An Ontology for Ice Hockey

Robin Keskisärkkä, Huanyu Li, Sijin Cheng,
Niklas Carlsson, and Patrick Lambrix

Linköping University, Linköping, Sweden
firstname.lastname@liu.se **

Abstract. Ice hockey is a highly popular sport that has seen significant increase in the use of sport analytics. To aid in such analytics, most major leagues collect and share increasing amounts of play-by-play data and other statistics. Additionally, some websites specialize in making such data available to the public in user-friendly forms. However, these sites fail to capture the semantic information of the data, and cannot be used to support more complex data requirements. In this paper, we present the design and development of an ice hockey ontology that provides improved knowledge representation, enables intelligent search and information acquisition, and helps when using information from multiple databases. Our ontology is substantially larger than previous ice hockey ontologies (that cover only a small part of the domain) and provides a formal and explicit representation of the ice hockey domain, supports information retrieval, data reuse, and complex performance metrics.

1 Introduction

While sports analytics in the past was limited to simple high-level statistics based on manually extracted data, the development of new technologies (e.g., optical object tracking) supporting the automatic annotation of games has led to increasing amounts of available play-by-play data, containing details about each play event and its context (e.g., detailed game state, player positions, puck/ball position, and timestamps). To gain a competitive advantage many teams are already continually analyzing this data, looking for an edge on their competitors.

Today, play-by-play data and other statistics are provided by many of the major ice hockey leagues, including the National Hockey League (NHL) in North America (US+Canada) and the Swedish Hockey League (SHL). There are also public websites that present statistics based on such data in human friendly formats; e.g., Corsica (`http://corsica.hockey`) and Natural Stat Trick (`http://www.naturalstattrick.com`). These sites typically show a limited range of performance metrics, and cannot support complex query requirements or detailed insights of play-by-play data.

Ontologies provide a formal and explicit representation of the domain knowledge, which can greatly benefit information retrieval and data reuse, and support

** Copyright © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

This paper appears in
ISWC 2019 Posters & Demonstrations, Industry and Outrageous Ideas Tracks, Auckland, New Zealand, 2019

advanced performance metrics, such as those proposed in recent research on ice hockey analytics (e.g., [3,4,1,2,5]). Prior work focusing on ice hockey ontologies is very limited, and existing ontologies cover only a small part of the domain.

In this paper, we present the development of an ice hockey ontology that extends the coverage of previous efforts, map play-by-play data to the Resource Description Framework (RDF), and validate that our solution easily can be used to effectively answer example questions (otherwise not easily accessible) using SPARQL queries. The ontology enables semantics-based access to existing data, as well as the integration of different data sources. In addition to being available on the web, the ontology will be used in applications related to ice hockey analytics and data visualization in cooperation with a professional ice hockey team. The design and development of the ontology and its current state are discussed in Sections 3 and 4, respectively.

2 Related work

There is little ontology-related work focusing on ice hockey. The International Press Telecommunications Council, which develops industry standards for the exchange of news data, has developed a sports ontology (`https://iptc.org/std/SportsML/3.0/documentation/`). The ontology defines general sports-related concepts. While the ontology benefits from many concepts being shared across sports, some terms end up being overloaded, leading to the exact interpretation often being dependent on the actual sport under consideration. The ice hockey-specific part of the ontology deals with traditional player and team statistics, and simple event states related to power play and scoring. Similarly, BBC developed a lightweight ontology (`https://www.bbc.co.uk/ontologies/sport`) for representing sports events with a focus on the organization of competitions. Finally, DBpedia (`https://wiki.dbpedia.org/`) contains some ice hockey-related terms such as ice hockey league and ice hockey player.

3 Ontology design and development

The design and development of the ontology included four high-level steps: (1) description of use cases, (2) specification of competency questions, (3) formalization, and (4) validation. The process was implemented in an iterative fashion, with refactoring, revisions, and refinements to ensure that the ontology was expressive enough to capture the competency questions.

Use cases: The use cases aim at providing ice hockey knowledge to general users and to support professionals in the domain, including head coaches, players, general managers, and team scouts. This includes role-specific uses cases requiring support for advanced data analytics of play-by-play data.

Competency questions: According to the use cases, a set of competency questions (CQs) were specified and categorized as either game related, event related, or performance-metrics related. Examples of representative CQs were: (1) *How many games end during the regular-time period?*, (2) *When did the*

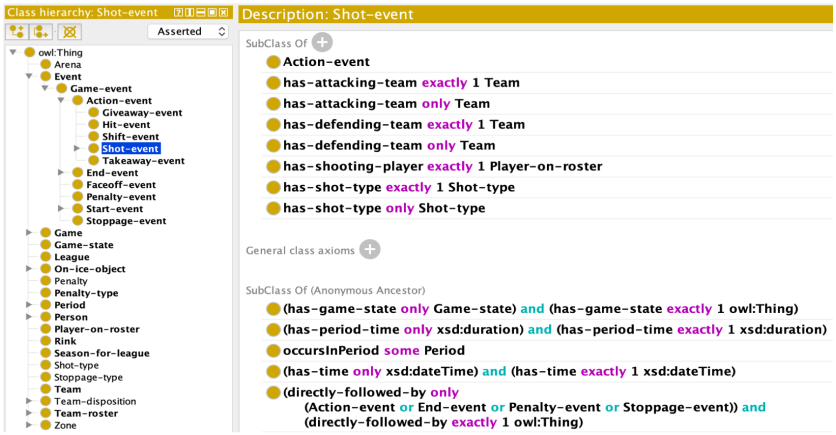

Fig. 1. Overview of the ice hockey ontology.

winning goal happen for a specific game?, and (3) *What is the faceoff winning percentage in the last X games of a specific player?*. The CQs were used both to provide the scope of the ontology, and to provide some way of validating the ontology with respect to the use cases.

Formalization: Based on the use cases and CQs, we conceptualized ice hockey related concepts, starting from NHL play-by-play data and the NHL rule book. We then used OWL to formalize the ontology in Protégé. Starting from a set of general concepts in ice hockey (e.g., game, event, person, team), we extended the concept hierarchy by specializing these concepts. For example, the event concept was specialized into penalty event, action event, etc. Furthermore, we defined class properties and constructed semantic relationships.

Validation: We validated the ontology using the OOPS! service, the HermiT reasoner, and RepOSE. We then mapped the play-by-play data to RDF using the RDF Mapping Language (RML), converted the data to RDF, and provided validation tests for each CQ using one or more SPARQL queries.

4 Current coverage

The current version of the ontology can be used to represent: (1) basic knowledge about the ice hockey domain, (2) game events and game sequences, and (3) describe the game context of events.

The ontology currently contains 125 concepts, 100 relations, and 892 axioms. Figure 1 shows an overview of the concept hierarchy, and a detailed description of *Shot-event*. The general concepts covered in the domain include, for example, *Arena*, *Game*, *League*, *Penalty*, *Period*, *Person*, *Team*, as well as concepts on the event level such as *Game-event*, with *Action-event* and *Faceoff-event* as sub-concepts, and *Game-state* to represent the event context.

Shot-event is defined as a sub-concept of *Action-event* which in turn is a sub-concept of *Game-event*. As shown by the axioms on the right-hand side of

Figure 1, we define a *Shot-event* from the perspective of the attacking team, and include the shooting player and the shot type (e.g., slap shot). The bottom half of the right-hand side of Figure 1 shows the axioms that *Shot-event* inherits from *Game-event*. The first axiom specifies that a *Shot-event* has a specific game context (*Game-state*), capturing the set of players on the ice for both the home and away team, and the team disposition, for example.

Mapping play-by-play data to RDF: The amount and quality of match-specific data available for ice hockey varies greatly between different leagues, and the features reported depend largely on the type of systems employed. Typical datasets referred to as play-by-play data include a discretized representation of ice hockey games, where timestamped events deemed relevant for post-game analysis have been recorded along with some contextual information. Also the formats used to represent such datasets differ greatly, varying from fully normalized database dumps with foreign key relations, to tables of semi-structured data. In this paper, we focus on the representation of (both public and private) datasets provided for NHL and SHL, and provide RML mappings to declaratively capture the data based on the proposed ontology.

Evaluation: After constructing the ontology and creating mappings from play-by-play data to RDF, we validated the ontology by answering CQs related to the use case. For each CQ, we created one or more SPARQL queries to validate if the ontology captured sufficient information to answer all CQs.

5 Conclusion

We have presented ongoing work on developing an ice hockey ontology that conceptualizes general ice hockey domain knowledge and events in play-by-play data. Based on the proposed ontology, we provided a mapping of play-by-play data to RDF using RML, and validated the ontology against a set of SPARQL queries solving competency questions derived from different role-specific use case. The ontology provides a formal and explicit representation of the domain knowledge that supports information retrieval, data reuse, and can help in the retrieval of more advanced performance metrics from play-by-play data in ice hockey.

References

1. Liu, G., Schulte, O.: Deep reinforcement learning in ice hockey for context-aware player evaluation. In: IJCAI. pp. 3442–3448 (2018)
2. Ljung, D., Carlsson, N., Lambrix, P.: Player pairs valuation in ice hockey. In: Machine Learning and Data Mining for Sports Analytics, LNCS 11330. pp. 82–92 (2019)
3. Macdonald, B.: An Expected Goals Model for Evaluating NHL Teams and Players. In: MIT Sloan Sports Analytics Conference (2012)
4. Routley, K., Schulte, O.: A Markov Game Model for Valuing Player Actions in Ice Hockey. In: UAI. pp. 782–791 (2015)
5. Sans Fuentes, C., Carlsson, N., Lambrix, P.: Player impact measures for scoring in ice hockey. In: MathSport International. pp. 307–317 (2019)