Referring Expressions in Ontologies and Query Answering (a tutorial)

David TOMAN [1] and Grant WEDDELL
Cheriton School of Computer Science, University of Waterloo, Canada

Abstract. How individuals are identified when cooperating agents need to communicate is an inherent issue faced by the designers of information systems. Solutions to this problem range from insisting on global often opaque identifiers, such as URIs, to application specific ways of externally identifying individuals, such as primary keys in relational systems. The goal of this tutorial is to introduce a flexible framework based on *referring expressions* that unifies approaches that address these issues.

Keywords. Knowledge representation, object identification in information systems, referring expressions.

1. Introduction and Goals

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 a knowledge base $\mathcal{K}$, commonly captured using a first order theory, *constant symbols* occurring in $\mathcal{K}$ are the artifacts usually used to identify a subset of the objects for which $\mathcal{K}$ captures knowledge.

In this tutorial, we explore how the class of objects that can be usefully identified can be extended by allowing a variety of more general formulas in the language of $\mathcal{K}$, called *singular referring expression*s, to replace constants as syntactic identifiers of such objects. In particular, we lay a foundation for admitting singular referring expressions in certain answer computation for queries over $\mathcal{K}$. An integral part of this foundation are characterization theorems for identification properties of singular referring expressions for queries annotated with a domain specific language for *referring expression types* [4]. We apply this framework in the context of tractable knowledge representation languages based on *description logics* that are fragments of first order logic, showing in these cases how identification properties can be determined at compile-time for conjunctive queries, and how off-the-shelf conjunctive query answering approaches for such logics, such as perfect rewriting, can be used in query evaluation [4]. We also show that the extension does not negatively impact the computational properties of the underlying logic.

We then apply the work on referring expression types to the issue of identification in conceptual modelling. In particular, we consider how such types yield a separation of concerns in a setting where an information system based on an ontology or concep-

[1] Corresponding Author: D. R. Cheriton School of Computer Science, University of Waterloo, 200 University Ave W., Waterloo, Ontario N2L3G1, Canada; e-mail: david@uwaterloo.ca.

tual schema is to be mapped to a relational schema that is then queried using relational queries. We start 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. We then allow the analyst to attach referring expression types to classes, and to specify appropriate uniqueness constraints to satisfy the requirements on referring expressions. Finally, we show: (a) how a number of well-formedness conditions concerning an assignment of referring expressions can be efficiently diagnosed, and (b) how the above types attached to classes allow an automatic synthesis of a concrete relational schema, and relational queries over that schema, from a combination of the conceptual schema, a choice of referring expression types, and abstract "logical" queries over the conceptual schema [5].

We conclude by considering variations on referring expressions in situations in which singularity can be relaxed, and therefore extend the applicability of the framework to weaker knowledge representation languages.

2. Outline of the Tutorial

The tutorial focuses on foundational issues relating to object identification in ontologies and information systems based on ontologies, and on how such issues can be comprehensively addressed. The approach to object identification discussed in the tutorial naturally and seamlessly complements standard approaches in ontology design.

What is a referring expression? We start with an introduction and overview of how *well formed formulae* that satisfy certain properties, in particular *singularity*—the property stating that the expression describes a single individual in every model of a knowledge base—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 then discuss how referring expressions can be computed, in particular when $\mathcal{K}$ conforms to a decidable fragment of first order logic, and more generic characterizations of varieties of referring expressions in the conceptual modelling of information systems. 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 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 $\mathcal{K}$, 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. How these types are used to adorn free variables in queries over $\mathcal{K}$ is presented. This includes an overview of cases in which $\mathcal{K}$ conforms to a description logic and how computation issues are then resolved by off-the-shelf technology for the logic.

Referring expressions in conceptual modelling. Next, we explore the benefits of adopting referring expression types for use in information systems derived from concep-

tual modelling. In particular, we show how this approach can separate the purely conceptual ontology design from issues connected with how objects are identified within an eventual information system based on the design. Results on how this enables a more wide spectrum use of SQL in defining information retrieval requirements is also presented.

Referring expressions for light knowledge representation formalisms. Finally, we discuss variations on the *singularity* requirement for referring expressions: we show that in certain situations, such as in the case of *certain answers*, the *singularity* condition can be weakened to require singularity of individual certain answers rather than singularity in *every* model of the knowledge base $\mathcal{K}$. This allows us to extend our results to knowledge bases formulated in weaker formalisms, such as the various light-weight Description Logics, e.g., those that lack the capability of enforcing functionality of binary predicates/roles.

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.

The tutorial also links the proposed framework to existing approaches to managing identities of objects in information systems, in particular of those based on knowledge bases or other logic-backed systems, e.g., [1,2,3,6,7,8,9,10,11].

References

[1] Carlos Areces, Santiago Figueira, and Daniel Gorín. Using logic in the generation of referring expressions. In *Logical Aspects of Computational Linguistics - 6th International Conference, LACL 2011*, pages 17–32, 2011.

[2] Carlos Areces, Alexander Koller, and Kristina Striegnitz. Referring expressions as formulas of description logic. In *INLG 2008 - Proceedings of the Fifth International Natural Language Generation Conference*, 2008.

[3] Alexander Borgida. Description logics in data management. *Knowledge and Data Engineering, IEEE Transactions on*, 7(5):671–682, 1995.

[4] Alexander Borgida, David Toman, and Grant Weddell. On referring expressions in query answering over first order knowledge bases. In *Principles of Knowledge Representation and Reasoning, Cape Town, SA*, pages 319–328, 2016.

[5] Alexander Borgida, David Toman, and Grant E. Weddell. On referring expressions in information systems derived from conceptual modelling. In *Conceptual Modeling - 35th International Conference, ER 2016, Gifu, Japan*, pages 183–197, 2016.

[6] Giancarlo Guizzardi, Gerd Wagner, Nicola Guarino, and Marten van Sinderen. An ontologically well-founded profile for UML conceptual models. In *Advanced Information Systems Engineering, CAiSE 2004*, pages 112–126, 2004.

[7] Terry A. Halpin. Modeling of linguistic reference schemes. *IJISMD*, 6(4):1–23, 2015.

[8] Tomasz Imielinski. Intelligent query answering in rule based systems. *The Journal of Logic Programming*, 4(3):229–257, 1987.

[9] Amihai Motro. Intensional answers to database queries. *Knowledge and Data Engineering, IEEE Transactions on*, 6(3):444–454, 1994.

[10] Kees van Deemter, Albert Gatt, Ielka van der Sluis, and Richard Power. Generation of referring expressions: Assessing the incremental algorithm. *Cognitive Science*, 36(5):799–836, 2012.

[11] Carlo Zaniolo. The database language GEM. In *ACM SIGMOD Int. Conf. on Management of Data*, pages 207–218, 1983.

About the Tutorial Authors

Dr. David Toman and Dr. Grant Weddell are professors of Computer Science at the University of Waterloo. Together with Alexander Borgida (Rutgers), they have introduced referring expressions in the area of Ontology-based data access (OBDA) [4] and received the *Ray Reiter Best Paper prize* at KR 2016 for this work. They subsequently extended this work to the area of conceptual modelling [5] and other areas connected with ontological reasoning and knowledge representation. They have published and presented results in the area of knowledge representation over the last 20 years at premier AI conferences (including another Reiter prize in 2010); Dr. Toman has also given tutorials in the area of temporal representation and reasoning that has led to an invited chapter in the *Handbook of Temporal Reasoning in Artificial Intelligence*.