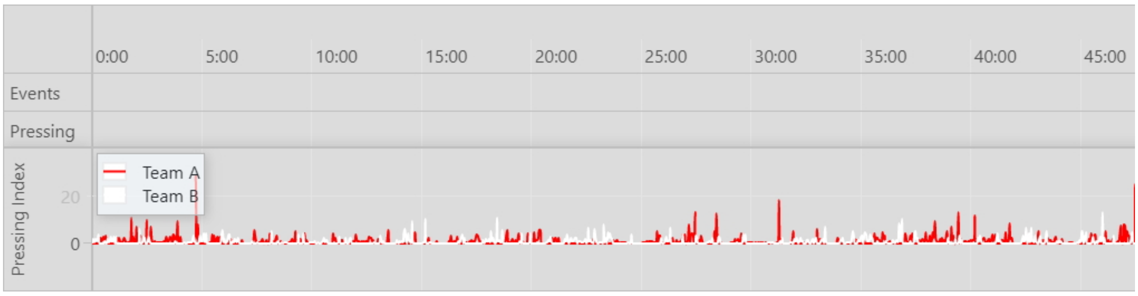
In this section we present the setup (see Section 5.2.3.1) and the results (see Section 5.2.3.2) of the first part of the performance evaluation of the SportSense

FOOTBALL system, the retrieval performance measurement.

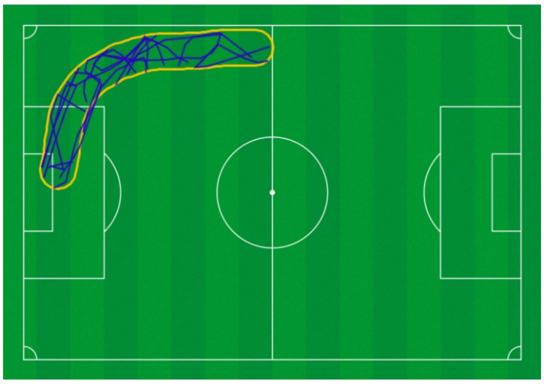
5.2.3.1 Setup

We evaluate the retrieval performance of SPORTSENSE FOOTBALL for one single match. More precisely, we evaluate the retrieval performance of each of the four different query types which we introduced in Section 3.3.1 and which are implemented in the system: (1) raw data query, (2) (raw) movement query, (3) event query, and (4) event pattern query. Additionally, we evaluate the retrieval performance of a quantitative analysis. A qualitative analysis is not part of the evaluation as video data are not stored in the MongoDB instance (see Section 4.1.1.1). For the evaluation we use one example for each query type and the quantitative analysis which all represent typical and useful queries for football practice.

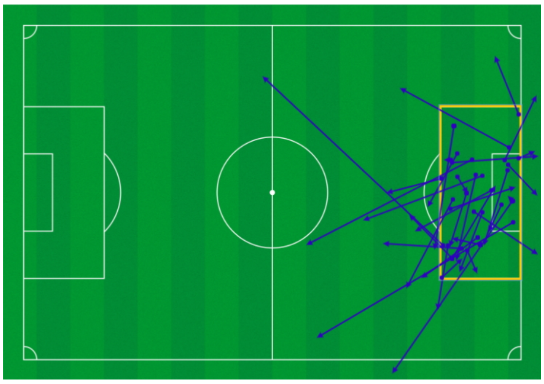
The pressing index feature of SPORTSENSE FOOTBALL (see Section 4.2.2.1) is used to evaluate the raw data query. Here, the pressing index values of both teams are retrieved (see Figure 5.8(a)). This query is useful for the analysis of a team's pressing behaviour during the match. To evaluate the raw movement query all ball trajectories which fit the sketched path (i.e., the orange polygon) are retrieved (see Figure 5.8(b)). This query is useful to search for a specific play which the coach taught in the last training sessions. For the evaluation of the event query, we specify the penalty box on the right side of the pitch (i.e., the orange box) and retrieve all events happened there (see Figure 5.8(c)). Events which happened inside the penalty box are often dangerous and thus worth to be analyzed in detail. Furthermore, we evaluate both implemented event pattern queries. First, a forward event cascade which consists of three steps. Here, all patterns are retrieved in which the first event happened in the rectangular area inside the right penalty box, the second event happened within five seconds after that event and inside the circular area on the lower part of the pitch, and the third event happened within five seconds after the second event and inside the rectangular area in front of the left penalty box (see Figure 5.8(d)). Second, a reverse event cascade which also consists of three steps. Here, all event patterns are retrieved in which the last event happened inside the defined rectangle within the left penalty box, and the second last event happened within five seconds before the last event and within the rectangular area on the top left side of the pitch (see Figure 5.8(e)). The forward and reverse event cascades are used to find specific plays which were part of the match plan. Thus, coaches and analysts can check if the team implemented the instructions. A formal



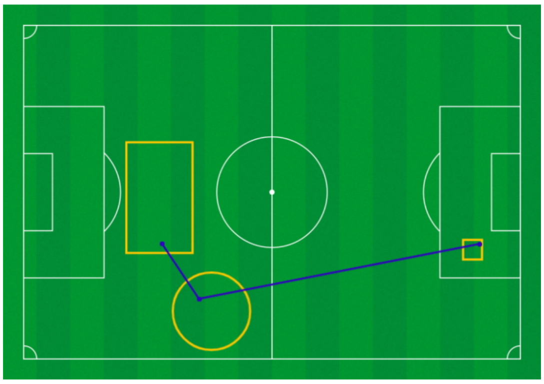
(a) Raw data query.



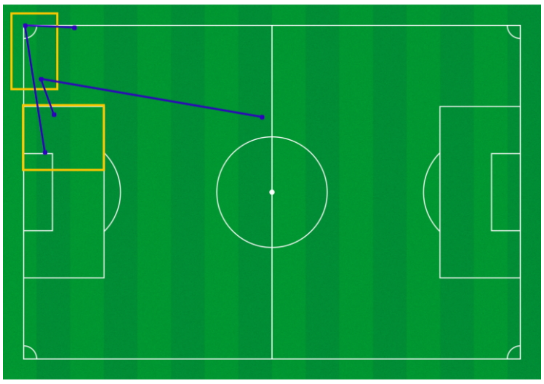
(b) Raw movement query.



(c) Event query.



(d) Event pattern query - Forward event cascade.



(e) Event pattern query - Reverse event cascade.

(f) Quantitative analysis.

Figure 5.8 Screenshots of the queries used for the performance evaluation.

Table 5.1 Formal query specifications of the evaluated query types.

Query Type	Query Specifications
Raw Data Query	$q_{sp}.R.J = "pressingState" \wedge q_{sp}.CID = "XYZ"$
Raw Movement Query	$q_{sp}.R.J = "TD" \wedge q_{sp}.AID = "Ball" \wedge q_{sp}.LOC = ((x_1, y_1), ..., (x_n, y_n))$
Event Query	$q_{sp}.LOC = ((x_1, y_1), (x_2, y_2))$
Event Pattern Query -	$q_{sp_1}.LOC = ((x_1, y_1), (x_2, y_2)) \wedge q_{sp_2}.LOC = ((x_3, y_3), radius) \wedge$
Forward Event Cascade	$q_{sp_3}.LOC = ((x_4, y_4), (x_5, y_5)) \wedge \delta_{after} = 5$
Event Pattern Query -	$q_{sp_1}.LOC = ((x_1, y_1), (x_2, y_2)) \wedge q_{sp_2}.LOC = ((x_3, y_3), (x_4, y_4)) \wedge$
Reverse Event Cascade	$\delta_{before} = 5$

representation of the query specifications of all evaluated queries is listed in Table 5.1. Finally, a quantitative analysis is evaluated as well. Here, we specify all performance parameters which are implemented in SportSense Football to retrieve data of four players (see Figure 5.8(f)). This is used by coaches and analysts to compare the players' performances quantitatively based on statistics and indicators.

The measurement procedure is similar to the one which was conducted to evaluate a previous version of SportSense [PRS+18] and works as follows. We captured the MongoDB REST proxy calls (i.e., the HTTP GET requests) which were issued during each of the different queries and the quantitative analysis. The exact calls which were issued are listed in Tables D.2 to D.9. Then, we measured the time between issuing the call and receiving the results. Here, we want to mention that the latency of the network is part of the measurement. However, this is inevitable. In contrast to that, the time which is needed to visualize the results is not part of the measurement. To measure the duration of a call we used cURL[3]. Each call was measured 40 times and the mean value and the standard deviation of these measurements were calculated.

5.2.3.2 Results

In this section we present the results of SportSense Football's retrieval performance measurements for the four different query types and the quantitative analysis. Tables 5.2 and 5.3 show the average retrieval time and the standard deviation (SD) for the different queries and the quantitative analysis. In the following, we take a closer look at the individual measurement results.

[3] `https://curl.haxx.se`

Table 5.2 Average retrieval time of evaluated queries for a single match. The table lists the average retrieval time (in milliseconds) and the standard deviation (SD) of the evaluated raw data query, raw movement query, event query, and of the quantitative analysis for one match.

Type	Time	SD
Raw Data Query	5,563.93	60.94
Raw Movement Query	457.88	98.13
Event Query	106.13	25.56
Quantitative Analysis	116.20	9.43

Raw Data Query. The first evaluated query type is the raw data query. The average retrieval time for the query was 5,563.93 ms (SD: 60.94 ms).

Raw Movement Query. Second, the raw movement query is evaluated. The average retrieval time for this query was 457.88 ms (SD: 98.13 ms).

Event Query. The third evaluated query type is the event query. The average retrieval time for the query was 106.13 ms (SD: 25.56 ms).

Event Pattern Query. Fourth, two event pattern queries are evaluated. The average retrieval time for the whole forward event cascade was 163.83 ms which results from the addition of the retrieval times of the first step (50.63 ms, SD: 3.24 ms), the second step (68.00 ms, SD: 5.18 ms), and the third step (45.20 ms, SD: 2.97 ms). The average retrieval time for the reverse event cascade was 171.70 ms. This value results from the addition of the average retrieval times of the first step (49.40 ms, SD: 3.24 ms), the second step (73.40 ms, SD: 7.72 ms), and the third step (48.90 ms, SD: 3.88 ms).

Quantitative Analysis. Finally, a quantitative analysis is evaluated. The average retrieval time was 116.20 ms (SD: 9.43 ms).

Table 5.3 Average retrieval time of evaluated event cascades for a single match. The table lists the average retrieval time (in milliseconds) and the standard deviation (SD) of the two evaluated event pattern queries for one match.

Type	Total	Call 1	SD	Call 2	SD	Call 3	SD
Event Pattern Query - Forward Event Cascade	163.83	50.63	3.24	68.00	5.18	45.20	2.97
Event Pattern Query - Reverse Event Cascade	171.70	49.40	3.24	73.40	7.72	48.90	3.88

Discussion. The call for the raw data query takes significantly longer compared to the other query types. With the response time being within the range of five to ten seconds the result however still lies within the user's "attention span" according to the categorization of Doherty and Sorenson (see Section 5.2.1). Nevertheless, the call should take as less time as possible to maximize the work efficiency of the user. The call takes longer because the whole *states* collection (see Section 4.1.1.2) is iterated to retrieve the pressing index values of both teams. However, an enhancement of the retrieval performance can easily be achieved. Currently, every second pressing state document from the collection is used to extract the pressing index for the final result. To reduce the retrieval time, it could already be sufficient to use a longer time interval (e.g., every fifth pressing state document) for the extraction of the pressing indexes. This is a quick and easy fix in the code of SportSense Football. Apart from this result, all other retrieval times are very fast and lie below the threshold of one second which means an "immediate" system response time according to the categorization of Doherty and Sorenson. Additionally, four measurements even lie below the threshold of 300 ms which means that the response time is "instantaneous" for the respective queries. In summary, these results confirm a very good retrieval performance and thus a pleasant user interaction with SportSense Football.

5.2.4 Scalability Measurement

In this section we present the setup (see Section 5.2.4.1) and the results (see Section 5.2.4.2) of the second part of the performance evaluation of the SportSense Football system, the scalability measurement.

5.2.4.1 Setup

As already mentioned at the beginning of Section 5.2, we want to check with the performance evaluation if SportSense Football is utilizable in sports practice and thus also if the system scales for a "real-world" scenario of seven analyzed matches. However, we evaluate the scalability of SportSense Football for up to ten matches and thus for even more matches than in the scenario above. This is important as other coaches or analysts might analyze even more matches than those who participated in our user studies.

The scalability measurement procedure is equal to the measurement of the retrieval performance presented in Section 5.2.3.1, just for more matches. However, we do not evaluate the scalability of the raw data query. This is because in football practice, the analysis of the pressing index values is only interesting for one specific match. Aggregated information on the pressing index like, for instance, the average number of pressing phases per match or the average exerted pressure, are covered anyway with the quantitative analysis of SportSense Football.

Additionally, the evaluation of the scalability of the event pattern queries (i.e., for forward and reverse event cascades) is not required. Only the first step of an event pattern query would be useful for the evaluation of the scalability. In reality however, there will never be two completely equal football matches where exactly the same happens. Also the match plan might differ fundamentally from match to match depending on the opponent. So it might be the case that for one match the coach wants his team to play over the wings whereas in the next match he wants the team to attack more central. Consequently, the event patterns would differ significantly and would not be comparable. Thus, the inclusion of event pattern queries in the scalability measurement is not necessary because we would not measure what we want. As the first step of an event pattern query is equal to a normal event query, we decided to not evaluate event pattern queries with respect to the scalability.

As already mentioned in Section 5.2.3.1 a qualitative analysis is not part of the performance evaluation because video data are not stored in the MongoDB. Therefore, a qualitative analysis is also not part of the scalability measurement.

In summary, two query types (i.e., the raw movement query and the event query) as well as the quantitative analysis remain for the evaluation of the scalability of SportSense Football. The respective MongoDB REST proxy calls which were issued are listed in Tables D.2 and D.3. As previously mentioned we use the same measurement procedure as described in Section 5.2.3.1 but this time for up to ten matches. This means that we measured the time of each MongoDB REST proxy call until the results are received for each number of matches.

5.2.4.2 Results

The results of the scalability measurement are depicted in Tables 5.4 to 5.6 and visualized in Figures 5.9 and 5.10. In the following, we take a detailed look at the results of each evaluated query and the quantitative analysis.

Raw Movement Query. The average retrieval time of the raw movement query for ten matches is 19.48 times higher than the average retrieval time for one match. When we take a look at Table 5.4 and Figure 5.9 we can see that the

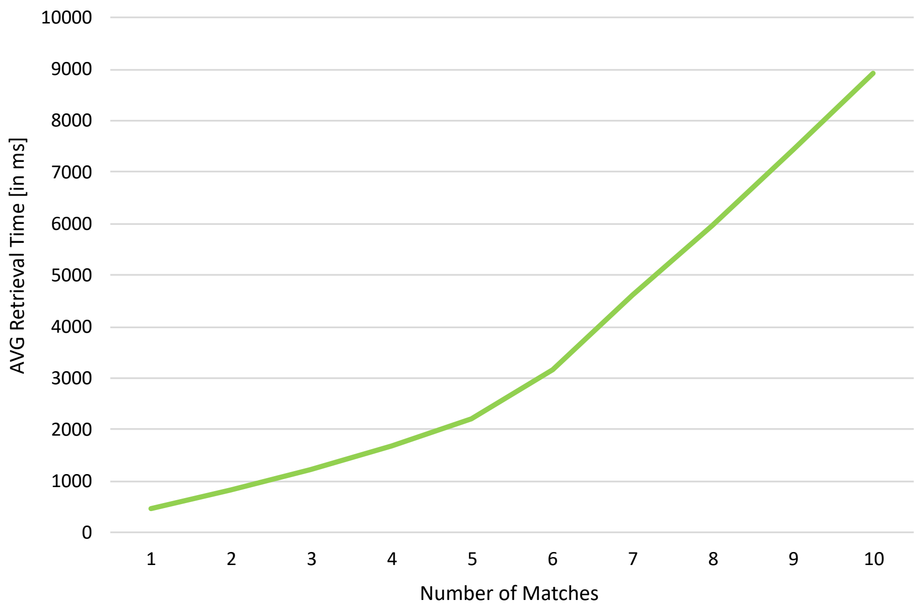


Figure 5.9 Scalability of the raw movement query w.r.t. the number of matches.

average retrieval times for up to five matches follow a linear development. From six to ten matches, the average retrieval times increase in bigger steps. However, this development again is linearly. We can see these developments also in the times per analyzed match. When analyzing up to five matches, the time to retrieve the results per match remains almost constant. From six matches we observe an increase. For ten matches, the time to retrieve the results per match

Table 5.4 Average retrieval times of the raw movement query.
The table lists the average retrieval time (in milliseconds) and the standard deviation (SD) of the raw movement query w.r.t. the number of matches. Additionally, the average retrieval times divided by the number of matches are listed as absolute values and in percentage to the average retrieval time for one match.

# Matches	Mean	SD	Time per Match (abs)	Time per Match (%)
1	457.88	98.13	457.88	100.00
2	835.48	199.04	417.74	91.23
3	1,225.93	283.89	408.64	89.25
4	1,682.73	497.39	420.68	91.88
5	2,197.80	720.02	439.56	96.00
6	3,169.98	1,001.91	528.33	115.39
7	4,621.53	1,044.25	660.22	144.19
8	5,948.28	922.97	743.53	162.39
9	7,438.03	1,136.82	826.45	180.50
10	8,920.58	1,281.36	892.06	194.83

Table 5.5 Average retrieval times of the event query.
The table lists the average retrieval time (in milliseconds) and the standard deviation (SD) of the event query w.r.t. the number of matches. Additionally, the average retrieval times divided by the number of matches are listed as absolute values and in percentage to the average retrieval time for one match.

# Matches	Mean	SD	Time per Match (abs)	Time per Match (%)
1	106.13	25.56	106.13	100.00
2	123.88	8.05	61.94	58.36
3	141.98	9.54	47.33	44.59
4	171.10	11.94	42.78	40.31
5	198.45	18.64	39.69	37.40
6	221.08	24.49	36.85	34.72
7	247.88	30.06	35.41	33.37
8	268.60	29.44	33.58	31.64
9	268.20	42.98	29.80	28.08
10	347.20	130.32	34.72	32.72

is almost twice the time to retrieve the results when analyzing only one match.

Event Query. The average retrieval time of the event query for ten matches is 3.27 times higher than the average retrieval time for one match. When we take a look at Table 5.5 and Figure 5.10 we observe that the average retrieval time scales linearly with the number of matches. The times per analyzed match show that from three to ten matches, the time to retrieve the results per match is less than half of the time to retrieve the results when analyzing only one match.

Quantitative Analysis. The average retrieval time of the quantitative analysis for ten matches is 4.24 times higher than the average retrieval time for one match. Similar to the event query results, we observe that the average retrieval time of the quantitative analysis scales linearly with the number of matches (see Table 5.6 and Figure 5.10). The times per analyzed match show that from five to ten matches, the time to retrieve the results per match is less than half of the time to retrieve the results when analyzing only one match.

Discussion. In summary, we can state that SportSense Football can be used to retrieve football data (i.e., event data and spatio-temporal tracking data) of multiple matches. The raw movement query scales linearly for up to five matches. However, from six to ten matches the scalability gets a bit worse. The average retrieval time for ten matches is almost 20 times higher than for one match. When analyzing more matches, the average retrieval time of the

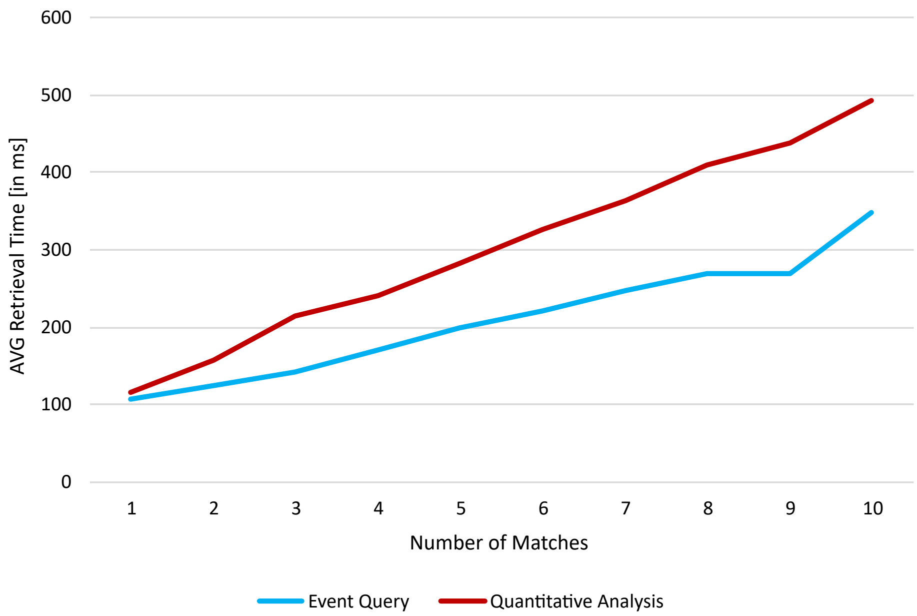


Figure 5.10 Scalability of the event query and the quantitative analysis w.r.t. the number of matches.

raw movement query reaches almost the threshold of ten seconds which is the maximum value of a user's "attention span" according to the categorization of Doherty and Sorenson (see Section 5.2.1). Here, optimizations are required to reduce the retrieval time of raw movement queries and to keep a linear scalability even for ten matches. This will ensure a pleasant user interaction with

Table 5.6 Average retrieval times of the quantitative analysis.
The table lists the average retrieval time (in milliseconds) and the standard deviation (SD) of the quantitative analysis w.r.t. the number of matches. Additionally, the average retrieval times divided by the number of matches are listed as absolute values and in percentage to the average retrieval time for one match.

# Matches	Mean	SD	Time per Match (abs)	Time per Match (%)
1	116.20	9.43	116.20	100.00
2	158.00	9.49	79.00	67.99
3	213.80	23.00	71.27	61.33
4	240.13	32.58	60.03	51.66
5	281.33	47.61	56.27	48.42
6	326.48	58.94	54.41	46.83
7	363.00	69.72	51.86	44.63
8	410.20	76.71	51.28	44.13
9	439.10	101.92	48.79	41.99
10	492.30	154.93	49.23	42.37

SportSense Football even when more matches are considered in the analysis process. Apart from these results, the retrieval times of the event query and the quantitative analysis look very good. The event query and the quantitative analysis both scale linearly for ten matches. Query results are returned “immediately” in less than a second even if ten matches are analyzed.

5.2.5 Discussion

Having presented the results of the performance evaluation in the previous sections we now discuss these results with respect to the hypotheses we introduced in Section 5.2.1.

Hypothesis 1. The results of SportSense Football's retrieval performance measurements for a single match as presented in Section 5.2.3.2 are very good. Except for the raw data query, all other evaluated query types and the quantitative analysis showed very fast retrieval times below the threshold of one second which means an “immediate” response time according to the categorization of Doherty and Sorenson. Thus, we mainly can confirm our first hypothesis despite an optimization for the raw data query is required.

Hypothesis 2. The results of the scalability measurements which we presented in Section 5.2.4.2 mainly confirm also our second hypothesis. The raw movement query scales linearly for up to five matches. Here, optimizations are required to reduce the retrieval times and to ensure a linear scalability also for seven matches. The event query and the quantitative analysis scale linearly with the number of matches. This is even the case for ten matches and not just for seven matches as formulated in the second hypothesis.

Summary. To conclude, we can state that the implementation of SportSense Football was successful. The retrieval performance and scalability measurements showed mainly very positive results. The weak points, namely the retrieval performance of the raw data query and the scalablity of the raw movement query should be considered and optimized in future work. If this is successful, then SportSense Football will be a very powerful tool and allow a pleasant user interaction in “real-world” scenarios of football practice.

6 Related Work

No research without action, no action without research.

— Kurt Lewin

The daily work of game sports analysts and coaches is very demanding and lots of decisions have to be made. Various factors are influencing and complicating the decision making processes. As we have presented in Section 2.1.1, interactions are one of these factors. Interactions complicate the drawing of conclusions on an individual's contribution to the overall (tactical) team performance. To know about these contributions, however, would be very important for analysts and coaches before they decide, for instance, on the line-up for an upcoming match. Physiological factors like health status, current fitness values, or stress levels of athletes are also important to know when it comes to decision making processes of analysts and coaches. Both, research and industry, thus try to support sports practice with different approaches and solutions.

In this section we take a closer look at both areas. First, we show some work from sport science and computer science researchers and research groups, which develop, for instance, new KPIs and innovative software tools (see Section 6.1). Second, we present some industrial solutions which are currently used in sports practice (see Section 6.2). By highlighting these two perspectives we see interesting approaches with a considerable potential to support analysts and coaches in their daily tasks and with their decision making processes. However, we also realize that there is no perfect solution for decision support in game sports so far.

6.1 Research Approaches

In this section we present different approaches from academia, involving sport scientific work as well as computer scientific work. Hereby, we mainly focus on two aspects. First we show some approaches which try to better evaluate athletes' and/or teams' (tactical) performances (see Section 6.1.1). Second, we present approaches which try to facilitate the analysis process and try to make it more time-efficient (see Section 6.1.2).

6.1.1 Sports Performance Analysis

The field of sports performance is actively researched [GH17]. Among other things, this is because today, most teams are collecting large amounts of data from training sessions and matches and are interested in exploiting these data in order to gain an advantage over their opponents [DVD18]. Even as statistics are known to be the most commonly used tool for the analysis of sports data [LK18] these nevertheless are mostly not useful to allow valid statements on players' or teams' (tactical) performances. This is confirmed by Cintia et al. [CGP+15]:

> "There is not yet a consolidated repertoire of statistics that are accepted as reference indicators for the various facets of team performance."

Moreover, the low scoring nature of some invasion games like football [AYS07] makes it hard to evaluate which team performed better and why a match was won or lost. Players' individual performances in collective sports furthermore must always be regarded within the specific context in which they are produced [PRB17]. So far, the evaluation of tactical performances thus remained a very subjective task. However, different types of data are nowadays generated and available for the analysis as we presented in Section 2.2.1. In the following, we thus present data-driven approaches which try to better evaluate players' or teams' (tactical) performances quantitatively (see Section 6.1.1.1). Additionally, we present approaches focusing more on visualizations to enhance and support the subjective analysis of tactical performances (see Section 6.1.1.2).

6.1.1.1 KPIs and Metrics

There are numerous works which present new KPIs and metrics to quantitatively represent player or team performances in game sports. We present some selected approaches with a great potential for an application in sports practice.

The work of Link et al [LLS16] shows an approach to quantify the attacking performance of teams in football on the basis of spatio-temporal tracking data. The model considers different factors (i.e, zone, control, pressure, and density) to calculate a so called dangerousity value which represents the probability of scoring a goal for every point in time a player has possession of the ball. Based on this value several other analysis options arise. One can derive, for instance, a team's dominance during a match as well as a player's individual contribution to create dangerous situations.

Cervone et al. [CDB+14] present the so called Expected Possession Value (EPV) using a Markovian assumption to evaluate player decisions in basketball, also based on spatio-temporal tracking data. Similar to the dangerousity metric presented above, EPV also represents a value which is computed and assigned to a player/team at each moment of a possession phase. In this case, the value is the number of points which the offensive team is expected to score by the end of the possession, taking into account the current spatial configuration of all players and the ball. The evolution of the EPV in turn allows coaches and analysts to quantitatively evaluate individual player actions.

Another approach which is very similar to the one of Cervone et al. is presented by Schulte et al. [SKG+17]. To evaluate player actions and team performance in ice hockey the authors also use a Markov Game model in which the context of an action is considered. The model includes "look-ahead" to evaluate long-term impacts of an action which might not immediately lead to a visible outcome like, e.g., a goal. This in turn allows then to evaluate quantitatively all on-ice actions and not just shots and goals as this is often the case for the evaluation of player performances in ice hockey.

There are many other interesting KPIs and metrics which would be worth to be presented in more detail. Just to mention a few approaches, for instance, from football, these include the off-ball scoring opportunity metric from Spearman [Spe18], an application of EPV presented by Fernández et al. [FBC19], or the off-ball advantage metric introduced by Llana et al. [LMF20]. With the spatial centre and stretch index, Bourbousson et al. [BSM10] present two metrics to measure space and time dynamics of basketball teams. Moreover, (social) network approaches are widely researched in academia to evaluate player and team performances and interactions [DWA10; CCM+15; CMK+15; RLA18; MR18].

6.1.1.2 Visualizations

In the following we present selected work with which researchers try to enhance the subjective evaluation of player and team performances by visualizing complex context information.

A data-driven "ghosting" approach to analyze the defensive behaviour of football teams is presented by Le et al. [LCY+17]. This approach allows complex movement analyses in which movements of individual players can be compared to average movements of "ghosts", i.e., players of the "league average" team, or a user-defined team, respectively. By visualizing the actual movements and the movements of the ghosts, the defensive behaviour of individual players or the whole team can better be evaluated.

The approach of Fernández and Bornn [FB18] introduces open spaces in football. Based on the authors' space control model two metrics can be calculated, namely space generation gain and space occupation gain. The corresponding visualizations on a 2D playing field can help coaches and analysts to better evaluate player and team performances.

The approach of Stein et al. [SJL+17] even integrates several visualizations directly in the video footage which football coaches and analysts use for their qualitative video analysis. These visualizations include, for instance, so called dominant regions, pass distances, player movements, or player reactions. With these enhancements, coaches and analysts get a valuable support and can better evaluate player and team performances qualitatively.

Many other examples can be found in which complex context information is visualized. Just to mention a few, these include the visualization of pressure [AAB+17], group or team centroids [BTL+18], or faulty movement behaviour [SSM+19].

6.1.2 Optimizing the Analysis Process

As we have introduced in Section 1.2 performance analysts need to work under an enormous time pressure. Therefore, academia also tries to provide support by automating parts of the analysis process (see Section 6.1.2.1) and developing new tools to allow for a flexible and time-efficient working (see Section 6.1.2.2).

6.1.2.1 Automation of the Analysis Process

In this section, we present selected work from researchers and research groups who try to automate parts of the analysis process with the goal to support the tedious and time-intensive task of performance analysts.

First, we again want to highlight the StreamTeam system [PBS+17; PRS+18; Pro20; PSS+20]. This worker-based data stream analysis system has an application in football and is able to detect and calculate events, states, and statistics on the basis of spatio-temporal raw tracking data in real-time. The definitions of the events, states, and statistics which are detected by the system were developed in close collaboration with sport scientists. An automatic event detection can help a lot as the tedious task of tagging matches live by human operators (e.g., analysts) is no longer necessary. Real-time analysis is extremely valuable for players and coaches because this helps them to maximize their potential and performances [Boj14].

Salim et al. [SHT+19] present another system which detects events in volleyball based on IMU data and automatically supplements video recordings. This allows coaches and analysts to focus on the match and to not lose concentration due to tagging tasks. Additionally, the web application of the system allows users to filter and search for actions, and to watch the video scenes by clicking on the respective item.

The approach of Decroos et al. [DVD18] focuses on the detection of team tactics in professional football matches like, for instance, common offensive strategies applied by teams. The authors used event data from a whole season as a basis for their model. For analysts and coaches this approach can be a huge advantage when they prepare for upcoming matches because the characteristics of opponents can quickly be analyzed.

Academic approaches to automatically measure and identify team formations in football based on spatio-temporal tracking data are presented, for instance, by Bialkowski et al. [BLC+14] or Shaw and Glickman [SG19]. The results can be used to describe and analyze team behaviour, which would be very valuable for analysts and coaches in sports practice.

Many other interesting work exists which would be worth to be presented here in more detail. There are, for instance, machine learning approaches in football to detect different types of ball possession [LH17], to identify counter-attacks or counter-pressing based on spatio-temporal tracking data [HPS+18], or to develop so called movement models [BLM19]. The identification of playing styles of tennis players as well as the prediction of the location of shots based on spatio-temporal data is shown in the work of Wei et al. [WLM+13]. The automatic detection of events, identification of players, as well as the classification of teams with examples from ice hockey and basketball is presented by Mehrasa et al. [MZT+18].

6.1.2.2 Innovative Analysis Tools and Techniques

Automation is only one option to optimize the analysis process. Another option is to make the analysis process more flexible and the information retrieval more intuitive. Here, also some academic approaches exist which propose new tools and techniques. In the following, we present selected work from these areas.

Pileggi et al. [PSB+12] present the system SnapShot for ice hockey analyses. This system was developed in close collaboration with a domain expert, i.e., a professional ice hockey analyst. Among other functionalities SnapShot provides users with the option to visualize shot data. A special feature, the so called radial heatmap, is introduced. This feature considers context information of the shot, in this case the shot length. Users therefore are able to not just quantitatively analyze shot data, but to also use the tool to visualize information on a 2D representation of the ice hockey rink.

Another software prototype is introduced by Benito Santos et al. [BTL+18]. The authors present a system for football, which supports various functionalities for collective group and team analyses. This includes, for instance, the calculation and visualization of the team center, team dispersion, or team synchrony. Also individual player analyses (e.g., heatmaps) are supported. Analysts and coaches can profit from the system because the analysis process gets more flexible and efficient.

A decision support tool to evaluate player performances in rugby union is presented by Calder and Durbach [CD15]. A tool to interactively perform exploratory analyses of football data is presented by Delibas et al. [DUI+19]. These tools represent only some examples and are just a small excerpt of many other innovative tools which are developed by academia. To just name a few other systems, there are, e.g., SoccerStories [PVF13], ForVizor [WXW+18], or PassVizor [XWL+21].

In the last part of this section we briefly present some approaches from academia with which researchers try to make the analysis process of coaches and analysts more intuitive. An elementary challenge here is how to design a query format in a way that is both, intuitive to use and rich enough to allow for precisely specify the information need [SLY+16]. This can be achieved, for instance, with the introduction of sketch-based methods for the information retrieval.

Sha et al. [SLY+18] introduce an intelligent interface which allows trajectory-based retrieval of specific plays and an interactive player and team analysis. One option to search for a certain play is to draw the corresponding query by hand on a "chalkboard". Another option for the retrieval is to sketch directly on the video footage (i.e., the broadcasting feed). The authors present their approach in the context of competitive basketball. Analysts and coaches can profit a lot as the search for specific plays is facilitated and more intuitive.

Another approach to simplify the time-intensive task of analysts and coaches in football is presented by Richly [Ric18]. Here, video data are combined with spatio-temporal tracking data. Users can query for patterns in a graphical, sketch-based manner by drawing on a virtual and interactive tactics board. This facilitates the analysis process of coaches and analysts because they can search for specific plays in an intuitive way. Additionally, the scenes can be watched directly in the video and thus analysts and coaches do not need to watch the whole video material of the match(es).

A similar approach in which users can search for football specific patterns is shown by Stein et al. [SJS+19]. Queries can be formulated interactively, again in a sketch-based manner. Additional filter options exist to further delimit the query results like, for instance, the specification of events, players, or pressure. This again allows the user an interactive and intuitive retrieval of relevant data for their analyses.

6.1.3 Discussion

In the previous sections we presented different academic approaches with which researchers try to support game sports analysts and coaches with their daily work. However, most of these promising approaches do not find their way into sports practice. This can be due to several reasons which we present in the following.

The first option to explain this problem is the semantic gap of sports analytics (see Section 2.1.3). Despite the academic concepts and technologies are very powerful, these often are developed without including coaches or other domain experts in the process. This in turn leads to a decreasing interest by sports practitioners because the understanding is missing and/or own concepts of domain experts are not included in the respective model. Researchers should thus respect, that the impact of any approach will likely be best achieved if the collaboration with coaches is very close [Cou14]. This can be achieved with our approach to bridge the semantic gap of sports analytics (see Section 3.1). The positive results of the user studies presented in Section 5.1 emphasize that our approach successfully can be applied in practice.

The second option which could explain that the approaches of academia often are not applied in sports practice are the limitations and weaknesses of some approaches. Despite some tools like SnapShot, presented by Pileggi et al., provide very powerful analysis options, the analysis process nevertheless is often limited. In the case of SnapShot, users are limited to a visual and a quantitative analysis of the results, i.e., shots. A qualitative analysis which includes to watch the video material is not provided. SportSense in turn provides a combination of all analysis options which are relevant for analysts and coaches to optimize the decision support.

SportSense needs to be highlighted again at this point as a very flexible and powerful tool. Among others, this is because approaches from research could easily be integrated into the architecture of the system. KPIs and metrics as, e.g., dangerousity or EPV could be modelled as continuous states (see Definition 3.2) and thus appropriately analyzed and visualized. However, additional KPIs, metrics, or functionalities should only be included if these are developed in close collaboration with coaches to avoid the semantic gap problem.

6.2 Industrial Solutions

We observe a remarkable professionalization in modern sports clubs. Amongst others, this can be seen in the kind and the amount of data which is gathered from athletes to maximize performances and to minimize the risk of injuries. This can be, for instance, information on an athlete's physiological status like health and fitness values, stress levels, or sleep data. Of course, also sport-specific event data, spatio-temporal tracking data, videos, as well as statistics and indicators are collected. With this increasing degree of professionalization, the number of analysts in clubs is increasing as well. Even whole analytical departments are established in many professional sports clubs and associations. The technical staff is expected to capture, process, analyze, and visualize data and provide relevant information fast for the coaching purposes [BTL+18]. To adequately handle all the data, various software tools are applied. Some of these tools which are used in sports practice are highlighted in this section.

We need to differentiate the tools with respect to the following four goals: (1) athlete management, (2) video content enhancement, (3) semi-automatic data generation, and (4) competition analysis. We introduce these different system types together with their strengths and weaknesses. Furthermore, we present some market leaders in the corresponding fields. Despite we assign industrial solutions to one of the four previously introduced categories, we want to state at this point that each system might contain elements from other categories in a more or less pronounced form as well. Nevertheless, we introduce the examples in the context of the category in which they have their key competences.

6.2.1 Athlete Management

Athlete management systems provide a complete suite of modules which are required to manage the performance and the development of athletes [SRP+20b]. Such modules focus different topics (e.g., health, training, performance analysis) and can include various information like, for instance, health, injury, and fitness documentations, but also training or match performance data.

Advantages. A big advantage of athlete management systems is the integration of many data from different sources into one platform. This allows a reduction of the number of different tools which otherwise would be needed by the clubs. Therefore, athlete management systems make the workflow in clubs very efficient. Furthermore, athlete management systems can reduce the financial expenditure for the clubs as less license fees for other tools need to be paid. The expenditure of time for club officials like coaches and managers can also be reduced because several tasks can be fulfilled efficiently with the same system. Another advantage of athlete management systems is that all the different departments and officials of sports clubs have access to the system, of course with some restrictions concerning data protection. This allows for a certain degree of transparency and therefore facilitates the communication in sports clubs. This can be advantageous when searching for new athletes during scouting processes or to justify certain decisions.

Disadvantages. One major drawback of athlete management systems is that these are not or at least not much used for the data generation process itself, but that most data need to be imported. This is where problems can occur. Many systems are limited because they are not compatible with some of the technology used by clubs. If, for instance, the tracking technology a club uses for the acquisition of spatio-temporal tracking data provides non-compatible data with the athlete management system, another tool for the analysis of these tracking data is required. If this is the case for other data sources as well, clubs have the problem of various systems being used in parallel. This means that each department has to perform own analyses and it will be extremely difficult to perform overarching analyses (see Section 2.2.2).

Examples. We identify SAP Sports One[1] as one of the preferred athlete management systems by sports clubs. This system is available for different game

[1] `https://www.sap.com/products/sports-one.html`

sports like, e.g., the invasion games basketball, football, handball, or ice hockey. It offers six different modules: (1) team management, (2) training management, (3) player fitness, (4) scouting insights, (5) performance insights, and (6) player health [SAP20]. The coaching and medical staff appreciate especially the option to supervise the players' internal loads of training sessions and the physiological reactions. This allows a good monitoring and enables the staff to proactively reduce the training load of athletes if they show any sign of physical overload. Therefore, SAP Sports One can play a key role for injury prevention. Furthermore, the system comes with cloud-based solutions allowing for data and information access from everywhere and with any mobile device. This improves the communication and collaboration within the whole team. To improve the preparation for upcoming matches, SAP Sports One also provides some options, for example, video imports of decisive match events [SAP20].

Another example of an athlete management system is Catapult AMS[2]. Because Catapult also sells products in the area of wearable technology (see Section 2.2.1), data from their own devices can easily be integrated into the Catapult AMS if clubs use the corresponding technology. Data can also be collected and accessed via mobile devices and information can be shared across departments and teams. The data analysis process is supported with visualizations in the form of customizable dashboards and reports which should help coaches to make decisions [Cat20]. However, the option for video analysis as a form of match preparation is not supported by Catapult AMS. To perform such analyses other products would be required like, for instance, Catapult Vision[3].

6.2.2 Video Content Enhancement

With previously introduced athlete management systems club officials can handle many topics in different areas of club management. Nevertheless, video analysis is often not covered adequately with these systems. As mentioned in Section 1.2 the video is essential for the coach's analysis process because the whole context of situations is maintained and important questions can be answered like, for example, "How is the player's orientation", "Where does the player look at during a specific event", or "With which foot does the player shoot at goal". These types of information are seldomly represented in match statistics, spatio-temporal tracking data, or event data. Thus, coaches and analysts spend many hours of watching videos to analyze the opponent's tactics

[2] https://www.catapultsports.com/products/catapult-ams
[3] https://www.catapultsports.com/products/vision

[DVD18]. To highlight the results of their analyses it is often necessary to enhance the pure video footage. In the following we focus on systems which are specialized on creating enhanced video content.

Advantages. With different options the video material can be processed and edited in a way that facilitates the analysis of certain scenes by highlighting important elements like, e.g., key players, open spaces, or running trajectories. This can help players and the coach to faster understand the playing style of the opponent and optimize the (tactical) training sessions for the upcoming match. Often, such systems are also used to generate event data through manual (live-) tagging [SRP+20b]. The tagging allows the connection of the video with own event data, which has the advantage of being independent of commercial event datasets. This further is advantageous because coaches' individual concepts can be integrated into the analysis. Additionally, high-level tactical patterns can then be analyzed by the coach in the corresponding video scenes. Nevertheless, most of the providers do support the import of commercial event datasets as well. This allows a very flexible way of working. With video content enhancement systems it is also possible to create video snippets which then can be shared with the players and other club officials [SRP+20b].

Disadvantages. Drawbacks of these kind of systems are that they are mostly limited to a pure video analysis functionality and therefore only support a qualitative analysis. Options to perform quantitative analyses like, e.g., creating aggregated statistics are often missing. Another drawback is the expenditure of time. The task of tagging events manually can be very time-intensive, especially with a lack of experience in this field. Game sports in general and invasion games in particular are very fast and many events can happen in a short time period. Therefore, it takes some time until users are able to perform appropriate (live-) tagging. However, this (live-) tagging can be essential especially if coaches' concepts (i.e., events and patterns of events) are not represented in commercial event datasets.

Examples. The first example of a video content enhancement system which is widely used in sports practice is Hudl Sportscode[4]. This software solution allows the manual live-tagging of training or match videos through completely customizable code windows. It does not matter which video material is used

[4] https://www.hudl.com/products/sportscode

and can be captured either with own cameras but also with camera solutions from Hudl which support multi-angle capturing on one computer [Spo20b]. Videos can also be imported and tagged afterwards. The generated event data are automatically synchronized with the video and individual clips for the analysis can be created. Additional drawing tools and graphics can be used to enhance the video content and to highlight specific elements in the respective scenes. The resulting clips can then be used, e.g., for presentations of the performance analysis staff with the coaches and athletes. These clips can also be shared with athletes or other people who need to watch the corresponding scenes of the video analysis via mobile devices [Spo20a].

Another example which we want to present in this section is myDartfish ProS[5]. This tool has lots of commonalities with Hudl Sportscode like a live-tagging feature as well as multi-camera capturing in a single recording [Dar20]. Additionally, data from third-party applications can be imported like, e.g., from InStat[6]. It further provides the options to draw, e.g., animated shapes, arrows, and trajectories, into the video scenes to highlight movements of players or the ball. A calibration of the different playing fields can be used to calculate positions, speed, accelerations, and distances of players which can be displayed in the video [Dar20]. The resulting video clips can provide support even during matches if they are shared, for instance, with coaches sitting on the bench.

6.2.3 Semi-Automatic Data Generation

The third category of industrial software solutions are systems which follow a semi-automatic data generation process. Such platforms are used most often by scouts, but of course also by analysts and coaches. The core of such systems is a large video database which contains tagged match video material of players and teams from around the world.

Advantages. These systems allow to search for certain actions of any athlete and thus facilitates the scouting process because travelling is no longer required to observe the development of talented players. Single video snippets of specific actions can also be downloaded and used later, for instance, to show strengths and weaknesses of opponent players to the own team. Therefore, these systems are suitable for a qualitative analysis. The time savings are a major advantage of such platforms. With only a few clicks one can find and select several video

[5] `https://www.dartfish.com/pro_s`
[6] `https://instatsport.com/`

clips and must not rely on the time-intensive task of watching whole videos and cutting scenes by oneself. Additionally, some systems provide reports with a large number of player and team statistics and thus can also support quantitative analyses. This in turn can help drawing conclusions on teams and player performances and thus can support pre-match as well as post-match analyses.

Disadvantages. Drawbacks of such platforms are that users have to rely on the already tagged actions or events. This can be problematic if scouts, coaches, and analysts have a different idea or definition of certain sport-specific events or patterns like, e.g., the definition of a long pass, a defensive transition phase, or pressure. If this is the case, then the usability of these systems is limited and coaches and analysts might need to perform their own tagging, e.g., with systems presented in Section 6.2.2. Some events or concepts of certain coaches might be missing completely in the database because the employees tagging the matches do not consider these concepts during their work. This is not the fault of the employees but might make the platform less attractive for coaches who rely on their own concepts. In these cases own tagging will then be the better and only choice, despite the fact that this is much more time-intensive. Otherwise, there will be no option for performing high-level tactical analyses. Another limitation are the statistical reports. Despite the big amount of analyses, data, and statistics, it is often not possible to conduct own analyses but to use only fixed reports. In sports practice it might be more useful to flexibly create own statistical analyses on-the-fly and to be independent of fixed statistics. Here too, if coaches have a different understanding of concepts, events or patterns, the resulting reports have only a limited value for these coaches.

Examples. The first example which we want to present here is the platform InStat Scout[7] which is available for football, ice hockey, and basketball. InStat Scout provides an interface to supply coaches and analysts with statistical reports on various team and player parameters, e.g., to compare performances [SRP+20b]. Additionally, videos of full matches or summaries, as well as individual player actions can be watched [Sco20]. InStat Scout provides the users also with interactive charts, e.g., shot charts which will help to prepare for upcoming matches. The information provided on the platform in either a video or a statistical format is updated mostly within only several hours after a match and sometimes even in real-time [Sco20]. An additional option which

[7] `https://instatsport.com/football/instat_scout`

makes the platform attractive is the connection of statistics with videos. Clicking on certain parameters will lead the users to a video playlist with corresponding scenes [Sco20].

Another system with comparable features to InStat Scout is Wyscout[8]. This football platform consists of a large video database which covers more than 250 different football competitions [Wys20]. Users can get professional reports with statistics, as well as further insights through charts, tables, heatmaps, and more. Player and team actions can be selected and the respective scenes can be watched and downloaded [Wys20].

6.2.4 Competition Analysis

Systems which are specialized on competition analysis are the fourth and last category of industrial software solutions which we introduce in this section. Such systems can be parts (modules) of athlete management systems and provide different tools and features for competition or match analysis. Nevertheless, software for competition analysis often also come as standalone-systems. Therefore, we decided to separate them into an own category. As the name already indicates these systems have their key competence in the context of competitions. The idea is to support coaches and analysts particularly during three of the seven different contexts we introduced in Section 2.2.3: (1) pre-match or competition analysis, (2) live-match or competition analysis, and (3) post-match or competition analysis.

Advantages. Competition analysis systems have several features like, for instance, a video player. Therefore, they often support qualitative analyses. Furthermore, they provide options to import data from different sources or to generate own data which allows for a quantitative analysis based on statistics and indicators. Often, also data visualizations like, e.g., bar charts, line graphs, or spider charts are supported and allow for a fast overview on athletes' performance data. The combination of different analysis options, if provided by the corresponding system, is one major advantage of these kind of systems. This combination allows the analysis of strengths and weaknesses of the opponents for the preparation of competitions, as well as the analysis of the performances of the own athletes or teams for the post-competition analysis. If data and video footage can be provided in real-time, such systems can also be used live and support decision making processes during competitions. All in all, working with

[8] `https://wyscout.com/`

competition analysis systems can be very efficient as they cover three different decision making contexts of coaches and analysts.

Disadvantages. The biggest problem of these systems is the lack of flexibility. It is often the case that systems from industrial providers offer fixed statistical features, visualizations, and maybe even own created KPIs. Additionally, the latter ones often are not transparent and thus have no or only a limited value for coaches and analysts. A generation of statistics on-the-fly is often not supported as well. Furthermore, individual concepts of coaches might not be represented in the analysis features which are provided by these systems. This in turn means that coaches are not able to perform high-level tactical performance analyses based on their own concepts (i.e., mental models) of the game.

Examples. One example which we want to present here is Match Tracker[9] from SBG Sports Software[10]. This football match analysis system is used by 70% of the clubs in English Premier League [Sof20] and is modularly structured. The software supports a qualitative analysis by importing one or several videos. Additionally, data from various sources can be imported and synchronized with the video like, e.g, event data, spatio-temporal tracking data, and even biometric data [Sof20]. Dashboards displaying statistics and other information provide the option for quantitative analyses and allow the comparison between player and team performances. Additional visualizations like pass trajectories or space control on a 2D representation of the pitch can help coaches and analysts to get deeper insights into certain match situations. Another interesting feature for the live match analysis are automatic alerts. These alerts are also customizable and allow the user to quickly pinpoint important situations in the match [Sof20].

6.2.5 Discussion

In the previous sections we presented a broad spectrum of industrial solutions which are currently in use in sports practice to support game sports analysts and coaches with their daily work and with their decision making processes. However, these systems also have their limitations in sports practice as we will discuss in the following.

The first reason why existing systems from industry often are of limited use in sports practice is their strong specialization like, for instance, in video content

[9] `https://sbgsportssoftware.com/product/matchtracker-for-teams/`
[10] `https://sbgsportssoftware.com/`

enhancement. Therefore, coaches and analysts often need to select specific tools which fit their own analysis preferences best. In our user studies we can observe the consequences, namely that many coaches and analysts use at least two tools for their analysis processes (see Figure 5.2(a)). Despite individual tools might be very powerful, the decision support nevertheless is not optimal. Additionally, using more tools in parallel is very time-consuming and thus often not possible due to the time pressure which analysts are exposed to.

The second very serious drawback are the costs. Many of the introduced industrial solutions come with license fees which clubs have to pay, e.g., every season. This can quickly exceed the financial capacities of clubs, especially clubs with a smaller budget. Therefore, these clubs need to use cheaper versions with less functionalities or less support and thus might be impelled to search for other solutions.

SportSense needs to be highlighted again at this point because the system has a big potential to cope with both introduced limitations. First, SportSense combines all analysis options which are relevant for game sports analysts and coaches. Additionally, the query types which are required for a holistic retrieval of game sports data are supported. Moreover, the generic architecture provides a good basis for the implementation of further functionalities. Second, the code of SportSense is freely available because the system is published open source. In summary, this further confirms the importance and the success of our approach how to conceptualize and design DASs to support decision making in game sports.

Part IV

Conclusions and Outlook

7

A man with new ideas is a madman, until his ideas triumph.

— Marcelo Bielsa

Conclusions

7.1 Summary

In this thesis, we have taken an attempt at providing data-driven support for the decision making processes of game sports analysts and coaches. We have seen in Chapters 1 and 2 that decision making in the context of sports, especially in game sports, is very complex. Additionally, we showed that coaches and analysts are facing lots of challenges and are exposed to an enormous pressure in their daily work. Therefore, we introduced an approach how to conceptualize and design DASs for an efficient support of the decision making processes in sports practice.

In this work, we presented an approach how to bridge the semantic gap of sports analytics. This can be achieved through an extraction of domain experts' (e.g., coaches) concepts and a modeling of these concepts into a performance model and later into a data model. This procedure is particularly important for the development of DASs because these systems only then will provide relevant information for the users, in our case for game sports analysts and coaches. The representation of their own concepts in the data furthermore allows coaches and analysts to conduct high-level tactical analyses which is of enormous value in their daily work. Furthermore, we introduced a model to describe the complexity levels of different sports and showed that a DAS needs to combine three analysis options to optimize the decision support of coaches and analysts, namely qualitative analyses, quantitative analyses, and visualizations. We also presented four different query types which need to be supported by a DAS for a holistic information retrieval of game sports data: raw data queries, movement queries, event queries, and event pattern queries. These query types are required to allow for the previously introduced qualitative and quantitative analyses, as well as visualizations based on the retrieved data. Additionally, we showed an approach how to proceed when planning and designing the UX of a DAS for game sports. For that, five steps need to be considered which all are based upon each other and assure a user-centered design. The latter point, the user focus, is essential to make the interaction with the system as intuitive and as flexible as possible for coaches and analysts which in turn allows a time-efficient working and further supports the decision making processes.

With SportSense we presented a generic implementation of a DAS for game sports. This system provides a combination of the three mentioned analysis options. Moreover, it supports the four different query types. This means that coaches and analysts are able to work very flexible and efficient with SportSense. We also presented two specific applications of the system, namely SportSense Football and SportSense Ice Hockey. We implemented some football-specific concepts into SportSense Football which we extracted from interviews with coaches. Though, we successfully applied our approach to bridge the semantic gap of sports analytics. The two applications further highlighted the enormous potential of the generic architecture which allows for quick and easy adaptations to the characteristics of the sport in question.

As SportSense Football is the more sophisticated application this system was evaluated qualitatively and quantitatively. The results showed that SportSense Football is a useful tool in "real-world" scenarios of football practice despite some optimizations are required.

7.2 Future Work

Over the course of this work, several ideas for further research emerged. In this section, we now discuss these ideas which could be part of upcoming projects based on the results of this thesis.

7.2.1 Optimization of SportSense

The first and most obvious point for future work is to take the present evaluation results and optimize the SportSense Football system, for instance, by implementing the suggestions from the participating domain experts (see Section 5.1.3.6). In the following, we present some points which have, from our point of view, the biggest improvement potential for SportSense.

General improvements include, for instance, the implementation of the playing direction. Game sports in general are played in several periods, where the teams change the sides at the beginning of each new period. Searching for certain movements in a sketch-based manner in the playing field component of SportSense then can be misleading if no period filters are set. Therefore, the idea is to fix the playing direction (e.g., attacking from left to right) and to support a mirrored search in the corresponding database collections.

Another point which should be considered in future work is a customization of the system according to the user preferences. This can be achieved, for instance, through a modular structure. Users then would be able to individually arrange the different components of SportSense.

Furthermore, the support of an export functionality would be valuable for users. Both qualitative analyses (e.g., video snippets or even video collections) and quantitative analyses (e.g., tables and reports) could then be exported and shared with athletes or other members of the coaching staff. This would add additional flexibility to the daily work of analysts and coaches.

Besides these major changes, lots of smaller points should also be considered to further improve SportSense. Among other things, this would mean to link the results of a raw data query (e.g., visualized as line graphs in the timeline) to the video, to support a dynamic switch between different video sources (e.g., tactical camera, broadcasting view, etc.), or to facilitate the qualitative video analysis part of coaches and analysts by supporting a more fine grained fast-forward and rewind functionality of the video player.

Adaptations which concern only the SportSense Football system can also be implemented based on the evaluation results. One very important point is to bring in more context into the analysis. This can mean, for instance, to link transitions to the subsequent action (i.e., to an event). This will then allow, for instance, to search for offensive transitions where the first pass after the ball possession gain was played vertically. Another example concerns the pass network. Here, more context could be added, for instance, by considering only passes during a specific match phase (see Section 5.1.3.4). Some smaller adaptations were also suggested by the participants of the user studies. Just to name a few, this would be to add the jersey numbers of the players to the playing field component or to introduce different zones (see Section 5.1.3.6). Additionally, there are a lot of football-specific concepts which we extracted from the interviews (see Table 3.2) which until now are not implemented in SportSense Football. A successful implementation of these concepts will make SportSense Football even more powerful and valuable for analysts and coaches in football practice.

Finally, the retrieval performance of raw data queries and the scalability of raw movement queries should be optimized as the results of our performance evaluation showed (see Sections 5.2.3.2 and 5.2.4.2).

7.2.2 Applications in Different Sports

In this thesis we showed how to conceptualize and design DASs to support decision making in game sports. With SportSense Football and SportSense Ice Hockey we presented two applications in the field of invasion games, which in turn represent the most complex subcategory of game sports (see Section 2.1.1). The two applications showed that our approach works very well in this complex context. We selected this category of sports types intentionally, as we assume that our approach will then also work for less complex game sports. So far, we however did not try to apply our approach to other sports with different characteristics compared to game sports. Therefore, it will be interesting to see how SportSense can be adapted, for instance, to duel sports like boxing (see Table 2.1). Nevertheless, we are convinced, that our generic architecture allows a quick and easy adaptation of SportSense to other, non-game sports.

The works of Maymin [May18] or Novak et al. [NBP⁺20] show clearly that performance analysis has already reached the field of eSports. Therefore, another interesting point for future work is an application of SportSense in this field. Some eSports games often have similarities with "real life" sports (e.g., FIFA 21 and football). Therefore, some of the already existing analysis features of SportSense could also be applied to eSports. However, other eSports games might reveal some interesting characteristics which "traditional" sports do not have. One example are team fights as they appear, for instance, in DotA 2 or other MOBA[1] games. Here, several events like, for instance, fights between players can happen in parallel. Tactics therefore can differ fundamentally compared to other sports. Consequently, this requires new analysis features which could be developed in future projects. In the work of Zumsteg [Zum19] a first attempt to apply SportSense in the field of eSports was undertaken. However, the focus of this work was more on the event detection part compared to the information retrieval and (tactical) performance analysis. As the global eSports market size was valued at 1.1 billion U.S. dollars in 2019 and is expected to expand at a compound annual growth rate of 24.4% from 2020 to 2027 [Res20] the economic potential lying in this field is enormous. Additionally, eSports is attracting large audiences. Today, hundreds of millions of people spectate eSports [HS17]. Thus,

[1] Multiplayer Online Battle Arena (MOBA)

the analysis of individual player performances and team tactics would be very valuable for sports practice. Furthermore, an analytical advantage of eSports over traditional sports is the number of games played [May18]. However, this just enforces the need for a DAS to accelerate and support the analysis process. Consequently, we propose a project which focuses on the application of SportSense in the field of eSports (e.g., e-SportSense).

7.2.3 Generalisation and Parameterization

As we have seen in Sections 4.2.1 and 4.3.1 there is not much effort needed when adapting the generic architecture of SportSense to the specific versions of football or ice hockey, respectively. In contrast, the implementation of sport-specific concepts and features is more elaborate. This is mainly because of the semantic gap of sports analytics. Here, it takes considerable effort to extract the knowledge of domain experts and to perform the subsequent modeling steps. Also the planning and designing of the UX might require more time due to the user-centered approach.

To facilitate the generation of future SportSense applications, the idea would be to further generalize the system and to allow quick and easy adaptations through the definition of certain parameters. A library which consists of, for instance, events, plays, and analysis features should be provided because some of these points might be interesting for more than just one sport. A specific SportSense application can then be generated and customized based on this library. Options to parameterize individual events, plays, or functionalities thus should be provided to further optimize the process.

In a very abstract way this could look like the following. To create a SportSense application for basketball, we would have to define the general characteristics of the game (e.g., four quarters), define and adapt basketball-specific events (e.g., passes, shots, goals), and specify certain analysis features (e.g., support of event queries). The definition of four quarters would then lead to an appropriate representation of the timeline component and corresponding filter options. Basketball-specific events could be adapted, e.g., from the game sports-generic definitions which would already be part of the library. The same holds for the analysis features, which could be selected from the library.

A deployment of SportSense in the Cloud would lead to an additional flexibility. Here, first attempts on a previous version of SportSense are shown successfully in the work of Al Kabary and Schuldt [AS20]. Future work in this area will allow an even more flexible and fast generation of SportSense applications.

7.2.4 Enhancement of Information Retrieval

SportSense was designed in a way to allow users an intuitive retrieval of relevant information for their analyses. In this section, we propose two ideas to even further improve the information retrieval process.

The first idea for future work in this area is an extension of the existing sketch-based retrieval features. When we think about a coach who draws various plays on the tactics board in team meetings, e.g., dashed lines for running trajectories, waved lines for dribblings, and solid lines for passes or shots, this could be starting points for further implementations into SportSense. By introducing new "vocabluary" via new sketches, this would further reduce the semantic gap of sports analytics because the system then provides a "language" which is even closer to the coach's one.

The second idea is to apply machine learning approaches to SportSense. The more a user interacts with the system the better it should provide support for the user's tasks. To name an example, implementations could go into the direction of automatic query completion or query recommendations. Another example can be the predefining of filters which users often apply for a certain query type.

This topic certainly provides many other points with a great potential for future work.

7.2.5 Support of Simulations

So far, SportSense Football and SportSense Ice Hockey are very valuable DASs for coaches and analyst. This holds particularly for the usage during the pre-match analysis and post-match analysis contexts. Users can analyze, e.g., *what* and *when* something happened in a match. However, a major benefit for coaches would be to know what would happen if a certain decision is made, i.e., a support of simulations. With so called "What-If" analyses a user would be able to check the consequences of certain decisions, e.g., of a formation change or a player's substitution. If coaches would be able, e.g., in their match preparation phase to simulate how the opponent team reacts to a certain change of formation, this would be a substantial benefit when planning the own tactics and generating the match plan. This would be even more interesting if the DAS could be used in real-time for live-match analyses.

Therefore, the support of simulations is an additional point which should be considered for future projects.

7.2.6 Applications in Other Contexts

We motivated in Chapter 1 that decision making is an elementary process in all areas of life. As sports is just one use case, another point for future projects is a transfer of our results beyond the field of sports. The approach we have presented in this thesis for data-driven analytics for decision making is generic and we are convinced that it can also be applied to other areas like, for instance, in business and organizational settings. A further use case can consider disaster management which is also proposed for future activities with the StreamTeam system in the work of Probst [Pro20]. Besides a live-analysis of a specific situation, e.g., the acting of fire fighters inside a burning house, it is also interesting and important to post-analyze the team behaviour for further operations. Therefore an application of SportSense in such a scenario would be a good tool for training sessions.

Appendix

A

Interview Guideline

General

1. Concerning the match preparation and the match plan: how large is the percentage of video analysis compared to the analysis based on statistical data?

Player Profiles

2. What are possible strengths of a team? Name three.
3. What are possible weaknesses of a team? Name three.
4. What are possible strengths of a player? Name three.
5. What are possible weaknesses of a player? Name three.
6. General player selection: which factors play an important role?
7. How do physical capabilities of a player influence the match plan?

Opponent Information

8. How is an opponent's key player characterized?

Offensive Organisation

9. Where and how can the offense be started?
10. What further variations exist during the attack?

Defensive Organisation

11. How can the defense be organized?
12. How (i.e., variations) and where (i.e., zone) is the opponent attacked?

Transition DEF-OFF

13. What options exist during transition DEF-OFF?

Transition OFF-DEF

14. What options exist during transition OFF-DEF?

Build-up Play

15. What options exist during build-up play?
16. What reactions to the opponent's build-up play exist?
17. How do you explain this to your players? (Sketch)

Set Plays

18. What variations exist for corner kicks?
19. How should the defense act in a corner kick situation?
20. How do you explain this to your players? (Sketch)
21. What variations exist for free kicks?
22. How should the defense act in a free kick situation?
23. How do you explain this to your players? (Sketch)

Principles

24. How often do you change the system during a match?
25. What are possible reasons for changing the system?

Depending on the time: Do you change the tactics (system, or another element of the match plan) ...

26. ... after a certain amount of time?
27. ... depending on the current score?

Diverse

28. Are there further points which you consider for the match plan?
29. Are off-ball movements integrated in the match plan as well?

B

GitHub Repositories

The code of all generic and application-specific implementations is published in the following repositories of the SportSense GitHub project (see https://github.com/sportsense) under the GNU Affero General Public License v3.0. All descriptions given in Chapter 4 as well as the evaluation presented in Chapter 5 refer to the code which is tagged with version 1.0.0.

– **sportsense-evaluation:** code/scripts for the performance evaluation (see Section 5.2)

– **sportsense-mongodb-rest-proxy:** code of the MongoDB REST proxy (see Section 4.1.2)

– **sportsense-web-client:** code of the football/ice hockey-specific UIs (see Sections 4.2 and 4.3)

C

Evaluation Questionnaire

Questions marked with an * are mandatory.

Part 1: General Questions

1. What is your age?*
   Text
2. What is your gender?*
   - ◯ Male
   - ◯ Female
   - ◯ Diverse
3. What is/was your role?*
   - ◯ Coach
   - ◯ Analyst
   - ◯ Other
4. If "Other" was selected in Question #3, which one?
   Text
5. On which level?*
   - ◯ Club
   - ◯ National Team
   - ◯ Other
6. If "Other" was selected in Question #5, which one?
   Text
7. Which league and/or age category?
   Text

8. Which coaching license (e.g., UEFA Pro) do you have?
   Text

Part 2: Combining Different Analysis Options

9. Do you use any software tools for your analyses in order to prepare for an upcoming match?*
   - ○ Yes
   - ○ No
10. Which options do you use for your analyses?*
    - □ Watching videos (Qualitative Video Analysis)
    - □ Reading reports (Quantitative Data Analysis)
    - □ Visualize heatmaps, movements, etc. (Visualizations)
    - □ Other options
11. If "Other options" was selected in Question #10, which one(s)?
    Text
12. How many different software tools do you use for your analyses?*
    - ○ 0
    - ○ 1
    - ○ 2
    - ○ > 2
13. Which software tools do you use for your analyses?
    Text
14. Would you profit from a software tool which combines different analysis options?*
    - ○ Yes
    - ○ No
15. Concerning Question #14, why or why not?
    Text
16. How many past matches do you consider for the match preparation?*
    Text

Part 3: Supporting Different Query Types

17. Which data are relevant for your analyses?*

- ☐ Raw data (e.g., GPS, heart rate)
- ☐ Event data (e.g., passes, shots)
- ☐ Statistics (e.g., average number of long passes per match)
- ☐ Indicators (e.g., expected goals, packing)
- ☐ Video

18. Demo - Event Query - How do you rate the usability of this query type?*
    - ○ very bad
    - ○ bad
    - ○ neutral
    - ○ good
    - ○ very good
19. Would you use this feature for your analyses?*
    - ○ Yes
    - ○ No
20. Event Query - Optional comment
    Text
21. Demo - Event Pattern Query - How do you rate the usability of this query type?*
    - ○ very bad
    - ○ bad
    - ○ neutral
    - ○ good
    - ○ very good
22. Would you use this feature for your analyses?*
    - ○ Yes
    - ○ No
23. Event Pattern Query - Optional comment
    Text
24. Demo - Raw Data Query - How do you rate the usability of this query type?*
    - ○ very bad
    - ○ bad
    - ○ neutral

○ good

○ very good

25. Would you use this feature for your analyses?*

    ○ Yes

    ○ No

26. Raw Data Query - Optional comment
    Text

27. Demo - Raw Movement Query - How do you rate the usability of this query type?*

    ○ very bad

    ○ bad

    ○ neutral

    ○ good

    ○ very good

28. Would you use this feature for your analyses?*

    ○ Yes

    ○ No

29. Raw Movement Query - Optional comment
    Text

30. Demo - Quantitative Analysis - How do you rate the usability of this analysis type?*

    ○ very bad

    ○ bad

    ○ neutral

    ○ good

    ○ very good

31. Would you use this feature for your analyses?*

    ○ Yes

    ○ No

32. Quantitative Analysis - Optional comment
    Text

33. Demo - Qualitative Analysis - How do you rate the usability of this analysis type?*

- ◯ very bad
- ◯ bad
- ◯ neutral
- ◯ good
- ◯ very good

34. Would you use this feature for your analyses?*
    - ◯ Yes
    - ◯ No
35. Qualitative Analysis - Optional comment
    Text

Part 4: Bridging the Semantic Gap of Sports Analytics

36. Demo - Pressing Index/Pressing Phases - How do you rate the representation of the concept “Pressure”?*
    - ◯ very bad
    - ◯ bad
    - ◯ neutral
    - ◯ good
    - ◯ very good
37. Would you use this feature for your analyses?*
    - ◯ Yes
    - ◯ No
38. Pressing Index/Pressing Phases - Optional comment
    Text
39. Demo - Transition Phases - How do you rate the representation of the concepts “Lose Possession” and “Gain Possession”?*
    - ◯ very bad
    - ◯ bad
    - ◯ neutral
    - ◯ good
    - ◯ very good
40. Would you use this feature for your analyses?*
    - ◯ Yes

- ◯ No

41. Transition Phases - Optional comment
    Text
42. Demo - Pass Network - How do you rate the representation of the concept "Cooperation"?*
    - ◯ very bad
    - ◯ bad
    - ◯ neutral
    - ◯ good
    - ◯ very good
43. Would you use this feature for your analyses?*
    - ◯ Yes
    - ◯ No
44. Pass Network - Optional comment
    Text

Part 5: Planning and Designing a DAS for Decision Support in Game Sports

45. How do you rate the representation of the timeline component?*
    - ◯ very bad
    - ◯ bad
    - ◯ neutral
    - ◯ good
    - ◯ very good
46. Would you use this component for your analyses?*
    - ◯ Yes
    - ◯ No
47. Timeline - Optional comment
    Text
48. How do you rate the representation of the playing field component?*
    - ◯ very bad
    - ◯ bad
    - ◯ neutral
    - ◯ good

- ◯ very good

49. Would you use this component for your analyses?*
    - ◯ Yes
    - ◯ No
50. Playing field - Optional comment
    Text
51. How do you rate the representation of the video player component?*
    - ◯ very bad
    - ◯ bad
    - ◯ neutral
    - ◯ good
    - ◯ very good
52. Would you use this component for your analyses?*
    - ◯ Yes
    - ◯ No
53. Video Player - Optional comment
    Text
54. How do you rate the clarity of the arrangement of the three main components?*
    - ◯ very bad
    - ◯ bad
    - ◯ neutral
    - ◯ good
    - ◯ very good
55. Arrangement of the three main components - Optional comment
    Text
56. How do you rate the clarity of the arrangement of all components within the user interface?*
    - ◯ very bad
    - ◯ bad
    - ◯ neutral
    - ◯ good
    - ◯ very good

57. Arrangement of all components - Optional comment
    Text

Part 6: Final Questions

58. How was your overall impression about the SportSense system?*
    - ○ very bad
    - ○ bad
    - ○ neutral
    - ○ good
    - ○ very good
59. Which additional features would you like to have supported by SportSense?
    Text
60. Optional final comments?
    Text

D

Additional Evaluation Data

Table D.1 Software tools used by the participants of the user studies. The table shows how many coaches and analysts use certain software tools for their analysis process w.r.t. the match preparation.

Software Tool	# Coaches	# Analysts	# Total
Hudl Sportscode	2	4	6
Wyscout	1	4	5
Coach Paint	1	3	4
InStat Scout	0	3	3
Tableau	1	2	3
Catapult AMS	2	0	2
SBG MatchTracker	0	2	2
Viz Libero	0	2	2
Adobe Premiere Pro	1	0	1
Catapult OpenField	1	0	1
Final Cut Pro	1	0	1
Internal application	0	1	1
iThlete	1	0	1
LongoMatch	1	0	1
MAGIX	1	0	1
Match Analysis Hub (DFL)	0	1	1
Metrica Play	0	1	1
Microsoft Excel	0	1	1
Polar Team Pro	1	0	1
SAP Sports One	0	1	1
Second Spectrum	0	1	1
SVEXA	1	0	1
Trumedia	0	1	1

Table D.2 MongoDB REST proxy calls issued during a raw data-, raw movement-, and event query. For the performance evaluation queries of different types are executed: a raw data query q_r, a raw movement query q_{rmov}, and an event query q_e. Each query is listed with its corresponding HTTP GET request (i.e., the MongoDB REST proxy call).

Query	HTTP GET Request
q_r	http://10.34.58.65:2222/analyzePressing2d?matchFilters={%22match0%22:%22209527%22}
q_{rmov}	http://10.34.58.65:2222/getMotionPath?shape=polygon&coordinates={%22vertices%22:[[-2.08,-24.48],[-3.45,-24.22],[-9.71,-24.22],[-10.45,-24.08],[-11.35,-24.08],[-11.41,-24.06],[-12.78,-23.80],[-13.50,-23.80],[-14.24,-23.66],[-17.15,-23.66],[-17.88,-23.52],[-22.69,-23.52],[-23.43,-23.38],[-23.89,-23.38],[-23.95,-23.36],[-24.68,-23.22],[-25.65,-23.04],[-26.14,-22.80],[-26.72,-22.66],[-27.45,-22.39],[-28.18,-22.11],[-29.34,-21.88],[-29.49,-21.83],[-30.08,-21.55],[-30.95,-20.99],[-31.06,-20.94],[-31.11,-20.89],[-32.27,-20.15],[-33.00,-19.87],[-33.07,-19.83],[-33.15,-19.77],[-33.88,-19.21],[-34.90,-18.23],[-35.63,-17.54],[-36.36,-16.56],[-36.82,-15.92],[-37.04,-15.57],[-37.33,-14.87],[-37.77,-13.89],[-37.91,-13.43],[-37.93,-13.25],[-38.08,-12.41],[-38.08,-11.43],[-38.35,-10.12],[-38.52,-9.54],[-38.52,-9.05],[-38.66,-8.21],[-38.66,-7.37],[-38.93,-6.06],[-38.95,-6.02],[-38.95,-5.98],[-39.10,-5.27],[-39.10,-5.14],[-39.37,-3.83],[-39.47,-3.64],[-39.66,-2.71],[-40.44,-1.60],[-41.60,-0.85], [-42.97,-0.59],[-44.34,-0.85],[-45.50,-1.60],[-46.27,-2.71],[-46.55,-4.02],[-46.55,-4.44],[-46.27,-5.75],[-46.13,-6.17], [-46.13,-6.17],[-45.98,-6.87],[-45.91,-7.08],[-45.82,-7.52],[-45.82,-8.21],[-45.67,-9.05],[-45.67,-9.89],[-45.40,-11.20], [-45.23,-11.78],[-45.23,-12.41],[-45.09,-13.25],[-44.94,-14.23],[-44.67,-15.54],[-44.38,-16.51],[-43.94,-17.49],[-43.65,-18.19], [-43.21,-18.89],[-42.43,-20.00],[-41.41,-21.40],[-40.68,-22.38],[-39.96,-23.08],[-38.93,-24.06],[-38.21,-24.62],[-37.48,-25.18], [-36.32,-25.92],[-35.73,-26.20],[-35.55,-26.27],[-34.57,-26.90],[-33.69,-27.32],[-32.82,-27.88],[-32.23,-28.15],[-31.50,-28.43], [-30.34,-28.66],[-30.19,-28.71],[-29.46,-28.99],[-28.88,-29.13],[-28.29,-29.41],[-26.93,-29.67],[-26.57,-29.73],[-25.82,-29.97],[-24.45,-30.23],[-23.58,-30.23],[-22.84,-30.37],[-18.03,-30.37],[-17.30,-30.51],[-14.68,-30.51],[-13.95,-30.65],[-13.45,-30.65],[-13.42,-30.67],[-12.05,-30.93],[-11.03,-30.93],[-10.30,-31.07],[-3.45,-31.07],[-2.08,-30.81],[-0.92,-30.07],[-0.14,-28.96],[0.13,-27.65],[-0.14,-26.34],[-0.92,-25.22]]}&eventFilters={}&teamFilters={}&playerFilters={filter0:BALL}&periodFilters={}&timeFilter={}&sportFilter={%27sport%27:football}&matchFilters={}
q_e	http://10.34.58.65:2222/getAreaEvents?shape=rectangle&coordinates={%22bottomLeftX%22:%2233.38%22,%22bottomLeftY%22:%2216.46%22,%22upperRightX%22:%2249.42%22,%22upperRightY%22:%22-16.67%22}&eventFilters={}&teamFilters={}&playerFilters={}&periodFilters={}&timeFilter={}&sportFilter={%27sport%27:football}&matchFilters={}

Table D.3 MongoDB REST proxy call issued during a quantitative analysis.
For the performance evaluation a quantitative analysis is executed. The query is listed with its corresponding HTTP GET request (i.e., the MongoDB REST proxy call).

HTTP GET Request
http://10.34.58.65:2222/analyzePlayers?user=Select%20User&discipline&players=A4,A5,A8,A9&parameters=gamesPlayed, gamesWon,gamesLost,gamesDrawn,winPercentage,successfulPassEvent,misplacedPassEvent,passAccuracy,longPasses, shortPasses,avgPassLength,avgPassVelocity,avgPacking,leftPasses,rightPasses,forwardPasses,backwardPasses,goalEvent, shotOnTargetEvent,shotOffTargetEvent,totalShots,avgShotLength,avgShotVelocity,successfulTakeOnEvent,failedTakeOnEvent, DribblingStatistic,interceptionEvent,playerFoulsEvent,playerGetFouledEvent,clearanceEvent,timeSpeedZone1,timeSpeedZone 2,timeSpeedZone3,timeSpeedZone4,timeSpeedZone5,timeSpeedZone6,cornerkickEvent,throwinEvent,freekickEvent,total Touches,playerOn,playerOff&matches=

Table D.4 MongoDB REST proxy call issued during the first step of a forward event cascade. For the performance evaluation an event pattern query (i.e., forward event cascade) is executed. The first step of the query is listed with its corresponding HTTP GET request (i.e., the MongoDB REST proxy call).

HTTP GET Request
http://10.34.58.65:2222/getAreaEvents?shape=rectangle&coordinates={%22bottomLeftX%22:%2238.05%22,%22bottomLeftY%22:%2212.83%22,%22upperRightX%22:%2241.84%22,%22upperRightY%22:%229.05%22}&eventFilters={}&teamFilters={}&playerFilters={}&periodFilters={}&timeFilter={}&sportFilter={%27sport%27:football}&matchFilters={}

Table D.5 MongoDB REST proxy call issued during the second step of a forward event cascade. For the performance evaluation an event pattern query (i.e., forward event cascade) is executed. The second step of the query is listed with its corresponding HTTP GET request (i.e., the MongoDB REST proxy call).

HTTP GET Request
http://10.34.58.65:2222/getEventCascade?reverse=false&timestamps=[{%22id%22:%22607b18dbaee15b0915190a9e%22,%22t%2 2:%22316960%22,%22matchId%22:%22209527%22},{%22id%22:%22607b18dbaee15b0915190c82%22,%22t%22:%22322080%22,%2 2matchId%22:%22209527%22},{%22id%22:%22607b18dbaee15b0915190dc6%22,%22t%22:%22324400%22,%22matchId%22:%2220 9527%22},{%22id%22:%22607b18dbaee15b0915190ef6%22,%22t%22:%22328200%22,%22matchId%22:%22209527%22},{%22id%22 :%22607b18dcaee15b0915191018%22,%22t%22:%22331000%22,%22matchId%22:%22209527%22},{%22id%22:%22607b18dcaee15b 0915191170%22,%22t%22:%22334200%22,%22matchId%22:%22209527%22},{%22id%22:%22607b18dcaee15b09151912ae%22,%22t %22:%22336480%22,%22matchId%22:%22209527%22},{%22id%22:%22607b18dcaee15b09151913a2%22,%22t%22:%22339280%22, %22matchId%22:%22209527%22},{%22id%22:%22607b18dcaee15b09151916f2%22,%22t%22:%22345480%22,%22matchId%22:%22 209527%22},{%22id%22:%22607b18ddaee15b0915191abc%22,%22t%22:%22349520%22,%22matchId%22:%22209527%22},{%22id %22:%22607b18ddaee15b0915191b2e%22,%22t%22:%22350800%22,%22matchId%22:%22209527%22},{%22id%22:%22607b18dfae e15b0915192d8c%22,%22t%22:%22376600%22,%22matchId%22:%22209527%22},{%22id%22:%22607b18dbaee15b0915190a9f%22, %22t%22:%223016960%22,%22matchId%22:%22209527%22},{%22id%22:%22607b18dbaee15b0915190c83%22,%22t%22:%2230220 80%22,%22matchId%22:%22209527%22},{%22id%22:%22607b18dbaee15b0915190dc7%22,%22t%22:%223024400%22,%22matchId %22:%22209527%22},{%22id%22:%22607b18dbaee15b0915190ef7%22,%22t%22:%223028200%22,%22matchId%22:%22209527%22 },{%22id%22:%22607b18dcaee15b0915191019%22,%22t%22:%223031000%22,%22matchId%22:%22209527%22},{%22id%22:%2260 7b18dcaee15b0915191171%22,%22t%22:%223034200%22,%22matchId%22:%22209527%22},{%22id%22:%22607b18dcaee15b09151 912af%22,%22t%22:%223036480%22,%22matchId%22:%22209527%22},{%22id%22:%22607b18dcaee15b09151913a3%22,%22t%22: %223039280%22,%22matchId%22:%22209527%22},{%22id%22:%22607b18dcaee15b09151916f3%22,%22t%22:%223045480%22,%2 2matchId%22:%22209527%22},{%22id%22:%22607b18ddaee15b0915191abd%22,%22t%22:%223049520%22,%22matchId%22:%22 209527%22},{%22id%22:%22607b18ddaee15b0915191b2f%22,%22t%22:%223050800%22,%22matchId%22:%22209527%22},{%22id %22:%22607b18dfaee15b0915192d8d%22,%22t%22:%223076600%22,%22matchId%22:%22209527%22}]&threshold=5000&shape= circle&coordinates={%22centerX%22:%22-12.42%22,%22centerY%22:%2222.75%22,%22radius%22:%227.750000000000002%22}& eventFilters={}&teamFilters={}&playerFilters={}&periodFilters={}&timeFilter={}&sportFilter={%27sport%27:football}&match Filters={}

Table D.6 MongoDB REST proxy call issued during the third step of a forward event cascade.
For the performance evaluation an event pattern query (i.e., forward event cascade) is executed. The third step of the query is listed with its corresponding HTTP GET request (i.e., the MongoDB REST proxy call).

HTTP GET Request
http://10.34.58.65:2222/getEventCascade?reverse=false&timestamps=[{%22id%22:%22607b18ddaee15b0915191b2c%22,%22t%22:%22350800%22,%22matchId%22:%22209527%22},{%22id%22:%22607b18ddaee15b0915191b2d%22,%22t%22:%223050800%22,%22matchId%22:%22209527%22}]&threshold=5000&shape=rectangle&coordinates={%22bottomLeftX%22:%22-29.48%22,%22bottomLeftY%22:%2211.57%22,%22upperRightX%22:%22-16.21%22,%22upperRightY%22:%22-9.68%22}&eventFilters={}&teamFilters={}&playerFilters={}&periodFilters={}&timeFilter={}&sportFilter={%27sport%27:football}&matchFilters={}

Table D.7 MongoDB REST proxy call issued during the first step of a reverse event cascade.
For the performance evaluation an event pattern query (i.e., reverse event cascade) is executed. The first step of the query is listed with its corresponding HTTP GET request (i.e., the MongoDB REST proxy call).

HTTP GET Request
http://10.34.58.65:2222/getAreaEvents?shape=rectangle&coordinates={%22bottomLeftX%22:%22-50.19%22,%22bottomLeftY%22:%22-4.37%22,%22upperRightX%22:%22-34.00%22,%22upperRightY%22:%22-16.81%22}&eventFilters={}&teamFilters={}&playerFilters={}&periodFilters={}&timeFilter={}&sportFilter={%27sport%27:football}&matchFilters={}

Table D.8 MongoDB REST proxy call issued during the second step of a reverse event cascade.
For the performance evaluation an event pattern query (i.e., reverse event cascade) is executed. The second step of the query is listed with its corresponding HTTP GET request (i.e., the MongoDB REST proxy call).

HTTP GET Request
http://10.34.58.65:2222/getEventCascade?reverse=true&timestamps=[{%22id%22:%22607b18c7aee15b0915184fd6%22,%22time%22:%2218640%22,%22matchId%22:%22209527%22},{%22id%22:%22607b18d6aee15b091518e1bc%22,%22time%22:%22241840%22,%22matchId%22:%22209527%22},{%22id%22:%22607b18f7aee15b09151a1184%22,%22time%22:%22789840%22,%22matchId%22:%22209527%22},{%22id%22:%22607b190baee15b09151acab2%22,%22time%22:%221117960%22,%22matchId%22:%22209527%22},{%22id%22:%22607b192caee15b09151c0400%22,%22time%22:%221693720%22,%22matchId%22:%22209527%22},{%22id%22:%22607b192daee15b09151c0edc%22,%22time%22:%221712760%22,%22matchId%22:%22209527%22},{%22id%22:%22607b192eaee15b09151c10c8%22,%22time%22:%221715120%22,%22matchId%22:%22209527%22},{%22id%22:%22607b1940aee15b09151cb256%22,%22time%22:%221977080%22,%22matchId%22:%22209527%22},{%22id%22:%22607b194caee15b09151d2154%22,%22time%22:%222171000%22,%22matchId%22:%22209527%22},{%22id%22:%22607b195faee15b09151dd45e%22,%22time%22:%222511440%22,%22matchId%22:%22209527%22},{%22id%22:%22607b195faee15b09151dd568%22,%22time%22:%222512960%22,%22matchId%22:%22209527%22},{%22id%22:%22607b18c7aee15b0915184fd7%22,%22time%22:%222718640%22,%22matchId%22:%22209527%22},{%22id%22:%22607b18d6aee15b091518e1bd%22,%22time%22:%222941840%22,%22matchId%22:%22209527%22},{%22id%22:%22607b18f7aee15b09151a1185%22,%22time%22:%223489840%22,%22matchId%22:%22209527%22},{%22id%22:%22607b190baee15b09151acab3%22,%22time%22:%223817960%22,%22matchId%22:%22209527%22},{%22id%22:%22607b192caee15b09151c0401%22,%22time%22:%224393720%22,%22matchId%22:%22209527%22},{%22id%22:%22607b192daee15b09151c0edd%22,%22time%22:%224412760%22,%22matchId%22:%22209527%22},{%22id%22:%22607b192eaee15b09151c10c9%22,%22time%22:%224415120%22,%22matchId%22:%22209527%22},{%22id%22:%22607b1940aee15b09151cb257%22,%22time%22:%224677080%22,%22matchId%22:%22209527%22},{%22id%22:%22607b194caee15b09151d2155%22,%22time%22:%224871000%22,%22matchId%22:%22209527%22},{%22id%22:%22607b195faee15b09151dd45f%22,%22time%22:%225211440%22,%22matchId%22:%22209527%22},{%22id%22:%22607b195faee15b09151dd569%22,%22time%22:%225212960%22,%22matchId%22:%22209527%22}]&threshold=5000&shape=none&coordinates={}&eventFilters={}&teamFilters={}&playerFilters={}&periodFilters={}&sportFilter={%27sport%27:football}&matchFilters={}&timeFilter={}

Table D.9 MongoDB REST proxy call issued during the third step of a reverse event cascade.
For the performance evaluation an event pattern query (i.e., reverse event cascade) is executed. The third step of the query is listed with its corresponding HTTP GET request (i.e., the MongoDB REST proxy call).

HTTP GET Request
http://10.34.58.65:2222/getEventCascade?reverse=true&timestamps=[{%22id%22:%22607b18d6aee15b091518e0b4%22,%22time%22:%22240520%22,%22matchId%22:%22209527%22},{%22id%22:%22607b18f7aee15b09151a114e%22,%22time%22:%22789480%22,%22matchId%22:%22209527%22},{%22id%22:%22607b18d6aee15b091518e0b5%22,%22time%22:%222940520%22,%22matchId%22:%22209527%22},{%22id%22:%22607b18f7aee15b09151a114f%22,%22time%22:%223489480%22,%22matchId%22:%22209527%22}]&threshold=5000&shape=none&coordinates={}&eventFilters={}&teamFilters={}&playerFilters={}&periodFilters={}&sportFilter={%27sport%27:football}&matchFilters={}&timeFilter={}

Bibliography

[AS13a] Ihab Al Kabary and Heiko Schuldt. SportSense: Using Motion Queries to Find Scenes in Sports Videos. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management (CIKM '13)*, pages 2489–2492, San Fransisco, CA, USA. ACM, 2013. DOI: 10.1145/2505515.2508211.

[AS13b] Ihab Al Kabary and Heiko Schuldt. Towards Sketch-Based Motion Queries in Sports Videos. In *Proceedings of the 2013 IEEE International Symposium on Multimedia (ISM '13)*, pages 309–314, Anaheim, CA, USA. IEEE, 2013. DOI: 10.1109/ISM.2013.60.

[AS14] Ihab Al Kabary and Heiko Schuldt. Enhancing Sketch-Based Sport Video Retrieval by Suggesting Relevant Motion Paths. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '14)*, pages 1227–1230, Gold Coast, Queensland, Australia. ACM, 2014. DOI: 10.1145/2600428.2609551.

[AS20] Ihab Al Kabary and Heiko Schuldt. Scalable Sketch-Based Sport Video Retrieval in the Cloud. In *Proceedings of the 13th International Conference on Cloud Computing (CLOUD '20)*, pages 226–241, Honolulu, HI, USA. Springer International Publishing, 2020. DOI: 10.1007/978-3-030-59635-4_16.

[Alm86] Len Almond. Reflecting on Themes: A Games Classification. In Rod Thorpe, David Bunker, and Len Almond, editors, *Rethinking Games Teaching*, pages 71–72. Loughborough University of Technology, Loughborough, UK, 1986.

[AAB+17] Gennady Andrienko, Natalia Andrienko, Guido Budziak, Jason Dykes, Georg Fuchs, Tatiana von Landesberger, and Hendrik Weber. Visual Analysis of Pressure in Football. *Data Mining and Knowledge Discovery*, 31(6):1793–1839, 2017. DOI: 10.1007/s10618-017-0513-2.

[AYS07] Vasilis Armatas, Athanasios Yiannakos, and Paraskevi Sileloglou. Relationship Between Time and Goal Scoring in Soccer Games: Analysis of Three World Cups. *International Journal of Performance*

Analysis in Sport, 7(2):48–58, 2007. DOI: 10.1080/24748668.2007.11868396.

[AWS21] Amazon Web Services AWS. What Is a Document Database? https://aws.amazon.com/nosql/document/, 2021. (Last accessed: 09-February-2021).

[BKL05] Michael Beetz, Bernhard Kirchlechner, and Martin Lames. Computerized Real-Time Analysis of Football Games. *IEEE Pervasive Computing*, 4(3):33–39, 2005. DOI: 10.1109/MPRV.2005.53.

[BTL+18] Alejandro Benito Santos, Roberto Theron, Antonio Losada, Jaime E Sampaio, and Carlos Lago-Peñas. Data-Driven Visual Performance Analysis in Soccer: An Exploratory Prototype. *Frontiers in Psychology*, 9:2416, 2018. DOI: 10.3389/fpsyg.2018.02416.

[BLC+14] Alina Bialkowski, Patrick Lucey, Peter Carr, Yisong Yue, Sridha Sridharan, and Iain Matthews. Identifying Team Style in Soccer Using Formations Learned from Spatiotemporal Tracking Data. In *Proceedings of the 2014 IEEE International Conference on Data Mining Workshop (ICDM Workshops '14)*, pages 9–14, Shenzhen, China. IEEE, 2014. DOI: 10.1109/ICDMW.2014.167.

[BL20] Thomas Blobel and Martin Lames. A Concept for Club Information Systems (CIS) – An Example for Applied Sports Informatics. *International Journal of Computer Science in Sport*, 19(1):102–122, 2020. DOI: 10.2478/ijcss-2020-0006.

[Boj14] Irena Bojanova. IT Enhances Football at World Cup 2014. *IT Professional*, 16(4):12–17, 2014. DOI: 10.1109/MITP.2014.54.

[BSM10] Jérôme Bourbousson, Carole Seve, and Tim McGarry. Space-Time Coordination Dynamics in Basketball: Part 2. The Interaction Between the Two Teams. *Journal of Sports Sciences*, 28(3):349–358, 2010. DOI: 10.1080/02640410903503640.

[BLM19] Ulf Brefeld, Jan Lasek, and Sebastian Mair. Probabilistic Movement Models and Zones of Control. *Machine Learning*, 108:127–147, 2019. DOI: 10.1007/s10994-018-5725-1.

[BKT+18] Michel S Brink, Jurian P Kuyvenhoven, Tynke Toering, Geir Jordet, and Wouter GP Frencken. What Do Football Coaches Want From Sport Science? *Kinesiology*, 50(1):150–154, 2018.

[Buc17] Martin Buchheit. Houston, We Still Have a Problem. *International Journal of Sports Physiology and Performance*, 12(8):1111–1114, 2017. DOI: 10.1123/ijspp.2017-0422.

[CD15] Jon M Calder and Ian N Durbach. Decision Support for Evaluating Player Performance in Rugby Union. *International Journal of Sports Science & Coaching*, 10(1):21–37, 2015. DOI: 10.1260/1747-9541.10.1.21.

[CBF+18] Valentina Camomilla, Elena Bergamini, Silvia Fantozzi, and Giuseppe Vannozzi. Trends Supporting the In-Field Use of Wearable Inertial Sensors for Sport Performance Evaluation: A Systematic Review. *Sensors*, 18(3):873, 2018. DOI: 10.3390/s18030873.

[CTW00] Fiona Carmichael, Dennis Thomas, and Robert Ward. Team Performance: The Case of English Premiership Football. *Managerial and Decision Economics*, 21(1):31–45, 2000. DOI: 10.1002/1099-1468(200001/02)21:1<31::AID-MDE963>3.0.CO;2-Q.

[Cat20] Catapult. Athlete Management. https://www.catapultsports.com/products/catapult-ams, 2020. (Last accessed: 24-November-2020).

[CDB+14] Dan Cervone, Alexander D'Amour, Luke Bornn, and Kirk Goldsberry. Pointwise: Predicting Points and Valuing Decisions in Real Time with NBA Optical Tracking Data. In *Proceedings of the 8th MIT Sloan Sports Analytics Conference*, Boston, MA, USA, 2014.

[Cha13] Philippe Chassy. Team Play in Football: How Science Supports F.C. Barcelona's Training Strategy. *Psychology*, 4(9):7–12, 2013. DOI: 10.4236/psych.2013.49A2002.

[CGP+15] Paolo Cintia, Fosca Giannotti, Luca Pappalardo, Dino Pedreschi, and Marco Malvaldi. The Harsh Rule of the Goals: Data-Driven Performance Indicators for Football Teams. In *Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA '15)*, pages 1–10, Paris, France. IEEE, 2015. DOI: 10.1109/DSAA.2015.7344823.

[CCM+15] Filipe Manuel Clemente, Micael Santos Couceiro, Fernando Manuel Lourenço Martins, and Rui Sousa Mendes. Using Network Metrics in Soccer: A Macro-Analysis. *Journal of Human Kinetics*, 45(1):123–134, 2015. DOI: 10.1515/hukin-2015-0013.

[CMK+15] Filipe Manuel Clemente, Fernando Manuel Lourenço Martins, Dimitris Kalamaras, P Del Wong, and Rui Sousa Mendes. General Network Analysis of National Soccer Teams in FIFA World Cup 2014. *International Journal of Performance Analysis in Sport*, 15(1):80–96, 2015. DOI: 10.1080/24748668.2015.11868778.

[Col13] Christian Collet. The Possession Game? A Comparative Analysis of Ball Retention and Team Success in European and International Football, 2007–2010. *Journal of Sports Sciences*, 31(2):123–136, 2013. DOI: 10.1080/02640414.2012.727455.

[Com20a] The International Olympic Committee. Beijing 2022. https://www.olympic.org/beijing-2022, 2020. (Last accessed: 05-November-2020).

[Com20b] The International Olympic Committee. Tokyo 2020. https://www.olympic.org/tokyo-2020, 2020. (Last accessed: 05-November-2020).

[Cou14] Aaron J Coutts. Evolution of Football Match Analysis Research. *Journal of Sports Sciences*, 32(20):1829–1830, 2014. DOI: 10.1080/02640414.2014.985450.

[Dar20] Dartfish. High-Performance Sports Video and Data Analysis Solution. https://www.dartfish.com/pro_s, 2020. (Last accessed: 25-November-2020).

[DVD18] Tom Decroos, Jan Van Haaren, and Jesse Davis. Automatic Discovery of Tactics in Spatio-Temporal Soccer Match Data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18)*, pages 223–232, London, United Kingdom. ACM, 2018. DOI: 10.1145/3219819.3219832.

[DM12] Juan Luis Delgado-Bordonau and Alberto Mendez-Villanueva. Tactical Periodization: Mourinho's Best-Kept Secret. *Soccer Journal*, 57(3):29–34, 2012.

[DUI+19] Emrullah Delibas, Ali Uzun, Mehmet Fatih Inan, Onur Guzey, and Ali Cakmak. Interactive Exploratory Soccer Data Analytics. *INFOR: Information Systems and Operational Research*, 57(2):141–164, 2019. DOI: 10.1080/03155986.2018.1533204.

[DBT+07] Valter Di Salvo, Ramon Baron, Harald Tschan, Francisco Javier Calderon Montero, Norbert Bachl, and Fabio Pigozzi. Performance Characteristics According to Playing Position in Elite Soccer. *International Journal of Sports Medicine*, 28(3):222–227, 2007. DOI: 10.1055/s-2006-924294.

[Doc21] MDN Web Docs. Ajax – Developer Guides. https://developer.mozilla.org/en-US/docs/Web/Guide/AJAX, 2021. (Last accessed: 25-January-2021).

[DS15] Rina A Doherty and Paul Sorenson. Keeping Users in the Flow: Mapping System Responsiveness with User Experience. *Procedia Manufacturing*, 3:4384–4391, 2015. DOI: 10.1016/j.promfg.2015.07.436.

[DG13] Barry Drust and Matthew Green. Science and Football: Evaluating the Influence of Science on Performance. *Journal of Sports Sciences*, 31(13):1377–1382, 2013. DOI: 10.1080/02640414.2013.828544.

[DWA10] Jordi Duch, Joshua S Waitzman, and Luís A Nunes Amaral. Quantifying the Performance of Individual Players in a Team Activity. *PLoS ONE*, 5(6):e10937, 2010. DOI: 10.1371/journal.pone.0010937.

[EN94] Martin Engelhardt and Georg Neumann. *Sportmedizin. Grundlagen für alle Sportarten*. BLV Buchverlag, München, Germany, 1994. ISBN: 9783405143251.

[FB18] Javier Fernández and Luke Bornn. Wide Open Spaces: A Statistical Technique for Measuring Space Creation in Professional Soccer. In *Proceedings of the 12th MIT Sloan Sports Analytics Conference*, Boston, MA, USA, 2018.

[FBC19] Javier Fernández, Luke Bornn, and Dan Cervone. Decomposing the Immeasurable Sport: A Deep Learning Expected Possession Value Framework for Soccer. In *Proceedings of the 13th MIT Sloan Sports Analytics Conference*, Boston, MA, USA, 2019.

[Fou21] Eclipse Foundation. The Eclipse Jetty Project. https://www.eclipse.org/jetty/, 2021. (Last accessed: 25-January-2021).

[FLD+11] Wouter Frencken, Koen Lemmink, Nico Delleman, and Chris Visscher. Oscillations of Centroid Position and Surface Area of Soccer Teams in Small-Sided Games. *European Journal of Sport Science*, 11(4):215–223, 2011. DOI: 10.1080/17461391.2010.499967.

[Gam20] Tokyo 2020 Olympic Games. Sports. https://tokyo2020.org/en/sports/, 2020. (Last accessed: 05-November-2020).

[Gar05] Jesse James Garrett. Ajax: A New Approach to Web Applications. https://web.archive.org/web/20080702075113/http://www.adaptivepath.com/ideas/essays/archives/000385.php, 2005. (Last accessed: 25-January-2021).

[Gar12] Jesse James Garrett. *Die Elemente der User Experience: Anwenderzentriertes (Web-) Design*. Addison Wesley Imprint in der Pearson Education Deutschland, München, Germany, 2012. ISBN: 9783827331168.

[Ghi15] Rakesh Ghildiyal. Role of Sports in the Development of an Individual and Role of Psychology in Sports. *Mens Sana Monographs*, 13(1):165–170, 2015. DOI: 10.4103/0973-1229.153335.

[Goa20] Goal. How Much Money do the Champions League 2019-20 Winners Get? https://www.goal.com/en-us/news/how-much-money-champions-league-2019-20-winners-get/1b78sfes1al5k1cxve9ld1wp48, 2020. (Last accessed: 27-October-2020).

[GKM+19] Floris R Goes, Matthias Kempe, Laurentius A Meerhoff, and Koen APM Lemmink. Not Every Pass Can Be an Assist: A Data-Driven Model to Measure Pass Effectiveness in Professional Soccer Matches. *Big Data*, 7(1):57–70, 2019. DOI: 10.1089/big.2018.0067.

[GMB+21] Floris R Goes, Laurentius A Meerhoff, Murilo JO Bueno, Daniele M Rodrigues, Felipe A Moura, Michael S Brink, Marije T Elferink-Gemser, Arno J Knobbe, Sergio A Cunha, Ricardo S Torres, and Koen APM Lemmink. Unlocking the Potential of Big Data to Support Tactical Performance Analysis in Professional Soccer: A Systematic Review. *European Journal of Sport Science*, 21(4):481–496, 2021. DOI: 10.1080/17461391.2020.1747552.

[GC04] Ryan Groom and Chris Cushion. Coaches Perceptions of the Use of Video Analysis: A Case Study. *Insight*, 7(3):56–58, 2004.

[GH17] Joachim Gudmundsson and Michael Horton. Spatio-Temporal Analysis of Team Sports. *ACM Computing Surveys*, 50(2):22:1–34, 2017. DOI: 10.1145/3054132.

[GRD19] Rohmat Gunawan, Alam Rahmatulloh, and Irfan Darmawan. Performance Evaluation of Query Response Time in The Document Stored NoSQL Database. In *Proceedings of the 16th International Conference on Quality in Research (QIR '19): International Symposium on Electrical and Computer Engineering*, pages 1–6, Padang, Indonesia. IEEE, 2019. DOI: 10.1109/QIR.2019.8898035.

[HS17] Juho Hamari and Max Sjöblom. What Is eSports and Why Do People Watch It? *Internet Research*, 27(2):211–232, 2017. DOI: 10.1108/IntR-04-2016-0085.

[Hei10] Andreas M Hein. Identification and Bridging of Semantic Gaps in the Context of Multi-Domain Engineering. In *Proceedings of the 2010 Forum on Philosophy, Engineering & Technology (fPET '10)*, pages 57–58, Golden, CO, USA, 2010.

[HKB+21] Mat Herold, Matthias Kempe, Pascal Bauer, and Tim Meyer. Attacking Key Performance Indicators in Soccer: Current Practice and Perceptions From the Elite to Youth Academy Level. *Journal of Sports Science & Medicine*, 20:158–169, 2021. DOI: 10.52082/jssm.2021.158.

[HPS+18] Jennifer Hobbs, Paul Power, Long Sha, and Patrick Lucey. Quantifying the Value of Transitions in Soccer via Spatiotemporal Trajectory Clustering. In *Proceedings of the 12th MIT Sloan Sports Analytics Conference*, Boston, MA, USA, 2018.

[KDJ+01] Y Antero Kesaniemi, Elliot Danforth, Michael D Jensen, Peter G Kopelman, Pierre Lefèbvre, and Bruce A Reeder. Dose-Response Issues Concerning Physical Activity and Health: An Evidence-Based Symposium. *Medicine and Science in Sports and Exercise*, 33(6):S351–S358, 2001. DOI: 10.1097/00005768-200106001-00003.

[KS10] Pat Kirwan and David Seigerman. *Take Your Eye Off the Ball: How to Watch Football by Knowing Where to Look*. Triumph Books, Chicago, IL, USA, 2010. ISBN: 9781600783913.

[Kla08] Andreas Klages. Politikfeld Sport. Die Gesellschaftspolitische Bedeutung des Gemeinwohlorientierten Sports. In Thomas von Winter and Volker Mittendorf, editors, *Perspektiven der Politischen Soziologie im Wandel von Gesellschaft und Staatlichkeit*, pages 185–202. Springer VS, Wiesbaden, Germany, 2008. DOI: 10.1007/978-3-531-90908-0_13.

[Kuc18] Udo Kuckartz. *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*. Beltz Verlagsgruppe, Weinheim, Germany, 2018. ISBN: 9783779936824.

[LM07] Martin Lames and Tim McGarry. On the Search for Reliable Performance Indicators in Game Sports. *International Journal of Performance Analysis in Sport*, 7(1):62–79, 2007. DOI: `10.1080/24748668.2007.11868388`.

[LCY+17] Hoang M Le, Peter Carr, Yisong Yue, and Patrick Lucey. Data-Driven Ghosting Using Deep Imitation Learning. In *Proceedings of the 11th MIT Sloan Sports Analytics Conference*, Boston, MA, USA, 2017.

[LK18] Eun-Jeong Lee and Hye-Young Kim. Optimal Regression Model for Predicting the Winning Game and Contributing Factors in Ice Hockey World Championship. *ISBS Proceedings Archive*, 36(1):183–186, 2018.

[LRM20] Erik Lignell, Vincenzo Rago, and Magni Mohr. Analysis of Goal Scoring Opportunities in Elite Male Ice Hockey in Relation to Tactical and Contextual Variables. *International Journal of Performance Analysis in Sport*, 20(6):1003–1017, 2020. DOI: `10.1080/24748668.2020.1823161`.

[Lin18] Daniel Link. *Data Analytics in Professional Soccer*. Springer Vieweg, Wiesbaden, Germany, 2018. ISBN: 9783658211776.

[LH17] Daniel Link and Martin Hoernig. Individual Ball Possession in Soccer. *PLoS ONE*, 12(7):e0179953, 2017. DOI: `10.1371/journal.pone.0179953`.

[LLS16] Daniel Link, Steffen Lang, and Philipp Seidenschwarz. Real Time Quantification of Dangerousity in Football Using Spatiotemporal Tracking Data. *PLoS ONE*, 11(12):e0168768, 2016. DOI: `10.1371/journal.pone.0168768`.

[LMF20] Sergio Llana, Pau Madrero, and Javier Fernández. The Right Place at the Right Time: Advanced Off-Ball Metrics for Exploiting an Opponent's Spatial Weaknesses in Soccer. In *Proceedings of the 14th MIT Sloan Sports Analytics Conference*, Boston, MA, USA, 2020.

[MAP20] MAPFRE. Data Analysis in Formula 1: The Difference Between Victory and Defeat. https://www.mapfre.com/en/insights/innovation/data-analysis-in-formula-1-the-difference-between-victory-and-defeat/, 2020. (Last accessed: 26-May-2021).

[Mar19] Guillermo Martinez Arastey. History of Performance Analysis: The Controversial Pioneer Charles Reep. https://www.sportperformanceanalysis.com/article/history-of-performance-analysis-the-controversial-pioneer-charles-reep, 2019. (Last accessed: 27-October-2020).

[MF07] Daniel S Mason and William M Foster. Putting Moneyball on Ice? *International Journal of Sport Finance*, 2(4):206–213, 2007.

[May18] Philip Z Maymin. An Open-Sourced Optical Tracking and Advanced eSports Analytics Platform for League of Legends. In *Proceedings of the 12th MIT Sloan Sports Analytics Conference*, Boston, MA, USA, 2018.

[MR18] Ian G McHale and Samuel D Relton. Identifying Key Players in Soccer Teams Using Network Analysis and Pass Difficulty. *European Journal of Operational Research*, 268(1):339–347, 2018. DOI: 10.1016/j.ejor.2018.01.018.

[MZT+18] Nazanin Mehrasa, Yatao Zhong, Frederick Tung, Luke Bornn, and Greg Mori. Deep Learning of Player Trajectory Representations for Team Activity Analysis. In *Proceedings of the 12th MIT Sloan Sports Analytics Conference*, Boston, MA, USA, 2018.

[MOG13] Stephen A Mitchell, Judith L Oslin, and Linda L Griffin. *Teaching Sport Concepts and Skills – A Tactical Games Approach for Ages 7 to 18*. Human Kinetics, Champaign, IL, USA, 2013. ISBN: 9781450411226.

[Mon21a] MongoDB. The Most Popular Database for Modern Apps. https://www.mongodb.com/, 2021. (Last accessed: 06-January-2021).

[Mon21b] MongoDB. What is MongoDB? https://www.mongodb.com/what-is-mongodb, 2021. (Last accessed: 06-January-2021).

[NJB+19] Scott B Nicholls, Nic James, Elizabeth Bryant, and Julia Wells. The Implementation of Performance Analysis and Feedback Within Olympic Sport: The Performance Analyst's Perspective. *International Journal of Sports Science & Coaching*, 14(1):63–71, 2019. DOI: 10.1177/1747954118808081.

[NBP+20] Andrew R Novak, Kyle JM Bennett, Matthew A Pluss, and Job Fransen. Performance Analysis in eSports: Modelling Performance at the 2018 League of Legends World Championship. *International Journal of Sports Science & Coaching*, 15(5-6):809–817, 2020. DOI: 10.1177/1747954120932853.

[Opt20] Opta. Data Feeds. https://www.optasports.com/services/data-feeds/, 2020. (Last accessed: 19-November-2020).

[PVF13] Charles Perin, Romain Vuillemot, and Jean-Daniel Fekete. SoccerStories: A Kick-Off for Visual Soccer Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2506–2515, 2013. DOI: 10.1109/TVCG.2013.192.

[PSB+12] Hannah Pileggi, Charles D Stolper, J Michael Boyle, and John T Stasko. SnapShot: Visualization to Propel Ice Hockey Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2819–2828, 2012. DOI: 10.1109/tvcg.2012.263.

[PRB17] Raffaele Poli, Loïc Ravenel, and Roger Besson. How to Evaluate Player Performance. https://football-observatory.com/IMG/pdf/note02en.pdf, 2017. (Last accessed: 05-May-2021).

[Pol02] Richard Pollard. Charles Reep (1904-2002): Pioneer of Notational and Performance Analysis in Football. *Journal of Sports Sciences*, 20(10):853–855, 2002. DOI: 10.1080/026404102320675684.

[Pro20] Lukas Probst. *Spatio-Temporal Multi Data Stream Analysis with Applications in Team Sports*. PhD thesis, University of Basel, 2020. DOI: 10.5451/unibas-007211799.

[PAL+18] Lukas Probst, Ihab Al Kabary, Rufus Lobo, Fabian Rauschenbach, Heiko Schuldt, Philipp Seidenschwarz, and Martin Rumo. SportSense: User Interface for Sketch-Based Spatio-Temporal Team Sports Video Scene Retrieval. In *Proceedings of the 1st Workshop on User Interfaces for Spatial and Temporal Data Analysis (UISTDA '18)*, Tokyo, Japan. CEUR-WS, 2018. http://ceur-ws.org/Vol-2068/uistda3.pdf.

[PBS+17] Lukas Probst, Frederik Brix, Heiko Schuldt, and Martin Rumo. Real-Time Football Analysis with StreamTeam: Demo. In *Proceedings of the 11th ACM International Conference on Distributed and Event-based Systems (DEBS '17)*, pages 319–322, Barcelona, Spain. ACM, 2017. DOI: 10.1145/3093742.3095089.

[PRS+18] Lukas Probst, Fabian Rauschenbach, Heiko Schuldt, Philipp Seidenschwarz, and Martin Rumo. Integrated Real-Time Data Stream Analysis and Sketch-Based Video Retrieval in Team Sports. In *Proceedings of the 2018 IEEE International Conference on Big Data (BigData '18)*, pages 548–555, Seattle, WA, USA. IEEE, 2018. DOI: 10.1109/BigData.2018.8622592.

[PSS+20] Lukas Probst, Heiko Schuldt, Philipp Seidenschwarz, and Martin Rumo. StreamTeam-Football: Analyzing Football Matches in Real-Time on the Basis of Position Streams. In *Proceedings of the 2020 IEEE International Conference on Big Data (BigData '20)*, pages 637–644, Atlanta, GA, USA. IEEE, 2020. DOI: 10.1109/BigData50022.2020.9377914.

[RLA18] João Ramos, Rui J Lopes, and Duarte Araújo. What's Next in Complex Networks? Capturing the Concept of Attacking Play in Invasive Team Sports. *Sports Medicine*, 48:17–28, 2018. DOI: 10.1007/s40279-017-0786-z.

[RM16] Robert Rein and Daniel Memmert. Big Data and Tactical Analysis in Elite Soccer: Future Challenges and Opportunities for Sports Science. *SpringerPlus*, 5:1410, 2016. DOI: 10.1186/s40064-016-3108-2.

[Res20] Grand View Research. Esports Market Size, Share, Growth Trends Report, 2020-2027. https://www.grandviewresearch.com/industry-analysis/esports-market, 2020. (Last accessed: 21-April-2021).

[Res21] Grand View Research. Sports Analytics Market Size Report, 2021-2028. https://www.grandviewresearch.com/industry-analysis/sports-analytics-market, 2021. (Last accessed: 26-May-2021).

[Ric18] Keven Richly. Leveraging Spatio-Temporal Soccer Data to Define a Graphical Query Language for Game Recordings. In *Proceedings of the 2018 IEEE International Conference on Big Data (BigData '18)*, pages 3456–3463, Seattle, WA, USA. IEEE, 2018. DOI: 10.1109/BigData.2018.8622159.

[Rob20] Sam Robertson. Man & Machine: Adaptive Tools for the Contemporary Performance Analyst. *Journal of Sports Sciences*, 38(18):2118–2126, 2020. DOI: 10.1080/02640414.2020.1774143.

[RPW+17] Hector Ruiz, Paul Power, Xinyu Wei, and Patrick Lucey. "The Leicester City Fairytale?": Utilizing New Soccer Analytics Tools to Compare Performance in the 15/16 & 16/17 EPL Seasons. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17)*, pages 1991–2000, Halifax, Nova Scotia, Canada. ACM, 2017. DOI: 10.1145/3097983.3098121.

[SAB20] SABR. A Guide to Sabermetric Research. https://sabr.org/sabermetrics, 2020. (Last accessed: 27-October-2020).

[SHT+19] Fahim Salim, Fasih Haider, Sena Busra Yengec Tasdemir, Vahid Naghashi, Izem Tengiz, Kubra Cengiz, Dees Postma, Robby van Delden, Dennis Reidsma, Saturnino Luz, and Bert-Jan van Beijnum. A Searching and Automatic Video Tagging Tool for Events of Interest during Volleyball Training Sessions. In *Proceedings of the 2019 International Conference on Multimodal Interaction (ICMI '19)*, pages 501–503, Suzhou, China. ACM, 2019. DOI: 10.1145/3340555.3358660.

[San18] Øyvind Sandbakk. Let's Close the Gap Between Research and Practice to Discover New Land Together! *International Journal of Sports Physiology and Performance*, 13(8):961, 2018. DOI: 10.1123/ijspp.2018-0550.

[SAP20] SAP. SAP Sports One. https://www.sap.com/products/sports-one.html, 2020. (Last accessed: 24-November-2020).

[SKG+17] Oliver Schulte, Mahmoud Khademi, Sajjad Gholami, Zeyu Zhao, Mehrsan Javan, and Philippe Desaulniers. A Markov Game Model for Valuing Actions, Locations, and Team Performance in Ice Hockey. *Data Mining and Knowledge Discovery*, 31(6):1735–1757, 2017. DOI: 10.1007/s10618-017-0496-z.

[Sco20] InStat Scout. The Most Accurate Data and Biggest Player Database. https://instatsport.com/football/instat_scout, 2020. (Last accessed: 26-November-2020).

[SJP+20] Philipp Seidenschwarz, Adalsteinn Jonsson, Michael Plüss, Martin Rumo, Lukas Probst, and Heiko Schuldt. The SportSense User Interface for Holistic Tactical Performance Analysis in Football. In *Proceedings of the 25th International Conference on Intelligent User Interfaces Companion (IUI '20 Companion)*, pages 45–46, Cagliari, Italy. ACM, 2020. DOI: 10.1145/3379336.3381473.

[SJR+19] Philipp Seidenschwarz, Adalsteinn Jonsson, Fabian Rauschenbach, Martin Rumo, Lukas Probst, and Heiko Schuldt. Combining Qualitative and Quantitative Analysis in Football with SportSense. In *Proceedings of the 2nd International Workshop on Multimedia Content Analysis in Sports (MMSports '19)*, pages 34–41, Nice, France. ACM, 2019. DOI: 10.1145/3347318.3355519.

[SRP+20a] Philipp Seidenschwarz, Martin Rumo, Lukas Probst, and Heiko Schuldt. A Flexible Approach to Football Analytics: Assessment, Modeling and Implementation. In Martin Lames, Alexander Danilov, Egor Timme, and Yuri Vassilevski, editors, *Proceedings of the 12th International Symposium on Computer Science in Sport (IACSS 2019)*, pages 19–27, Moscow, Russia. Springer International Publishing, 2020. DOI: 10.1007/978-3-030-35048-2_3.

[SRP+20b] Philipp Seidenschwarz, Martin Rumo, Lukas Probst, and Heiko Schuldt. High-Level Tactical Performance Analysis with SportSense. In *Proceedings of the 3rd International Workshop on Multimedia Content Analysis in Sports (MMSports '20)*, pages 45–52, Seattle, WA, USA. ACM, 2020. DOI: 10.1145/3422844.3423053.

[SLY+16] Long Sha, Patrick Lucey, Yisong Yue, Peter Carr, Charlie Rohlf, and Iain Matthews. Chalkboarding: A New Spatiotemporal Query Paradigm for Sports Play Retrieval. In *Proceedings of the 21st International Conference on Intelligent User Interfaces (IUI '16)*, pages 336–347, Sonoma, CA, USA. ACM, 2016. DOI: 10.1145/2856767.2856772.

[SLY+18] Long Sha, Patrick Lucey, Yisong Yue, Xinyu Wei, Jennifer Hobbs, Charlie Rohlf, and Sridha Sridharan. Interactive Sports Analytics: An Intelligent Interface for Utilizing Trajectories for Interactive Sports Play Retrieval and Analytics. *ACM Transactions on Computer-Human Interaction*, 25(2):13, 2018. DOI: 10.1145/3185596.

[SG19] Laurie Shaw and Mark Glickman. Dynamic Analysis of Team Strategy in Professional Football. *Barça Sports Analytics Summit*, 2019. https://www.sportperformanceanalysis.com/s/Dynamic-analysis-of-team-strategy-in-professional-football-By-Laurie-Shaw-And-Mark-Glickman.pdf.

[Sof20] SBG Sport Software. MatchTracker for Teams. https://sbgsportsoftware.com/product/matchtracker-for-teams/, 2020. (Last accessed: 27-November-2020).

[Spe18] William Spearman. Beyond Expected Goals. In *Proceedings of the 12th MIT Sloan Sports Analytics Conference*, Boston, MA, USA, 2018.

[Spo20a] Hudl Sportscode. Fully Customizable Performance Analysis. https://www.hudl.com/products/sportscode, 2020. (Last accessed: 25-November-2020).

[Spo20b] Hudl Sportscode. Hudl Sportscode is Here. https://www.hudl.com/products/hudlsportscode, 2020. (Last accessed: 25-November-2020).

[Spo17] SportsPodium. How many sports are there in the world. https://steemit.com/sport/@sportspodium/how-many-sports-are-there-in-the-world, 2017. (Last accessed: 05-November-2020).

[SJL+17] Manuel Stein, Halldor Janetzko, Andreas Lamprecht, Thorsten Breitkreutz, Philipp Zimmermann, Bastian Goldlücke, Tobias Schreck, Gennady Andrienko, Michael Grossniklaus, and Daniel A Keim. Bring It to the Pitch: Combining Video and Movement Data to Enhance Team Sport Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):13–22, 2017. DOI: `10.1109/TVCG.2017.2745181`.

[SJS+19] Manuel Stein, Halldor Janetzko, Tobias Schreck, and Daniel A Keim. Tackling Similarity Search for Soccer Match Analysis: Multimodal Distance Measure and Interactive Query Definition. *IEEE Computer Graphics and Applications*, 39(5):60–71, 2019. DOI: `10.1109/MCG.2019.2922224`.

[SSM+19] Manuel Stein, Daniel Seebacher, Rui Marcelino, Tobias Schreck, Michael Grossniklaus, Daniel A Keim, and Halldor Janetzko. Where to Go: Computational and Visual What-If Analyses in Soccer. *Journal of Sports Sciences*, 37(24):2774–2782, 2019. DOI: `10.1080/02640414.2019.1652541`.

[TOF+20] Matt Taberner, Jason O'Keefe, David Flower, Jack Phillips, Graeme Close, Daniel Dylan Cohen, Chris Richter, and Christopher Carling. Interchangeability of Position Tracking Technologies; Can We Merge the Data? *Science and Medicine in Football*, 4(1):76–81, 2020. DOI: `10.1080/24733938.2019.1634279`.

[vdKR18] Eline van der Kruk and Marco M Reijne. Accuracy of Human Motion Capture Systems for Sport Applications; State-of-the-Art Review. *European Journal of Sport Science*, 18(6):806–819, 2018. DOI: 10.1080/17461391.2018.1463397.

[WLM+13] Xinyu Wei, Patrick Lucey, Stuart Morgan, and Sridha Sridharan. Sweet-Spot: Using Spatiotemporal Data to Discover and Predict Shots in Tennis. In *Proceedings of the 7th MIT Sloan Sports Analytics Conference*, Boston, MA, USA, 2013.

[WXW+18] Yingcai Wu, Xiao Xie, Jiachen Wang, Dazhen Deng, Hongye Liang, Hui Zhang, Shoubin Cheng, and Wei Chen. Forvizor: Visualizing Spatio-Temporal Team Formations in Soccer. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):65–75, 2018. DOI: 10.1109/TVCG.2018.2865041.

[Wys20] Wyscout. The Platform. https://wyscout.com/football-platform/, 2020. (Last accessed: 26-November-2020).

[XWL+21] Xiao Xie, Jiachen Wang, Hongye Liang, Dazhen Deng, Shoubin Cheng, Hui Zhang, Wei Chen, and Yingcai Wu. PassVizor: Toward Better Understanding of the Dynamics of Soccer Passes. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1322–1331, 2021. DOI: 10.1109/TVCG.2020.3030359.

[Zum19] Patrick Zumsteg. *Complex Event Detection and Event Based Retrieval in eSports*. Master's thesis, University of Basel, 2019.