

# Οργάνωση Βάσεων Βιοϊατρικών Δεδομένων Εξόρυξη Γνώσης Βιοϊατρικών Δεδομένων

Ακαδημαϊκό Έτος 2016-2017

## Τεχνολογίες Πληροφορικής στην Ιατρική και τη Βιολογία

### Άσκηση 3<sup>η</sup>

---

#### Ανάκτηση μέσω SPARQL.

##### Ερώτημα 1 (30%)

- Ποια είναι η βασική ιδέα πίσω από τον σημασιολογικό ιστό;
- Ποιες είναι οι εννοιολογικές διαφορές μεταξύ των RDF, RDFS και OWL; Δώστε παραδείγματα πραγμάτων που μπορούν (ή δεν μπορούν) να μοντελοποιηθούν από κάθε μία από τις 3 τεχνολογίες.
- Περιγράψτε τι σημαίνουν τα «domain» και «range» στο RDFS, και δώστε ένα παράδειγμα για το καθένα.
- Είναι οι ακόλουθες δηλώσεις αληθείς ή όχι;
  - Μια κυριολεκτική τιμή (literal) μπορεί να είναι κατηγορημα (predicate) σε μια δήλωση (statement) RDF.
  - Δεν υπάρχει τρόπος μη-δυναδικές σχέσεις να μοντελοποιηθούν σε RDF.
  - Το «κάθε σκύλος είναι θηλαστικό» μπορεί να αναπαρασταθεί σε RDFS.
  - Το «`rdfs:subClassOf`» σημαίνει ότι κάθε στιγμιότυπο (instance) της μιας κλάσης είναι ταυτόχρονα και στιγμιότυπο της άλλης.

##### Ερώτημα 2 (70%)

Η βάση γνώσης NeXtProt<sup>1</sup> (<http://www.nextprot.org/>) περιλαμβάνει πληροφορίες σχετικές με τις ανθρώπινες πρωτεΐνες. Η βάση αυτή έχει μοντελοποιηθεί μέσω RDF, ενώ μια απλοποιημένη εκδοχή του μοντέλου εμφανίζεται στην εικόνα 1. Καλείστε να μελετήσετε το σχήμα αυτό, και να εκτελέσετε ανακτήσεις από την βάση αυτή, χρησιμοποιώντας τα εργαλεία που παρέχονται στον ακόλουθο σύνδεσμο:

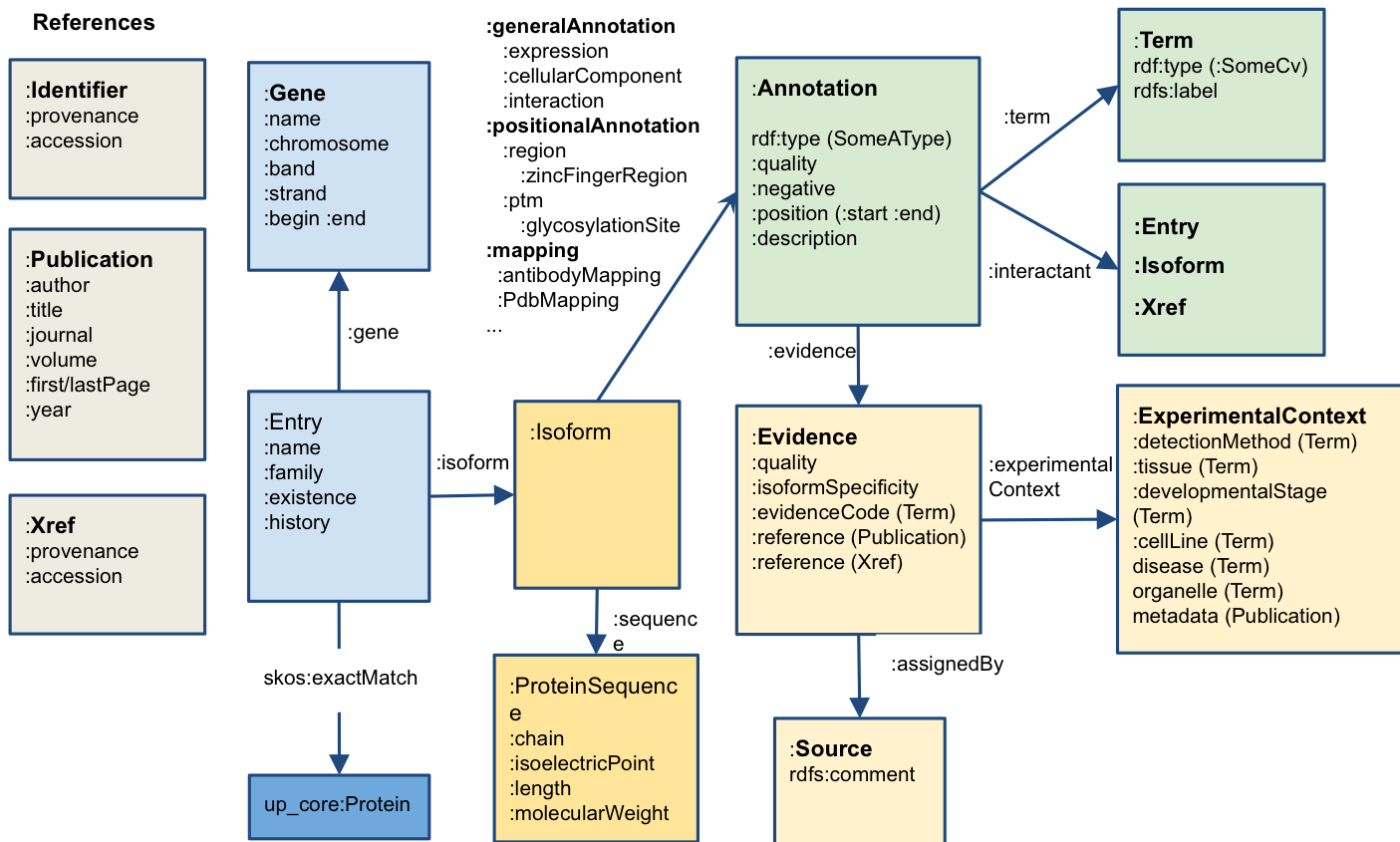
<http://sparql-playground.nextprot.org/>

<http://snorql.nextprot.org/>

- Επιλέξτε το 1<sup>ο</sup> ερώτημα SPARQL που εμφανίζεται στο δεξί μέρος της διεπαφής του παραπάνω συνδέσμου, με τίτλο «Proteins phosphorylated and located in the cytoplasm». Περιγράψτε τι ανακτά το συγκεκριμένο ερώτημα SPARQL. Στην συνέχεια αλλάξτε το, ώστε να επιστρέφει τις γλυκοπρωτεΐνες που βρίσκονται στο κυτταρόπλασμα (glycoproteins located in the cytoplasm). Περιγράψτε το τροποποιημένο ερώτημα SPARQL.
- Επιλέξτε το 99<sup>ο</sup> ερώτημα SPARQL που εμφανίζεται στο δεξί μέρος της διεπαφής του παραπάνω συνδέσμου, με τίτλο «Proteins that have at least one PTM in a position of a variant». Περιγράψτε τι ανακτά το συγκεκριμένο ερώτημα SPARQL. Στην συνέχεια αλλάξτε το, ώστε να επιστρέφει τις πρωτεΐνες που περιέχουν τροποποίηση σχετιζόμενη με λιπίδιο (proteins with a lipidation site having a variant on it). Περιγράψτε το τροποποιημένο ερώτημα SPARQL.

---

<sup>1</sup> <https://en.wikipedia.org/wiki/NeXtProt>



Εικόνα 1: Μια απλοποιημένη εκδοχή της βάσης γνώσης NeXtProt<sup>2</sup>.

Ημερομηνία Παράδοσης: Τρίτη, 30 Μαΐου 2017

Η άσκηση είναι ατομική.

Πρέπει να παραδοθεί στην διεύθυνση ηλ. ταχυδρομείου: [petasis@iit.demokritos.gr](mailto:petasis@iit.demokritos.gr)

<sup>2</sup> <http://sparql-playground.nextprot.org/help/doc/introduction>