Taxonomic Partitioning of the Gene Ontology

Wacław Kuśnierczyk
Department of Information and Computer Science
Norwegian University of Science and Technology
email: waku@idi.ntnu.no

Dagstuhl Seminar 07132, March 28, 2007

Presentation Outline

Motivation

Observe the following:

- "The Gene Ontology (GO) project is a collaborative effort to construct and use ontologies to facilitate the biologically meaningful annotation of genes and their products in a wide variety of organisms" [GOC06]
- Some GO terms are universal — they represent molecular functions, biological processes, or cellular components ('features') found in (all?) organisms of all taxa; others are not.
- The sense of 'sensu'.
- If GO terms are appropriately linked to the corresponding terms in the Taxonomy of Species (TS), the GO can be partitioned on demand, based on various taxonomic criteria.

Goals

What might be useful:

1. To investigate the patterns of relations between organisms and their features.
2. To logically formalize these patterns.
3. To investigate the possibility of automated reasoning involving these patterns.

Organism-Feature Relations

There are a few organism-feature relation patterns:

1. The all-some pattern.
   - *Cells are structures present in all vertebrates — all vertebrates have some cells as parts.*

Organism-Feature Relations

There are a few organism-feature relation patterns:

1. The all-some pattern.
   - *Cells are structures present in all vertebrates — all vertebrates have some cells as parts.*
2. The some-some pattern.
   - *Wings are structures present in some vertebrates — some vertebrates have some wings as parts.*

Organism-Feature Relations (contd)

There are a few organism-feature relation patterns:

3. The recursive some-some pattern.
   - *Vertebrae are structures present in some organisms of every vertebrate species — all vertebrate species include some organisms that have some vertebrae as parts.*

Organism-Feature Relations (contd)

There are a few organism-feature relation patterns:

3. The recursive some-some pattern.
   - *Vertebrae are structures present in some organisms of every vertebrate species — all vertebrate species include some organisms that have some vertebrae as parts.*
4. The only-some pattern.
   - *Plant cell walls are structures present only in plants — only plants have some plant cells walls as parts.*

Organism-Feature Relations (contd)

We can also specify negative (complementary) organism-feature relation patterns:

5. The negative all-some pattern.
   - *It is not the case that all formicidae (ants) have some wings as parts — some ants do not have wings (though some ants of all ant species do have wings).*

Organism-Feature Relations (contd)

We can also specify negative (complementary) organism-feature relation patterns:

5. The negative all-some pattern.
   - *It is not the case that all formicidae (ants) have some wings as parts — some ants do not have wings (though some ants of all ant species do have wings).*
6. The negative some-some pattern.
   - *It is not the case that some vertebrates have plant cell walls as parts — no vertebrate has a plant cell wall as a feature.*

Organism-Feature Relations (contd)

We can also specify negative (complementary) organism-feature relation patterns:

7. The negative recursive some-some pattern.
   - *It is not the case that some organisms of all vertebrate species have wings as features — in some vertebrate species there are no wings to be found in any organism.*

Organism-Feature Relations (contd)

We can also specify negative (complementary) organism-feature relation patterns:

7. The negative recursive some-some pattern.
   - *It is not the case that some organisms of all vertebrate species have wings as features — in some vertebrate species there are no wings to be found in any organism.*
8. The negative only-some pattern.
   - *It is not the case that cell walls are found only in plants — fungi and bacteria do have cell walls as well (although not plant cell walls).*

Logical Formalization

We introduce a simple representation language, $\mathscr{L}_{\mathrm{OF}}$.

Syntax:

- Φ_{AS}-sentences: (all-some t g)
- Φ_{SS}-sentences: (some-some t g)
- Φ_{RSS}-sentences: (recursive-some-some t g)
- Φ_{OS}-sentences: (only-some t g)

where $\mathrm{t} \in V_{\mathrm{T}}$, $\mathrm{g} \in V_{\mathrm{G}}$, and $V_{\mathrm{T}}, V_{\mathrm{G}}$ are sets of taxon names (TS terms) and feature names (GO terms), respectively.

Logical Formalization (contd)

We introduce a simple language, $\mathscr{L}_{OF}$

Semantics:

- $\mathscr{I} = \langle U, I, \triangleright \rangle$, an interpretation.
- U, the universe of $\mathscr{I}$, with $U = U_O \cup U_o \cup U_F \cup U_f$, and U_O and U_o sets of classes of organisms and of individual organisms, respectively, and U_F, and U_f sets of classes of features and of individual features, respectively.

Logical Formalization (contd)

We introduce a simple language, $\mathscr{L}_{OF}$

Semantics:

- $\mathscr{I} = \langle U, I, \triangleright \rangle$, an interpretation.
- U, the universe of $\mathscr{I}$, with $U = U_O \cup U_o \cup U_F \cup U_f$, and U_O and U_o sets of classes of organisms and of individual organisms, respectively, and U_F, and U_f sets of classes of features and of individual features, respectively.
- I, a mapping from the vocabulary $V = V_T \cup V_G$ into the universe U, with $I : V_T \to U_O$, $I : V_G \to U_F$.

Logical Formalization (contd)

We introduce a simple language, $\mathscr{L}_{OF}$

Semantics:

- $\mathscr{I} = \langle U, I, \triangleright \rangle$, an interpretation.
- U, the universe of $\mathscr{I}$, with $U = U_O \cup U_o \cup U_F \cup U_f$, and U_O and U_o sets of classes of organisms and of individual organisms, respectively, and U_F, and U_f sets of classes of features and of individual features, respectively.
- I, a mapping from the vocabulary $V = V_T \cup V_G$ into the universe U, with $I : V_T \rightarrow U_O$, $I : V_G \rightarrow U_F$.
- $\triangleright$, a (vaguely specified) relation that holds between an organism and a feature ('has as a feature', as in "*vertebrates have vertebrae as features*" or "*mammals have suckling as a feature*").

Logical Formalization (contd)

We introduce a simple language, $\mathscr{L}_{OF}$

Semantics:

- $\mathscr{I} = \langle U, I, \triangleright \rangle$, an interpretation.
- U, the universe of $\mathscr{I}$, with $U = U_O \cup U_o \cup U_F \cup U_f$, and U_O and U_o sets of classes of organisms and of individual organisms, respectively, and U_F, and U_f sets of classes of features and of individual features, respectively.
- I, a mapping from the vocabulary $V = V_T \cup V_G$ into the universe U, with $I : V_T \rightarrow U_O$, $I : V_G \rightarrow U_F$.
- $\triangleright$, a (vaguely specified) relation that holds between an organism and a feature ('has as a feature', as in "*vertebrates have vertebrae as features*" or "*mammals have suckling as a feature*").
- T and F, the usual Boolean logical truth values.

All-Some Semantics

Definition (Semantics of Φ_{AS}-sentences)

Let σ be a $\mathscr{L}_{OF}$ sentence of the form (all-some t g), with $t \in V_T$ and $g \in V_G$. Then:

$$\sigma^{\mathscr{I}} = \begin{cases} \text{T}, & \text{if } \forall o \in I(t)\ \exists f \in I(g) : o \triangleright f, \\ \text{F}, & \text{otherwise.} \end{cases}$$

where o ranges over individual organisms in U_o, f ranges over individual features in U_f.

All-Some Semantics

Definition (Semantics of Φ_{AS}-sentences)

Let σ be a $\mathscr{L}_{OF}$ sentence of the form (all-some t g), with $t \in V_T$ and $g \in V_G$. Then:

$$\sigma^{\mathscr{I}} = \begin{cases} \text{T}, & \text{if } \forall o \in I(t)\ \exists f \in I(g) : o \triangleright f, \\ \text{F}, & \text{otherwise.} \end{cases}$$

where o ranges over individual organisms in U_o, f ranges over individual features in U_f.

Note:

The all-some semantic pattern is the usual (default) in most knowledge representation languages that allow relational assertions about classes.

Some-Some Semantics

Definition (Semantics of Φ_{SS}-sentences)

Let σ be a $\mathscr{L}_{OF}$ sentence of the form (some-some t g), with $t \in V_T$ and $g \in V_G$. Then:

$$\sigma^{\mathscr{I}} = \begin{cases} \text{T}, & \text{if } \exists o \in I(t)\ \exists f \in I(g) : o \triangleright f, \\ \text{F}, & \text{otherwise.} \end{cases}$$

Some-Some Semantics

Definition (Semantics of Φ_{SS}-sentences)

Let σ be a $\mathscr{L}_{OF}$ sentence of the form (some-some t g), with $t \in V_T$ and $g \in V_G$. Then:

$$\sigma^{\mathscr{I}} = \begin{cases} \text{T}, & \text{if } \exists o \in I(t)\ \exists f \in I(g) : o \triangleright f, \\ \text{F}, & \text{otherwise.} \end{cases}$$

Note:

The some-some pattern is based on an existential claim: there are organisms of the class ... such that ...

Recursive Some-Some Semantics

Definition (Semantics of Φ_{RSS}-sentences)

Let σ be a $\mathscr{L}_{OF}$ sentence of the form (recursive-some-some t g), with $t \in V_T$ and $g \in V_G$. Then:

$$\sigma^{\mathscr{I}} = \begin{cases} \text{T}, & \text{if } \forall O \in U_O : (O \subseteq I(t) \ \wedge \ c(O)) \Rightarrow \\ & \qquad\qquad \exists o \in O \ \exists f \in I(g) : o \triangleright f, \\ \text{F}, & \text{otherwise,} \end{cases}$$

where O ranges over classes of organisms in U_O, and c, the condition of recursion, is a boolean function $c : U_O \to \{\text{T}, \text{F}\}$. (Here, $c(O) = \text{T}$ iff O is of the taxonomic rank called 'species' or above).

Only-Some Semantics

Definition (Semantics of Φ_{OS}-sentences)

Let σ be a $\mathscr{L}_{OF}$ sentence of the form (only-some t g), with $t \in V_T$ and $g \in V_G$. Then:

$$\sigma^{\mathscr{I}} = \begin{cases} \text{T}, & \text{if } \forall f \in I(g)\ \forall o \in U_o : o \triangleright f \Rightarrow o \in I(t), \\ \text{F}, & \text{otherwise,} \end{cases}$$

Only-Some Semantics

Definition (Semantics of Φ_{OS}-sentences)

Let σ be a $\mathscr{L}_{OF}$ sentence of the form (only-some t g), with $t \in V_T$ and $g \in V_G$. Then:

$$\sigma^{\mathscr{I}} = \begin{cases} \text{T}, & \text{if } \forall f \in I(g)\ \forall o \in U_o : o \triangleright f \Rightarrow o \in I(t), \\ \text{F}, & \text{otherwise,} \end{cases}$$

Note:

The only-some pattern is roughly the inverse of the all-some pattern:

$$\text{(only-some } t\ g) \approx \text{(all-some } g\ t)$$

with $\triangleright$ replaced by its inverse in the semantics of the right-hand sentence.

Inference Patterns

Monotonic reasoning with $\mathscr{L}_{\mathsf{OF}}$-sentences:

A number of rules may be defined to perform monotonic reasoning in $\mathscr{L}_{\mathsf{OF}}$.

Example (Inference with Φ_{SS}-sentences)

$$\frac{(\text{some-some } t\ g),\ (\text{subclass-of } t\ t')}{(\text{some-some } t'\ g)}\,R_{SS|T}$$

The rule $R_{SS|T}$ says that the relation $\triangleright$ propagates with the some-some semantics upwards along the hierarchy of the Taxonomy of Species.

Inference Patterns (contd)

Summary of monotonic inference patterns

	Taxonomy of Species	Gene Ontology
AS	$\downarrow$	$\uparrow$
SS	$\uparrow$	$\uparrow$
RSS	$\downarrow$	$\uparrow$
OS	$\uparrow$	$\downarrow$
NAS	$\uparrow$	$\downarrow$
NSS	$\downarrow$	$\downarrow$
NRSS	$\uparrow$	$\downarrow$
NOS	$\downarrow$	$\uparrow$

Problems

The framework is incoherent:

- Sentences of the forms Φ_{SS} and Φ_{RSS} make unconditioned existential claims: "*there exists* such an instance such that ...".
- Sentences of the forms Φ_{AS} and Φ_{OS} do not make such claims.

Problems (contd)

Alternative solutions:

- Avoid existential claims in Φ_{SS} and Φ_{RSS} by conditioning on the existence: "*If there exists an instance, then there exists an instance such that . . .*".

Problems (contd)

Alternative solutions:

- Avoid existential claims in Φ_{SS} and Φ_{RSS} by conditioning on the existence: "*If there exists an instance, then there exists an instance such that . . .*".
- Augment Φ_{AS} and Φ_{OS} to make existential claims: "*There exists an instance, and every instance is such that . . .*".

Questions

The question of existential claims:

If an existential claim is made, what does it mean, precisely?

- That at all times there existed and will exist an organism such that . . . ?

Questions

The question of existential claims:

If an existential claim is made, what does it mean, precisely?

- That at all times there existed and will exist an organism such that . . . ?
- That at any time during some particular interval (our century? this year? today?) there exists such an organism?

Questions

The question of existential claims:

If an existential claim is made, what does it mean, precisely?

- That at all times there existed and will exist an organism such that . . . ?
- That at any time during some particular interval (our century? this year? today?) there exists such an organism?
- That such an organism existed at the moment when the observation was made?

Questions

The question of existential claims:

If an existential claim is made, what does it mean, precisely?

- That at all times there existed and will exist an organism such that . . . ?
- That at any time during some particular interval (our century? this year? today?) there exists such an organism?
- That such an organism existed at the moment when the observation was made?
- That such an organism existed at the moment when the statement was made?

Questions

The question of existential claims:

If an existential claim is made, what does it mean, precisely?

- That at all times there existed and will exist an organism such that . . . ?
- That at any time during some particular interval (our century? this year? today?) there exists such an organism?
- That such an organism existed at the moment when the observation was made?
- That such an organism existed at the moment when the statement was made?
- That such an organism exists at any time the statement is read?

Questions (contd)

The question of existential claims:

Do not ontologies intended to describe the real world rather than a synthetic model-theoretic 'reality' make existential claims implicitly?

Questions (contd)

The question of existential claims:

Do not ontologies intended to describe the real world rather than a synthetic model-theoretic 'reality' make existential claims implicitly?

- If they do, the problem is solved: make such claims explicit.

Questions (contd)

The question of existential claims:

Do not ontologies intended to describe the real world rather than a synthetic model-theoretic 'reality' make existential claims implicitly?

- If they do, the problem is solved: make such claims explicit.
- If they do not, an ontology describing, e.g., unicorns (i.e., an ontology specifying what something would have to be like to be a unicorn) is trivially correct in its description of the reality.

Acknowledgments

Thanks to:
Agnar Aamodt, Jöerg Cassens, Melissa Haendel, Chris Mungall, Barry Smith, David Sutherland, and the anonymous reviewers of a related publication — for comments and suggestions.

Bibliography


The Gene Ontology Consortium.
The Gene Ontology (GO) project in 2006.
Nucleic Acids Research, 34:D322–D326, 2006.

Thank you.