

# Exemplo para utilizar com funções

Filipe Santana

21 de dezembro de 2015

## 1 Exemplificação

O exemplo que podemos criar para a conversão de consultas pode ser retirado da planilha (final-data), no caminho (.../GitHub/ontos/Examples/Example - Axioma Função).

Na planilha owl\_elements há uma estrutura de substituição em que variáveis nomeadas e escritas entre símbolos \$ ... \$ são substituídas pelo conteúdo das linhas da tabela final\_data. Se você tiver conhecimento sobre VBA macro, pode recuperar a ideia do macro e distribuir na ferramenta. A tarefa realizada pelo macro é a de substituir.

O conteúdo da tabela é convertido e é gerado um arquivo chamado integrativo\_new.owl.

A tabela é construída no formato da tabela sobre o Uniprot (tabela 1) e Ensembl (tabela 2) estão somados, e resumido em uma estrutura genérica na tabela 3.

Tabela 1: Uniprot Table example.

Entry	Protein	Organism	GO (bp)	GO (mf)	GO (cc)	Ensembl ID
F1MEW4	CBS	<i>Bos taurus</i>	blood ves- sel remo- deling; ...	cystathionine $\beta$ -synthase activity ...	cytoplasm ...	ENSBTAT00000000184 ...
A6H5Y3	MS	<i>Mus musculus</i>	methionine biosynthe- tic pro- cess; ...	cobalamin binding; ...	cytoplasm ...	ENSMUST000000099856

GO (bp) , GO (mf) and GO (cc) represents rows from UniProt that include annotations for GO classes '*Biological process*', '*Molecular function*' and '*Cellular component*' respectively.

Os axiomas que geram a conversão são escritos no formato das letras das tabelas e estão escritos nas tabelas 4,5, 6, 7, 8, 9, 10.

Tabela 4: Defined Subclasses of proteins  $P$ .

$P$ subclassOf pr: <i>Protein</i>
$P\_sensu\_O$ subclassOf $P$

Tabela 2: Ensembl Table example.

Entry	Description	Gene	Organism	Phenotype
ENST00000376583	methylenereductase	t. MTHFR	<i>Homo sapiens</i>	Methotrexate poisoning; Homocysteinemia; ...
ENSBTAT00000000184	homocystathionine $\beta$ -synthase (CBS)	CBS	<i>Bos taurus</i>	No phenotype associated

Tabela 3: Template table.

#	$P$	$O$	$Bp$	$Mf$	$C$	$Ph$	$M$
$k$	$P_k$	$O_k$	$Bp_1 \dots n$	$Mf_1 \dots n$	$C_1 \dots n$	$Ph_1 \dots n$	$M_1 \dots n$
...	...	...	...	...	...	...	...
$l$	$P_l$	$O_l$	$Bp_1 \dots n$	$Mf_1 \dots n$	$C_1 \dots n$	$Ph_1 \dots n$	$M_1 \dots n$
...	...	...	...	...	...	...	...
$m$	$P_m$	$O_m$	$Bp_1 \dots n$	$Mf_1 \dots n$	$C_1 \dots n$	$Ph_1 \dots n$	$M_1 \dots n$

The symbol # represents record IDs;  $P$  proteins;  $G$  genes;  $O$  organisms;  $Bp$  biological processes;  $Mf$  molecular function;  $C$  cellular component;  $Ph$  phenotype; and,  $M$  the associate molecules.

Tabela 5: Defined Subclasses of biological process  $Bp$ .

$Bp$ subclassOf go:'biological_process'
$Bp\_in\_O\_with\_P\_and\_M$ subclassOf $Bp$
$Dysfunctional\_Bp\_in\_O\_with\_P\_and\_M$ subclassOf $Bp\_in\_O\_with\_P\_and\_M$

Tabela 6: Cellular component  $C$  union classes.

$C$ subclassOf go:'cell_component'
$C_1\_or\_C_n$ subclassOf go:'cell_component'
$C_1\_or\_C_n$ equivalentTo $(C_1 \text{ or } C_2 \text{ or } C_n)$

Tabela 7: Definitions for biological process subclasses defines for specific proteins, organisms and molecules (*Bp\_in\_O\_with\_P\_and\_M* together with their *Dysfunctional\_Bp\_in\_O\_with\_P\_and\_M*.

---


$$\begin{aligned}
 & \text{Bp\_in\_O\_with\_P\_and\_M equivalentTo Bp} \\
 & \text{and ('has participant' some M)} \\
 & \text{and ('has participant' some (P and} \\
 & \quad \text{('is bearer of' some (bt12:Function and} \\
 & \quad \quad \text{('is realization of' only Mf) ) )} \\
 & \text{and ('is included in' some C}_1\text{\_or\_C}_n\text{))} \\
 & \text{and ('is included in' some O)}
 \end{aligned}$$


---

Tabela 8: Dysfunctional phenotypes of *Bp\_in\_O\_with\_P\_and\_M*.

---


$$\begin{aligned}
 & \text{Dysfunctional\_Bp\_in\_O\_with\_P\_and\_M equivalentTo} \\
 & \quad \text{Bp\_in\_O\_with\_P\_and\_M} \\
 & \quad \text{and ('is bearer of' some 'Dysfunctional Quality')} \\
 & \text{Dysfunctional\_Bp\_in\_O\_with\_P\_and\_M subClassOf} \\
 & \quad \text{Bp\_in\_O\_with\_P\_and\_M} \\
 & \quad \text{and ('is realization of' only (Risk and} \\
 & \quad \quad \text{(causes some Ph)))}
 \end{aligned}$$


---

Tabela 9: Subclasses created for the organism specific protein (*P\_sensu\_O*) classes in database records

---


$$\begin{aligned}
 & \text{P\_sensu\_O equivalentTo P} \\
 & \quad \text{and ('is included in' some O)} \\
 & \text{P\_sensu\_O subClassOf P} \\
 & \quad \text{and ('is bearer of' some (Function and} \\
 & \quad \quad \text{('has realization' only Mf)))}
 \end{aligned}$$


---

Tabela 10: Axioms generated for organisms *O* in database records

---


$$\begin{aligned}
 & \text{O subClassOf bt12:Organism} \\
 & \quad \text{and ('is bearer of' some (Disposition and} \\
 & \quad \quad \text{('has realization' only Bp)))}
 \end{aligned}$$


---

Esses axiomas foram gerados a partir da interpretação das tabelas do UniProt/Ensembl. Nesse meio, foram definidas as seguintes consultas que recuperam os respectivos resultados.

**CQ1: Which kinds of biological process are included in organisms of a specific type A** This query is intended to retrieve all biological process classes that takes place in organisms. To perform this query, we used an enzyme as example and substitute the placeholder *A* by mouse (*Mus musculus*).

Tabela 11: Competency Question #1

CQ#	DL Query
1	' <i>biological_process</i> ' and (' <b>is included in</b> ' some <i>A</i> )

The following classes are retrieved from the ontology (Table 12).

Tabela 12: Result of Competency Question #1

'*amino acid betaine catabolic process in Mus musculus with Betaine homocysteine S methyltransferase 1* a  
'*blood vessel remodeling in Mus musculus with Cystathionine beta-synthase and Homocysteine*';  
'*cartilage development involved in endochondral bone morphogenesis in Mus musculus*  
with Cystathionine beta-synthase and Homocysteine'; and 36 more classes.

These results are expected as they match the content represented in data, without changing any domain or upper domain ontology. Here we see how a query on potentialities can be expressed as a simple DL query. However, the correct interpretation hinges on the assumption that none of these specific subclasses is empty.

**CQ2: Which are the kinds of protein that, while in organisms, are capable of participating in a certain process of type B?** This query is meant to retrieve kinds of proteins capable of performing a specific biological process.

To illustrate, we substitute B by '*Methylation*'.

Tabela 13: Competency Question #2

CQ#	DL Query
2	<i>Protein</i> and ( ' <b>is bearer of</b> ' some ( <i>Function</i> and ( ' <b>is realized by</b> ' only <i>B</i> ) ) )

Using DL query and reasoning, we obtain the following results (Table 14):

Tabela 14: Result of Competency Question #2

'*Betaine homocysteine S-methyltransferase 1 sensu Homo sapiens*';  
'*Cystathionine beta-synthase sensu Homo sapiens*';  
'*Cystathionine gamma lyase sensu Homo sapiens*';  
'*Methionine synthase sensu Homo sapiens*';  
'*Methylenetetrahydrofolate reductase sensu Homo sapiens*'.

These are the proteins described for the *Methylation* process in specific organisms.

Tabela 15: Competency Question #3

CQ#	DL Query
3	' <i>biological_process</i> ' and ( ' <b>has participant</b> ' some ( <i>A</i> and ( ' <b>is bearer of</b> ' some ( <i>Function</i> and ( ' <b>is realization of</b> ' some <i>B</i> ) ) ) ) )

**CQ3: Which are the kinds of biological processes in which a specific protein of type A participates that has the function of performing molecular activities of type B?** This query is related to the identification of biological processes (e.g. reactions) that involve a specific protein, a protein that should be able to performing this reaction. The relevance of this query is related to the capability of retrieving specific biological processes by means of proteins from specific reactions. We substitute A by *Cystathionine\_gama\_lyase* and B by 'carbon-sulfur lyase activity'. The result is in Table 16.

Tabela 16: Result of Competency Question #3

' <i>cellular nitrogen compound metabolic process in Homo sapiens with Cystathionine gamma lyase and Homocysteine</i> ';
' <i>cysteine biosynthetic process in Homo sapiens with Cystathionine gamma lyase and Homocysteine</i> ';
' <i>cysteine metabolic process in Homo sapiens with Cystathionine gamma lyase and Homocysteine</i> ';
and 14 more classes.

Tabela 17: Competency Question #4

CQ#	DL Query
4	' <i>biological_process</i> ' and ( ' <b>realization of</b> ' only ( <i>Risk</i> and ( ' <b>causes</b> some <i>C</i> ) ) )

**CQ4: Which are the kinds of biological processes that entail some risk of causing a specific dysfunctional state C (phenotype)?** This query retrieves biological processes that entail the risk of developing a dysfunctional phenotype. This query is relevant in the sense whether it enables the identification of any abnormal situations regarding an specific process from an organism. We replace C by 'Atherosclerosis' and obtain the results in Table 18.

Tabela 18: Result of Competency Question #4

---

' <i>Dysfunctional homocysteine metabolic process in <u>Rattus norvegicus</u> with</i>
<i>Methylenetetrahydrofolate reductase and Homocysteine</i> ';
' <i>Dysfunctional methionine biosynthetic process in <u>Rattus norvegicus</u> with</i>
<i>Methylenetetrahydrofolate reductase and Homocysteine</i> ';
' <i>Dysfunctional one carbon metabolic process in <u>Rattus norvegicus</u> with</i>
<i>Methylenetetrahydrofolate reductase and Homocysteine</i> ';
and 12 more classes.

---

Tabela 19: Competency Question #5

---

CQ#	DL Query
5	<i>Organism</i> and ( 'is bearer of' some ( <i>Disposition</i> and ( 'is realized by' only <i>A</i> ) ) )

---

**CQ5: Which kinds of organisms are capable of performing a specific biological process of type A?** This query retrieves organisms that are capable of performing specific biological processes. This query is relevant because not all biological processes for organisms are fully described. Even two different organisms that include same proteins under same conditions may not include similar processes.

We used as example 'Cysteine biosynthetic process' to substitute the placeholder A. The results are given in Table 20.

Tabela 20: Result of Competency Question #5

---

' <i>Homo sapiens</i> ' and ' <i>Mus musculus</i> '
---

---

These are the organisms that are able to perform 'Cysteine biosynthetic process'.