Language Interpretation and Generation for Football Commentary

A. Nijholt, R. op den Akker & F. de Jong

Center for Telematics and Information Technology
TKI-Parlevink Research Group
University of Twente,
PO Box 217, 7500 AE Enschede
The Netherlands

Abstract

The objective of the EU/IST funded MUMIS project is to develop technology for automatic indexing of multimedia programme material (texts, news streams, speech transcripts) and to develop a user interface that supports the conceptual querying and browsing of related video content over the internet. One of the innovative features of MUMIS is that it aims at the disclosure of video archives by applying information extraction techniques, originally developed for the textual domain. In MUMIS the extraction technology is applied to the domain of football.

In this report we survey a number of research efforts that all deal with football commentary but for which the technology focus differs: retrieval, interpretation or generation of commentary and related, not necessarily language oriented, research. There were two reasons for this investigation. One obvious reason is to find out how the MUMIS technology and tools can be adapted to similar applications. The other is to see how the domain knowledge obtained for extraction purposes can be employed for other intelligent applications in the same domain.

Language Interpretation and Generation for Football Commentary

A. Nijholt, R. op den Akker & F. de Jong

Center for Telematics and Information Technology
TKI-Parlevink Research Group
University of Twente,
PO Box 217, 7500 AE Enschede
The Netherlands

Abstract

The objective of the EU/IST funded MUMIS project is to develop technology for automatic indexing of multimedia programme material (texts, news streams, speech transcripts) and to develop a user interface that supports the conceptual querying and browsing of related video content over the internet. One of the innovative features of MUMIS is that it aims at the disclosure of video archives by applying information extraction techniques, originally developed for the textual domain. In MUMIS the extraction technology is applied to the domain of football.

In this report we survey a number of research efforts that all deal with football commentary but for which the technology focus differs: retrieval, interpretation or generation of commentary and related, not necessarily language oriented, research. There were two reasons for this investigation. One obvious reason is to find out how the MUMIS technology and tools can be adapted to similar applications. The other is to see how the domain knowledge obtained for extraction purposes can be employed for other intelligent applications in the same domain.

1 Introduction

Our interest in the computer processing of football commentary was at first given in by the EU/IST funded project MUMIS (Multimedia Indexing and Searching) that started in 2000 [11, 18, 19]. This project's objective is to develop technology for automatic indexing of multimedia programme material (texts, news streams, speech transcripts) and to develop a user interface that supports the conceptual querying and browsing of related video content over the internet. One of the innovative features of MUMIS is that it aims at the disclosure of video archives by applying information extraction techniques, originally developed for the textual domain. Information extraction is a technique that is typically suited for content in specific domains. In MUMIS the extraction technology is applied to the domain of football. The project requires the integration of lexicons, ontology and information extraction tools for this domain, and the development of merging algorithms to integrate the (incomplete) information coming from different sources. For example, the information extraction components should be able to extract some thirty different

event types, using methods such as part of speech tagging, syntactic parsing, semantic tagging, and discourse analysis. Typical football events to be detected are: kick-off, penalty, goal, half-time, free kick, etc. The user interface should help users to formulate queries that can be matched on the annotations generated by the extraction component and linked to the time-codes of the corresponding video fragments.

In parallel to the MUMIS project we surveyed a number of research efforts that deal all with football commentary but for which the technology focus differs: retrieval, interpretation or generation of commentary and related (but not necessarily language oriented) research. There were two reasons for this investigation. One obvious reason is to find out how the new technology and tools can be adapted to similar applications. The other is to see how the domain knowledge obtained for extraction purposes can be employed for other intelligent applications in the same domain. Since at the same time several of our M.Sc. students got interested in learning multi-agent systems for developing teams for the RoboCup leagues we decided to stick close to the football domain in this survey as well.

In this paper we start with the survey of the domain specific research. The aim of the survey was to obtain a comprehensive view of the field that could guide the selection of new research themes. In the second part of the paper we give a short introduction to the MUMIS project as it can be embedded in the general football related language technology research.

2 Background

If we look at football (soccer) we first can distinguish between interpretation and generation (including presentation during retrieval sessions). Depending on the role of the computer (does it take part in the game as a robot or a software agent, is the game more or less computer-controlled or observable) we can also give the computer more or less online or offline capabilities to interpret what is or has been going on in the stadium (players, spectators, coaches, referees, ball, commentators in different roles, etc.).

The online interpretation requires the observation of events and the translation from visual to verbal representations or the online translation from a news stream produced by a mediator (e.g., a human commentator or the maker of a ticker report) to a textual representation that can be used for future analysis and retrieval purposes. Retrieval does not necessarily have to yield text results. If in a multimedia search environment the indexes may contain links to video or audio time codes, the objects retrieved can also be spoken (radio) commentary or video fragments.

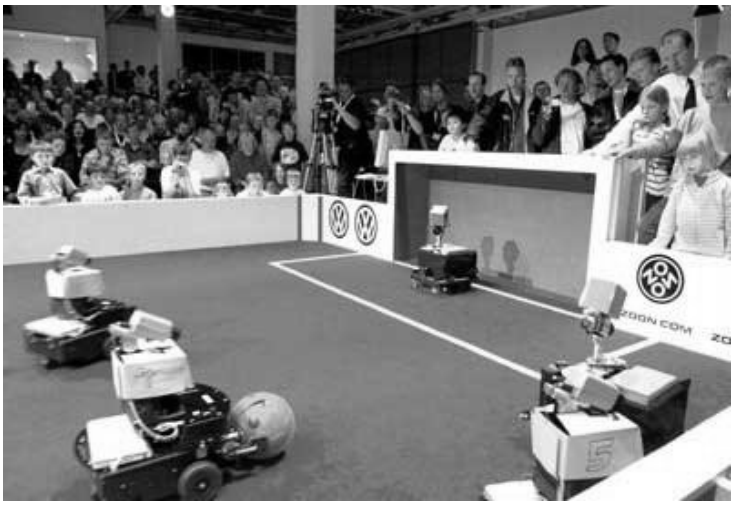

Figure 1: RoboCup, middle-size league

It may also be possible to regenerate part or all of the game (action replay) in a virtual world with 3D players, allowing interaction be-

tween the user and the synthetic players, including interactive performance. Games that can be distinguished are the football games that take place in stadiums in the real world, the RoboCup leagues, where the human players are replaced by physical robots (see Figure 1) and the RoboCup simulation leagues (virtual football), where we have 2D or 3D virtual robots and the field only exists on the computer screen (see Figure 2).

'Soccer Server' [14] is a simulator of football that allows a team of players to play matches against other teams. It provides also an environment to gather information about the performance of multi-agent systems and to apply learning algorithms to improve performance.

For completeness we mention also the football computer games that can be played on play stations (FIFA 2001) and the possibility that in the future we can have teams of human players versus robot players or mixed teams of robot and human players.

3 Commentary Systems

Automatic commentator systems have been designed and implemented for the RoboCup and the RoboCup simulation leagues. Obviously, in these cases, although the agents and the teams have individual and team behavior and strategies, actions of teams and players can be tracked by the computer(s) (with e.g. camera and sensor support in the RoboCup situation) and this can be the basis for a textual description or an automatic commentary system. However, commentary is often more than a ticker report. There is a difference between on the one hand the mentioning of actions and analysis of what is going on and on the other hand providing background information. For real games this is often done by different commentators, a play-by-play commentator and a color commentator.

Figure 2: RoboCup simulation

An automatic commentator should have access to the implemented strategies of the teams or the individual goals of the players in order to interpret the actions correctly. Obviously, as in real games, a commentator may interpret different actions in view of some assumed general goals and agent-specific goals [6] and reason about them and about player interaction. When due to a severe collision two robots break down, a commentator that relies on information coming from the individual robots only cannot provide further information.

An example of an automated commentary system is MIKE (Multi-agent Interactions Knowledgeably Explained). It comments in real-time using spoken or plain English, French or Japanese.

Figure 3: MIKE as an embodied agent

It was first used in a simulator league in 1998 (as a single commentator). Later it was extended to a commentary team (announcer and expert) for real-world robotic football, allowing the expert to comment on a higher level using statistics (e.g., pass patterns) and knowledge of the current situation (e.g., the defense area of a player is too large) [13, 21]. A more general remark, when there is a one-one relation between the commentary software and the team's play-plan selection, an almost perfect commentary can be given. Obviously, a spectator can be much more amused when commentators make errors.

A second example we like to mention here is the RoCCo (RoboCup Commentator) system [1]. RoCCo is based on an interesting example of research (VITRA: Visual Translator) that allows the translation of visual data obtained by high-level scene analysis towards multimedia reports, including visualization of inferred information in virtual-reality-like replays of events and natural language output. RoCCo does not utilize all aspects of the translation system. Rather it uses the existing strategies and architecture to obtain scene interpretations that are input to offline plan-based reports (allowing organization of the data) and live reports (only local, partially planning of the information presentation) on a RoboCup simulation game.

Events that take place during a (simulated) RoboCup game have to drive the automatic commentary. Therefore natural language generation and speech synthesis is driven by the events that take place. A quite common method to be used are domain-dependent template methods where the templates have slots that need to be filled by – this case – player names, football actions, positions, etc. More advanced method should make use of text generation grammars or concept-to-speech synthesis methods. For particular applications other methods have been used, see e.g. [22] for a table-driven method to generate football reports. In the case of football a play-by-play input can be used, making simple commentary, or a more advanced (online) game-analysis system – e.g. the systems underlying MIKE or RoCCo, can be used as the basis for providing commentary. In [3] different commentary

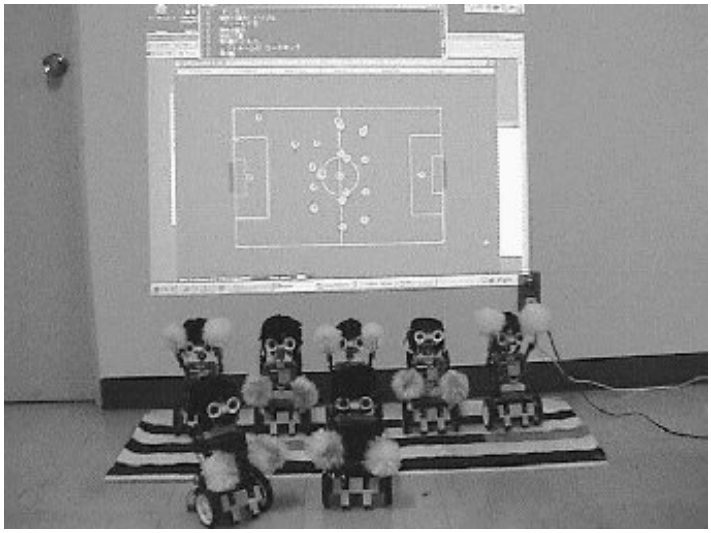

Figure 4: Seven cheerleader robots performing

systems are compared.

Football games played on play stations don't seem to add essential new viewpoints to the generation of commentary. Clearly, as argued in [15, 16], also here the quality of the generated commentary is very much crucial for the fun experience players have. We will not elaborate on language technology related to mixed human-robot football games.

4 Embodied Agents and Commentary

Various research groups have been working on embodied agents or talking heads that allow more expressive running commentary during a simulation league match. For example, in [4, 5] the Byrne architecture is discussed. Byrne does not only provide commentary (using templates based on play-by-play information), but has also appropriate linguistically motivated affective speech and facial expression (using part-of-speech information).

In the case of MIKE some extensions have been made that can amuse the spectators of real-world robotic football. E.g., a physical embodiment (SIG) that also takes into account social interaction with the audience of the game using audio and video data streams ([8], see also Figure 3). A further step has been the adding of a robot cheerleader team that performs cheerleading commentary on the game in order to elicit responses of the audience ([12], see also Figure 4).

In [2] the work on team commentary in the RoCCo system is generalized to presentation teams where the members have different personalities. These teams are not necessarily commentators, they can be car sellers or football fans commenting on a game and modeled as semi-autonomous agents whose actions are triggered by events during the game.

5 Merging, Extraction and Video Retrieval in MUMIS

In this section the MUMIS project will be discussed in more detail. In particular the problems of domain specific language processing will be focused on, as well as the approaches for handling them within an information extraction module within the football domain.

One of the MUMIS project goals is realization of a prototype search environment. For this prototype the European Football Championships 2000 has been selected as the main source for the multimedia content database. For the textual part it will eventually contain various types of documents: newspaper articles, automatically generated speech transcripts and ticker texts. Tickers are a football match report type consisting of minute-by-minute brief free text descriptions of the most remarkable things that happened during the match. A ticker describes a scene and starts with a time stamp.

This is the description of what happened in the third minute of Nederland-Joegoslavie (Netherlands – Yugoslavia):

```
03: Nederland dreigt meteen. Bergkamp kan een mooi opgezette aan-
val net niet afwerken. Zijn schot vindt het lichaam van
Mihajlovic.
```

(Immediately the Dutch start threatening. Bergkamp fails to complete a nicely set up attack successfully. His shot finds the body of Mihajlovic.)

Thirty-one event types have been selected that are to be extracted from the ticker texts, the newspaper reports and the spoken reports of the matches of the European Championships. In Figure 5 these events are collected.

Final whistle, Offside, Goal kick, Interception, Kick-off, Chested ball, One-two, Dribbling, Corner, Own goal, Throw-in, Cross, Foul, Free-kick, Yellow card, Halftime signal, Hand ball, Header, Hitting crossbar, Pass, Hitting post, Clearance, Red card, Overhead kick, Tackle, Penalty, Goal, Shot on goal, Save, Injury, Substitution

Figure 5: MUMIS event types

One of the events that we can identify in the example ticker text above is of type Shot-on-goal, with actor attribute value Bergkamp. Although the text does not mention the event literally, the second and third sentences together more or less imply that Bergkamp tried a shot on the goal. A second event, to be inferred from the third sentence, is of the type Interception, with actor value Mihajlovic.

Within MUMIS the extraction of events takes place via the automatic generation of formal annotations that reflect the typical output of IE systems, i.e. instantiated domain-specific templates or event tables. Tools are needed to identify and interpret scene descriptions in terms of these filled tables and templates. To be able to build an extraction module along these lines, a corpus of typical football report idioms, idiosyncratic sayings and phrases has to be collected, together with their semantics in terms of football events. Also more general syntactic analysis tools are needed. Anaphora resolution for example is needed to detect that in the above ticker text that His refers to Bergkamp and that he shot at the goal rather than finishing the attack by a header ball. Cf. [24].

The same reporter wrote:

```
11: Mihajlovic toont z'n specialiteit: een banaanschot op doel.
Van der Sar plukt de banaan.
```

(Mihajlovic demonstrates his specialty: a banana shot at the goal. Van der Sar picks the banana.)

Van der Sar is the Dutch goalkeeper who catches the banana. The word 'banana' refers to the shot on goal that is typified as a banana shot. Hence the following two events can be identified: Shot-on-goal by Mihajlovic and a Save by Van der Sar.

It could very well be that in another report the same event would be described as "Mihajlovic took a free kick in the 10th minute". To detect that the scene in described in the ticker text for minute 11 and the scene described in other source as a free-kick event in minute 11 are one and the same event, the MUMIS system incorporates a so-called merging component. The role of this merging component is to generate one possibly complex event descriptions on the basis of multiple event descriptions and to complete incomplete descriptions. In the case of conflicting information sources, a decision is made on the basis of a reliability score.

In minute 31 the reporter returns to the banana metaphor:

```
31: Mihajlovic zendt nog wel een van zijn bananen richting doel
maar weer redt Van der Sar.
```

(Mihajlovic sends another of his bananas in the direction of the goal, but again Van der Sar rescues.)

In minute 67 when “de zoveelste banaan van M. naast de paal valt” (another of the many bananas of M. falls beside the goal-post) the reporter writes: “De banaan valt niet ver van de paal.” (“The banana does not fall far from the post”, making a reference to the Dutch version of the proverb “as the tree, so the fruit”.) As the use of this metaphor is not characteristic for the ticker reporters in general, but an idiosyncratic feature of the style of one reporter it is unlikely that the domain model that is the driver of the extraction module would cover it. This *Shot-on-goal* description is therefore more likely detected by the merging component.

As said, an automatic information extraction system has to be trained on a corpus of domain specific texts. There are several methods to abstract and generalize the knowledge and language in a corpus; either by making a conceptual or semantic network coupled with a lexicon of domain specific phrases or by generalizing directly from the text, by matching a new text with already seen and annotated text fragments.

Within the MUMIS ontology, the most important concepts are events and entities. For a large part of the data the is_a relationship suffices to describe the dependencies between the concepts (a GoalKeeper isa Player isa Person). The concepts in the ontology are linked with lexicons for the three project languages: English, German and Dutch. Other aspect of the functionality of the information extraction system are covered by modules for named entity recognition (using name lists of players, teams, referees, stations, countries, cities), POS-tagging, shallow parsing, and discourse modeling. The final merging component needs knowledge from the football domain; typical events, scene descriptions, scripts and the roles that players have in these scenes to fill missing information output by the information extraction components, to identify the correct order of events, and to make decisions where conflicting information is provided.

6 Interpretation and Generation, Retrieval and Replay: Mutual Benefits

Whether we talk about videos of real-world football, real-world robot football (RoboCup) or simulation league football, there is a translation from visible events to a digital representation that can be manipulated later. Human mediators, microphones, cameras and sensors help to obtain this digital representation and further analysis and merging of information coming from distant sources is necessary to obtain input for further processing. For instance, to become input to a game-analysis system, to generate a match report, provide a running commentary, or to retrieve and replay information offline.

In order to be able to understand by computer what is going on or to give (when necessary real-time) computer support to commentators, match report makers, or users interested in retrieval or replay of events, video scenes need to be identified and localized. Obviously, real-time information can also be useful for referees, but until now this has not been allowed for real-life football. In RoboCup information about positions and movements of players is automatically available, but not sufficient to generate entertaining reports. Scene analysis includes tracking of the players, the ball and the referees. This can be based on analyzing successive frames, but underlying model-based approaches (motion analysis, ball trajectories, etc.) can improve results. As an example, in [9] these approaches are part of a real-time scene analysis that aims at continuous tracking of the 3D position of the ball, whether it is hidden behind a player or not.

Clearly, methods like these can also provide input for video retrieval and automatic indexing, for game-analysis systems underlying commentary systems and for match report generation. In many research groups, including the MUMIS consortium, more or less explicit ontology's for the domain of football are being developed in order to be able to describe objects, events and even more high-level concepts (intentions, strategies) from which we can build tools and systems.

For commentary and match report systems it is as well useful to analyze corpora of newspaper articles and commentaries as for MUMIS-like retrieval purposes. In [15] comprehensive analysis of play-by-play and (the interplay with) color commentary can be found. As an other example, in [7] a global statistics-based analysis of a corpus of newspaper match reports (from *The Times*) is presented. The analysis is meant to give insights to the designers of RoboCup teams and to obtain formalizations to be used for algorithms that improve the interaction between team members.

We assume that much deeper insights and formalizations can be obtained with the type of linguistics analysis tools made available in the MUMIS project. On the other hand, Soccer Server, mentioned above, seems to be an excellent environment to develop, compare, evaluate and improve commentary systems or systems that automatically generate match reports. Moreover, it allows comparisons between human and computer-generated commentary and reports. Data gathered by such a server allows the application of learning algorithms, not only for improving team member's interaction and the team's strategy, but also for improving commentary and generating match reports. In general we assume that there are mutual benefits when event interpretation, event retrieval and event generation are based on the same formal representation models and tools based on these models. Rather isolated research efforts that now occur in the domain of football can be brought together in order to obtain these benefits.

7 References

[1] E. Andre, G. Herzog & T. Rist. Generating multimedia presentations for RoboCup soccer games. http://www.dfki.de/vitra/papers/robocup97/.

[2] E. Andre, T. Rist, S. van Mulken, M. Klesen & S. Baldes. The automated design of believable dialogues for animated presentation teams. *Embodied Conversational Agents*. J. Cassell, S. Prevost, J. Sullivan & E. Churchill (eds.), The MIT Press, 2000, 220-255.

[3] E. Andre, K. Binsted, K. Tanaka-Ishii, S. Luke, G. Herzog & T. Rist. Three robocup simulation league commentary systems. *AI Magazine* 21 (1), 2000, 57-66.

[4] K. Binsted & S. Luke. Character design for soccer commentary. *RoboCup-98: Robot Soccer World Cup II*, M. Asada & H. Kitano (eds.), LNAI 1604, Springer, 1998, 23-35.

[5] K. Binsted. A talking head architecture for entertainment and experimentation. AAAI Fall Symposium *Emotional and Intelligent: The Tangled Knot of Cognition*, 1998.

[6] H.-D. Burkhard, M. Hannebauer & J. Wendler. BDI Deliberation in Artificial Soccer. *AI Magazine* 19 (3), 1998, 87-93.

[7] I. Frank. Football in recent times: What we can learn from the newspapers. *RoboCup '97: Robot Soccer World Cup I*, H. Kitano (ed.), Springer, 1997, 216-230.

[8] I. Frank, K. Tanaka Ishii, H. G. Okuno, K. Nakadai, Y. Nakagawa, K. Maeda & H. Kitano. And The Fans are Going Wild! SIG plus MIKE. *Fourth Workshop on RoboCup (RoboCup-2000)*, Melbourne, LNAI, Springer, 2000, 267--276.

[9] I. Haas, R. Maierhofer, R. Sendlhofer & W. Polzleitner. Real-time scene analysis for the localization of the ball position in soccer games. *SCI 2002*, Orlando, 2002.

[10] G. Herzog. From Visual Input to Verbal Output in the Visual Translator. TR 124, Universität des Saarlandes, July 1995.

[11] F. de Jong & T. Westerveld. MUMIS: multimedia indexing and searching. *Content-Based Multimedia Indexing (CBMI 2001),* 2001, 423-425.

[12] N. Mastunami, K. Tanaka-Ishii, I. Frank & H. Matsubara. Cheerleading soccer commentary robots. International workshop on *Intelligent Media Technology*, T. Nishida (ed.), *PRICAI-02* workshop, Tokyo, 2002.

[13] H. Matsubara, I. Frank, K. Tanak-Ishii, I. Noda, H. Nakashima & K. Hasida. Automatic soccer commentary and RoboCup. *RoboCup-98: Robot Soccer World Cup II*, M. Asada & H. Kitano (eds.), LNAI 1604, Springer, 1998, 7-22.

[14] I. Noda, H. Matsubara, K. Hiraki & I. Frank. Soccer server: a tool for research on multi-agent systems. *Applied Artificial Intelligence*, 12, 2-3, 1998, 233-250.

[15] S. Nowson. Being John Motson: Toward a computational model of football commentary. M.Sc. thesis, U. of Edinburgh, 2001.

[16] S. Nowson. Increase the fun. In: Proc. *CHI '02* Workshop - *Funology: Designing Enjoyment*. April, Minneapolis, 2002.

[17] F. Ribeiro, C. Machado, S. Sampaio & B. Martins. Image processing applied to a robotic football team. In: Proc. of *EuRoboCup 2000* Workshop, G. Adorni & W. Hoeke (eds.), CD Rom, Amsterdam, 2000.

[18] H. Saggion, H. Cunningham, D. Maynard & Y. Wilks. Using the GATE architecture for named entity recognition in the football domain. 2nd *Spanish Natural Language Processing Workshop*, 2001.

[19] H. Saggion et al. MUMIS Deliverable D4 (VI): Text annotation tool. 2001.

[20] K. Tanaka-Ishii, K. Hasida & I. Noda. Reactive content selection in the generation of real-time soccer commentary. *COLING '98*, 1998.

[21] K. Tanaka-Ishii, I. Noda, I. Frank, H. Nakashima, K. Hasida & H. Matsubara. MIKE: An Automatic Commentary System for Soccer. Third International Conference on *Multi-Agent Systems*, Y. Demazeau (ed.), 1998, 285-292.

[22] M. Theune. GoalGetter: predicting contrastive accent in data-to-speech generation. In: *Computational Linguistics in the Netherlands*. Papers from the Seventh CLIN Meeting, Eindhoven, The Netherlands, 1997, 177-190.

[23] D. Voels, E. Andre, G. Herzog & T. Rist. Rocco: A RoboCup soccer commentator system. *RoboCup-98: Robot Soccer World Cup II*, M. Asada & H. Kitano (eds.), LNAI 1604, Springer, 1998.

[24] Th. Declerck, P. Wittenburg & H. Cunningham. The Automatic Generation of Formal Annotations in a Multimedia Indexing and Searching Environment. ACL/EACL Workshop on Human Language Technology and Knowledge Management, 129-136, 2001.