

Presenting Ontologies in Natural Language

Chris Mellish, Department of Computing Science

The standard tool for articulating the shared assumptions of a community for knowledge representation in a domain is the *ontology*. An ontology is a list of standard terms to be used (usually based on words or phrases in some natural language), together with logical axioms about their intended interpretation (e.g. two terms might name classes, one of which is a subclass of the other or which are disjoint). Ordinary users of e-services will need to be able to understand the ontologies they are using whilst being insulated from the details of the logic sitting in the background. This will be especially necessary when a new or casual user wishes to access knowledge using terms in the relevant ontology. Large ontologies are typically built by multiple authors over a period of years, and can be very inaccessible.

Existing ontology tools concentrate on graphical access for expert users. Get many communities of expertise will only be able to survive in the semantic grid if they can effectively present their own ontologies to ordinary people.

This project seeks to develop a generic approach for building *natural language* presentations of parts of ontologies. It will build a prototype system that can answer several types of question involving presentations of parts of ontologies, assess the impact of natural language presentability on principles of ontology design and define a flexible system for attaching linguistic information to an ontology to support presentability. The project is supported by EPSRC grant GR/S62932.

The Basic Approach

The designer of an ontology has chosen one of many possible logically equivalent ways to axiomatise their domain, and this is important information. Therefore our approach works from the axioms themselves (possibly slightly normalised) rather than the set of logical consequences.

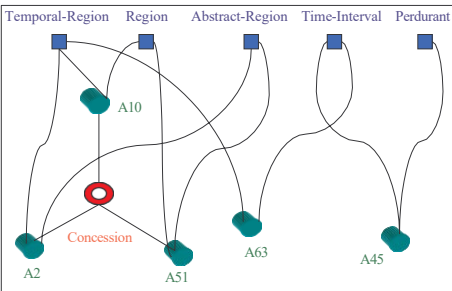
The axioms can be seen as forming a graph, where each axiom is connected to the concepts it mentions (and where there may also be other links for relations between axioms) - see figure. In this graph, routes between axioms correspond to different possible transitions in a coherent text - a text proceeds from one sentence to another by exploiting shared entities or by virtue of a rhetorical relation between the sentences.

Planning a text then involves two stages: selecting a subgraph from the graph of all possible axioms, and then ordering the result to produce a linear structure that maximises coherence. The planned text is then further processed to decide on aggregation/sentence and paragraph boundaries, generate referring expressions, choose among possible words and syntactic structures etc.

What is a Temporal-Region?

Axioms include:

- 1F: SubClassOf(Temporal-Region Region)
- 2: DisjointClasses(Abstract-Region Temporal-Region)
- 63: SubClassOf(Time-Interval Temporal-Region)
- 45: SubClassOf(Perdurant restriction(HAPPEN-AT allValuesFrom(Time-Interval)))
- 51: SubClassOf(Perdurant restriction(HAPPEN-AT allValuesFrom(Time-Interval)))



One Possible Text

Temporal-Region is a kind of Region. An Abstract-Region is also a kind of Region but nothing is both a Temporal-Region and an Abstract-Region. One kind of Temporal-Region is a Time Interval. A Perdurant can happen at a Time-Interval.

Issues

- Mapping from concept names to words
- Describing constraints on roles
- Choosing which way to say an axiom
- Answering other questions, e.g. "What is the difference between Endurant and Perdurant?"
- Preferring to use concepts the user understands
- Characterising sensible relations between axioms (e.g. when could there be a "concession"?)

Supporting the User of E-Services

The vision of e-Science, e-Commerce and e-Government, as well as requiring important technological infrastructure for its realisation, necessitates a move from the collaborative construction and exchange of *data* to the collaborative construction and exchange of *knowledge*. The generation, exchange and maintenance of knowledge is essential for scientific progress, as well as for achieving excellence in the modern business environment and for ensuring that government is socially-inclusive and efficient. From these requirements follows the natural conclusion that the communication and computational infrastructure developed for e-Science, e-Commerce and e-Government will be a *Semantic Grid*.

The move from existing data-based communication systems to a semantic grid will, however, not be easy. The number of users of the services provided, and the variation in their backgrounds, will be very large (especially for e-Government, though even e-Science will need to remain open to potential contributions from and use by initially unanticipated sectors of the international community). To gain from what is offered, and to make their own contributions, these users will have to understand (and contribute to) the conventions and assumptions underlying the representation of the knowledge they wish to interact with.

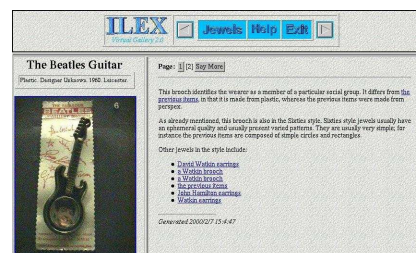
Natural language processing technology can play a key role in making this possible, because:

- users require no special training to understand their native languages (unlike for a complex graphical interface)
- natural languages have evolved sophisticated ways of conveying information about e.g. time, quantification, negation and disjunction, which can be hard to present graphically.

The ILEX System

The ILEX project developed, in collaboration with the National Museums of Scotland, a system to produce descriptions of museum exhibits that were sensitive to what the user had previously seen. It had to solve problems of content selection and text structuring which may be very similar to those arising in presenting ontologies using natural language. A key innovation in ILEX was the notion of a *content potential*, a graph representing the possible content items that can be included in a text and the rhetorical connections between them. The above example is a first attempt to reuse this idea for presenting ontology axioms.

The ILEX project was supported by EPSRC grant GR/K53321 to Jon Oberlander and Chris Mellish at the University of Edinburgh.



O'Donnell, M., Mellish, C., Oberlander, J. and Knott, A., 7ILEX: An Architecture for a Dynamic Hypertext System", *Natural Language Engineering* 7(3) pp225-25F, 2FF1.

Related Work

The use of natural language generation to express the definition of concepts in an ontology was pioneered in the GALEN-IN-USE project, which produced descriptions of surgical procedures in several European languages (Wagner *et al* 99). This project found that the natural language descriptions were very valuable in helping a person to spot errors in definitions which were often very complex. However, this work concentrated on supporting serious ontology developers rather than casual users. Thus the emphasis was on the precision, rather than on the readability, of the phrases. The actual concept definitions used in GALEN had a frame-like structure which made it possible to map a definition into a single natural language phrase in a compositional way. For ontologies defined in languages such as OWL, a concept has less a *definition* than a *set of constraints*. Here a natural language description of a concept will have to take the form of a *text* with multiple sentences, the overall structure having to be planned so as to be coherent as a discourse.

Related work in natural language generation has produced descriptions of database structures (McKeown 85) and object-oriented models (Lavoie *et al* 97).

McKeown, K., *Text Generation*, Cambridge University Press, 1985.

Lavoie, G., Rambow, O. and Reiter, E., "Customizable Descriptions of Object-Oriented Models", *Proc of the 5th Applied NLP Conference*, 1997.

Wagner, J., Rogers, J., Gaud, R. and Scherrer, J-R., "Natural Language Generation of Surgical Procedures", *Medical Informatics* 53, 1999.