Referring Expressions in Artificial Intelligence and Knowledge Representation Systems

(a tutorial proposal for ECAI 2024 in Santiago de Compostela, Spain)

David Toman and Grant Weddell

Cheriton School of Computer Science, University of Waterloo
200 University Ave W., Waterloo, ON N2L3G1, Canada
{david, gweddell}@cs.uwaterloo.ca

Abstract

The tutorial introduces the audience to the concept of *referring expressions*, formulæ that can be used to communicate identities of otherwise abstract objects. The formalism provides foundations for a successful and unambiguous exchange of information about individuals between agents sharing common knowledge about such individuals, a task that is indispensable in most modern applications of knowledge representation and semantic technologies.

Proposed Length and Format of the Tutorial

The tutorial is designed for 1/2 day. This provides sufficient time for an in-depth presentation on the topic based on our experience in delivering earlier versions both in-person and on-line (see below). However, the tutorial structure anticipates alternative timing arrangements and can be presented in an abbreviated form as a 1/4 day tutorial (i.e., 90-120 mins), should that be preferable for the overall ECAI schedule. For an ECAI-sized conference we expect about 30-40 attendees (based on past editions of this tutorial).

Description and Objectives of the Tutorial

A *referring expression* in linguistics is any noun phrase identifying an object in a way that will be useful to interlocutors. In the context of knowledge representation and information systems, *constant symbols* occurring in an underlying knowledge base are the artifacts usually used to identify a subset of the objects for which the knowledge base captures knowledge.

This tutorial explores how the set of objects that can be usefully identified can be extended by allowing more general expressions in the underlying language of the knowledge base, called *singular referring expressions*, to replace constants as syntactic identifiers of such objects. Expanding the possibilities of identifying possibly implicitly defined objects serves numerous purposes, ranging from enabling query answers to contain additional results which would otherwise not be possible due to a lack of constant symbols denoting (components of) such results, to answers that are more informative, to decisions on how to communicate references to objects among various cooperating agents, and to identification issues related to physical data representation in computer storage, for example, to *addresses* of records located in main-memory databases. The idea of referring expressions itself circumvents the need for artificially-invented identifiers that are commonly opaque to the user that interacts with the knowledge base.

The tutorial has two main goals: (1) to review the history of approaches to identifying individuals/objects, in particular in the area at the intersection between logic-based approaches to knowledge representation and the areas of natural languages and philosophy, approaches that can be traced back to Russell [Russell, 1905], Frege, and Strawson [Strawson, 1950, Strawson, 1956] including the use of such expressions in NLP systems [Dale and Haddock, 1991, Reiter and Dale, 1992, Dale and Reiter, 1995], and (2) to present a modern unifying approach to referring expressions based on logical methods in AI and Knowledge Representation with clear and direct applications to many practical areas of Semantic Technologies, AI, and Computer Science in general.

Relevance to ECAI

The tutorial focuses on foundational issues relating to object identification in knowledge bases, ontologies, and information systems based on ontologies, and on how such issues can be comprehensively addressed. Since every design of an information system faces decisions relating to how external entities will be identified within such a system (in addition to representing various properties of such entities), a general approach to this problem is of interest to ontology developers/engineers and data scientists. Interestingly, the approach to object identification discussed in the tutorial naturally and seamlessly complements standard approaches in conceptual and ontology design methodologies. The tutorial is thus of interest both to researchers in knowledge representation and to practitioners in the wide area of information management.

Outline of the Tutorial

Attendees of the tutorial will be able to identify and analyze issues related to identification of abstract entities, information about which is captured using various semantic technologies, such as knowledge graphs and other knowledge representation systems. The following outlines the major take-home points that the participants will learn about:

What is a referring expression? We start with an introduction and overview of how *well formed formulae* that satisfy certain properties can serve as *referring expressions* in information systems whose underlying ontologies correspond to first order knowledge bases.

Background. We introduce formal properties of referring expressions and show how they can be determined. We discuss how referring expressions can be computed, in particular when the knowledge base conforms to a decidable fragment of first order logic. We also review past work on determining referring expressions in the context of knowledge bases and position these approaches among other approaches designed to indirectly and/or symbolically capture identities of (sets of) objects.

Referring expressions, types, and query answering. We show how referring expressions can be used to enrich query answers over knowledge bases by allowing to refer to answer objects that may not have an explicit name within the knowledge base, or for which a more *preferred* way of communicating its identity is available. To control the form of the answers, we define a type language that describes varieties of referring expressions desired in query answers.

Referring expressions and types in conceptual modelling. We then we explore the benefits of adopting referring expression types for use in information systems derived from conceptual modelling. In particular, we show how this approach can separate purely conceptual ontology design from issues connected with how objects are identified within an eventual information system based on the design.

Open problems. We conclude the tutorial with an outline of directions for further research, and with a list of open issues related to the use of referring expressions in ontology-based information systems.

Topics Covered

The tutorial discusses the efficacy of referring expressions in several settings:

1. The first considers the issues with query answering over knowledge bases over which (parts) of query answers may not have an explicit identifier, in particular, are not denoted by a constant symbol. The tutorial shows how *referring expressions* can be used to enable a wider range of answers in comparison with standard approaches. In the context of tractable knowledge representation languages, e.g., based on *description logics* that are fragments of first order logic, the tutorial outlines how identification properties can be determined at compile-time for conjunctive queries, and how off-the-shelf conjunctive query answering approaches for these logics, such as perfect rewriting, can be used in query evaluation.
2. In the context of conceptual modelling, referring expression types yield a separation of a concerns in conceptual design, in particular the choice of how conceptual objects are identified in the final information system. The tutorial starts from a simple object-centered representation common in ontologies and in virtually all semantic data models where, since objects are self identified, naming is not an issue, and then proceeds to show how an analyst can attach *referring expression types* to classes, and to specify appropriate uniqueness constraints to satisfy the requirements on referring expressions.

3. The tutorial also discusses HCI experiments that validate the utility of a referring expression-based approach to modelling object identity. The experimental setup measures the performance and correctness of both authoring and understanding queries in comparison to the standard state of the art approaches.

In all cases, the tutorial lays a foundation for the use of singular referring expressions in knowledge representation and on the computation of answers to queries. An integral part of this foundation are characterization theorems for identification properties of referring expressions for queries annotated with a domain specific language for referring expression types. Parts of the tutorial rely on results published in [Borgida et al., 2016a], [Borgida et al., 2016b], and [Ma et al., 2018].

Audience and Background

The topics covered in the tutorial are of interest to wide range of AI researchers and to members of the general public with an interest in knowledge representation. The tutorial targets the following groups:

- Undergraduate and graduate students and junior researchers: the tutorial introduces this group to state-of-the-art approaches to addressing identification issues in knowledge bases and to modern techniques that address these issues;
- Researchers in the area of knowledge representation and other areas of AI: the tutorial provides bridges to many areas of AI, ranging from natural language issues, where the idea of *referring expressions* has originated, to philosophical underpinnings of object identification to unambiguous communication of information between agents, and to identity management in information and semantic WEB systems;
- Industry practitioners and developers: the tutorial provides ideas on how development of software systems, in particular in the critical phase of conceptual modelling, can be improved and what tools are available to aid this goal;
- Members of the general public, with an interest in logical underpinnings of logic-based information management and in technologies based on these ideas.

The tutorial assumes the audience is familiar with the basics of first order logic and of conceptual modelling formalisms (such as ER or UML) at the introductory university course level. No knowledge of particular ontology/KR languages such as *description logics* and other formalisms is assumed.

Previous Presentations of the Tutorial

The authors, together with Alexander Borgida (Rutgers), have introduced referring expressions in the area of Knowledge Representation and Ontology-based Data Access [Borgida et al., 2016a] and were awarded the *Ray Reiter Best Paper prize* at KR 2016 for this work. They have extended this work to the area of conceptual modelling [Borgida et al., 2016b] and other areas connected with ontological reasoning and knowledge representation. Subsequently, with their coauthors, they were awarded the *2018 Bob Wielinga Best Paper Award* for the [Ma et al., 2018] paper furthering the use of referring expressions in conceptual modelling.

The authors have recently presented tutorials on the topic of referring expressions in knowledge representation and information systems:

1. **Referring Expressions in Ontologies and Query Answering** at the 10th International Conference on Formal Ontology in Information Systems, FOIS 2018 (in Cape Town, South Africa, September 2018).
2. **Managing and Communicating Object Identities in Knowledge Representation and Information Systems** at the 31st Australasian Joint Conference on Artificial Intelligence AI 2018 (in Wellington, New Zealand, December 2018).
3. **Referring Expressions in Knowledge Representation Systems** (T3) at 28th International Joint Conference on Artificial Intelligence, IJCAI 2019 (in Macao, SAR China, August 2019).
4. **Referring Expressions in Knowledge Representation Systems** at the 16th Pacific Rim International Conference on Artificial Intelligence, PRICAI 2019 (in Cuvu, Fiji, August 2019).

5. **Referring Expressions in Knowledge Representation Systems** at the 24th European Conference on Artificial Intelligence, ECAI 2020 (Santiago de Compostela, Spain, on-line).
6. **Referring Expressions in Artificial Intelligence and Knowledge Representation Systems** at the 8th Federated Logic Conference, FLoC 2022 (Haifa, Israel, August 2022).

The tutorial proposed for ECAI 2024 differs from the above editions in its focus on applications of the technology based on referring expressions rather than on issues connected with logical foundations of Knowledge Representation (which was the major theme of the tutorial so far).

About the Tutorial Presenters

Presenters' Affiliation

Name:	David Toman and Grant Weddell
Affiliation:	Cheriton School of Computer Science, University of Waterloo
Postal address:	200 University Ave W., Waterloo, ON N2L3G1, Canada
E-mail address:	{david,gweddell}@uwaterloo.ca
Home page:	https://cs.uwaterloo.ca/~{david,gweddell}

Presenters' Short Bio

Dr. David Toman and Dr. Grant Weddell are professors of Computer Science at the University of Waterloo, Canada. They have published and presented results in the area of knowledge representation over the last 20 years at premier AI conferences (including another *Ray Reiter Prize* in 2010 [Kontchakov et al., 2010] and *Best Paper Prize* and ISWC 2013 [Lutz et al., 2013]); Dr. Toman has also given tutorials in the area of temporal representation and reasoning and temporal databases and information systems that has led to an invited chapter in the *Handbook of Temporal Reasoning in Artificial Intelligence*.

Auxiliary files (e.g., photos of presenters) at https://cs.uwaterloo.ca/~david/ecai24/.

References

[Borgida et al., 2016a] Borgida, A., Toman, D., and Weddell, G. (2016a). On referring expressions in query answering over first order knowledge bases. In *Principles of Knowledge Representation and Reasoning, Cape Town, SA*, pages 319–328.

[Borgida et al., 2016b] Borgida, A., Toman, D., and Weddell, G. E. (2016b). On referring expressions in information systems derived from conceptual modelling. In *Conceptual Modeling - 35th International Conference, ER 2016, Gifu, Japan*, pages 183–197.

[Dale and Haddock, 1991] Dale, R. and Haddock, N. J. (1991). Generating referring expressions involving relations. In Kunze, J. and Reimann, D., editors, *EACL 1991, 5th Conference of the European Chapter of the Association for Computational Linguistics, April 9-11, 1991, Congress Hall, Alexanderplatz, Berlin, Germany*, pages 161–166.

[Dale and Reiter, 1995] Dale, R. and Reiter, E. (1995). Computational interpretations of the gricean maxims in the generation of referring expressions. *Cogn. Sci.*, 19(2):233–263.

[Kontchakov et al., 2010] Kontchakov, R., Lutz, C., Toman, D., Wolter, F., and Zakharyaschev, M. (2010). The combined approach to query answering in DL-Lite. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference, KR 2010, Toronto, Ontario, Canada, May 9-13, 2010*.

[Lutz et al., 2013] Lutz, C., Seylan, I., Toman, D., and Wolter, F. (2013). The combined approach to OBDA: taming role hierarchies using filters. In *The Semantic Web - ISWC 2013 - 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part I*, pages 314–330.

[Ma et al., 2018] Ma, W., Keet, C. M., Oldford, W., Toman, D., and Weddell, G. E. (2018). The utility of the abstract relational model and attribute paths in SQL. In *Knowledge Engineering and Knowledge Management - 21st International Conference, EKAW 2018, Nancy, France, November 12-16, 2018, Proceedings*, pages 195–211.

[Reiter and Dale, 1992] Reiter, E. and Dale, R. (1992). A fast algorithm for the generation of referring expressions. In *14th International Conference on Computational Linguistics, COLING 1992, Nantes, France, August 23-28, 1992*, pages 232–238.

[Russell, 1905] Russell, B. (1905). On denoting. *Mind*, 14(56):479–493.

[Strawson, 1950] Strawson, P. F. (1950). On referring. *Mind*, 59(235):320–344.

[Strawson, 1956] Strawson, P. F. (1956). Singular terms, ontology and identity. *Mind*, 65(260):433–454.