

GROUNDING ONTOLOGIES WITH SOCIAL PROCESSES AND NATURAL LANGUAGE

Christophe DEBRUYNE

ABSTRACT

The formal semantics of a (computer-based) system quite simply is the correspondence between this system and some real world as perceived by humans. It is usually given by a formal mapping of the symbols in the system's description to objects in that real world, such that relationships and logical statements in the specification language can be assigned a truth-value depending on whether a certain state of affairs among objects exists in the real world. As the real world is not usually directly accessible inside a computer, storing and reasoning about semantics requires the world to be replaced by an agreed specification of a conceptualization, which is often in the shape of a formal (mathematical) construct. A computer-based, shared, agreed formal conceptualization is what is known as an ontology.

The creation of ontologies for enabling semantic interoperability between two or more autonomously developed and maintained information systems is far from trivial, as it requires the necessary agreements by all representatives as stakeholders – which is called a community – on concepts and relations to make this interoperation possible.

As those agreements are the result of social interactions, appropriate ontology engineering methods should take into account the natural language used by the community during those interactions. In this thesis, a fact-oriented formalism is extended for the construction of so-called hybrid ontologies. In hybrid ontologies, all concepts, terms, etc. are represented not just on their own formal structures (e.g. by means of fact-orientation), but are always to be interpreted in a given context, which is the community that agrees on those formal structures. Agreements are made possible and are supported by glosses in natural language of which the shared understanding is implicit. In hybrid ontology engineering, communities consequently are promoted to “first-class citizens” by formalizing the social interactions that evolve the hybrid ontologies and by declaring the community as the context in which all processes take place.

Next to proposing a framework for hybrid ontologies, this thesis also presents a collaborative ontology engineering method called GOSPL. GOSPL – which stands for Grounding Ontologies with Social Processes and natural Language – allows prescribing how the social interactions should be orchestrated and furthermore describes how agreements on formal and informal descriptions are complementary, and how they interplay.

A part of this thesis is devoted to show how glosses can drive the hybrid ontology construction process and the annotation of aforementioned information systems are used to steer the social interactions and agreement processes.

All of the ideas presented in this work have been implemented in a tool. Both method and tool were used in two experiments, of which a discussion is presented in this thesis.

SAMENVATTING

De formele semantiek van een (computer-gebaseerd) systeem kan het best omschreven worden als de correspondentie tussen dat systeem en een realiteit zoals deze door mensen wordt waargenomen. Deze correspondentie wordt gewoonlijk gegeven door een formele afbeelding van symbolen in de beschrijving van het systeem naar objecten in die realiteit. Zo wordt aan relaties en logische uitdrukkingen in de specificatietaal een waarheidswaarde toegekend, die afhangt van de stand van zaken tussen de objecten in die realiteit. Omdat de echte wereld gewoonlijk niet direct toegankelijk is vanuit een computer, vergt het opslaan en redeneren over de semantiek een abstracte benadering, m.a.w. het vervangen van de realiteit door een overeengekomen specificatie van een conceptualisatie, vaak in de vorm van een formeel (wiskundig) construct. Een computer-gebaseerde, gedeelde en overeengekomen formele conceptualisatie is wat gekend staat als een ontologie.

De creatie van ontologieën die semantische interoperatie tussen twee of meer autonoom ontwikkelde en beheerde informatiesystemen moet bewerkstelligen, is ver van triviaal. Zij vereist de nodige akkoorden over de concepten en relaties tussen alle belanghebbenden – wat we een community noemen – om die interoperatie mogelijk te maken.

Omdat die akkoorden het resultaat zijn van sociale interacties, moeten geschikte methoden voor het bouwen van ontologieën de natuurlijke taal in acht nemen die door de community tijdens die interacties wordt gebruikt. In deze thesis wordt een feit-georiënteerd formalisme aangewend en uitgebreid voor het maken van zogenaamde hybride ontologieën. In hybride ontologieën worden alle concepten, termen, ... niet alleen voorgesteld door hun formele structuren (bijvoorbeeld door middel van feit-oriëntatie), maar moeten altijd binnen een bepaalde context geïnterpreteerd worden. Deze context is de community die akkoorden over die formele structuren bereikt. Die akkoorden worden mogelijk gemaakt door en zijn ondersteund door glosses in natuurlijke taal waarvan het gedeelte verstaan impliciet is. In de constructie van hybride ontologieën worden communities bijgevolg gepromoveerd tot “first-class citizens”, enerzijds door de sociale interacties te formaliseren die de hybride ontologieën doen evolueren en anderzijds door het declareren van de community als de context waarin alle akkoorden plaats vinden.

Naast het voorstellen van een raamwerk voor hybride ontologieën, zal deze thesis ook de collaboratieve ontologie-constructiemethode GOSPL voorstellen. GOSPL staat voor “Grounding Ontologies with Social Processes and natural Language” en voorziet een kader waarin de verschillende sociale interacties kunnen worden voorgeschreven. Verder beschrijft GOSPL hoe akkoorden over formele en informele beschrijvingen complementair zijn en op elkaar inspelen.

Een deel van deze thesis is ook gewijd aan het tonen hoe glosses het hybride ontologie-constructie proces aandrijven en hoe de annotaties van bovengenoemde informatiesystemen worden gebruikt om de sociale interacties en akkoorden bij te sturen.

Alle ideeën voorgesteld in dit werk werden in een tool geïmplementeerd. De methode en tool werden in twee experimenten gebruikt waarop in dit werk dieper zal ingegaan