ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

«Βελτιώσεις του συστήματος παραγωγής φυσικής γλώσσας NaturalOWL»

Κωνσταντία Μαρκαντώνη (Α.Μ: 3040112)

Επιβλέπων: Ίων Ανδρουτσόπουλος

Αθήνα 2009

Περιεχόμενα

Περιεχόμενα 2
1 Εισαγωγή 3
1.1 Αντικείμενο της εργασίας 3
1.2 Διάρθρωση της εργασίας 3
1.3 Ευχαριστίες 4
2 Βελτιώσεις του συστήματος NaturalOWL 5
2.1 Απαλοιφή περιττών μηνυμάτων στην επιφανειακή μορφή 5
2.2 Αυτόματη εισαγωγή εγγραφών στο λεξικό 7
2.3 Χρήση λίστας κουκκίδων σε ιδιότητα με πολλαπλές τιμές 9
2.4 Χρήση κατάλληλης φράσης στην παραγωγή κειμένων για οντότητες που είναι μοναδικές σε μια τάξη της οντολογίας 12
2.5 Χωρισμός ιδιοτήτων σε ενότητες και ορισμός προτεραιότητας των ενοτήτων 14
2.6 Χειρισμός περιττών αντίστροφων γεγονότων 15
2.7 Δυνατότητα παραγωγής κειμένων με επισημειώσεις RDFa 17
2.8 Αυτόματη εισαγωγή πληροφοριών για νέο τύπο χρηστών 19
2.9 Προβολή λεπτομερειών εγγραφής του λεξικού σε αναδυόμενο παράθυρο 21
2.10 Βελτίωση προβολής των Multilingual Personalized Canned Texts 24
3 Δοκιμές με την οντολογία SWRC 26
4 Συμπεράσματα 30
Αναφορές 31

1 Εισαγωγή

1.1 Αντικείμενο της εργασίας

Στη διάρκεια προηγούμενης διπλωματικής εργασίας [1], αναπτύχθηκε το σύστημα παραγωγής κειμένων φυσικής γλώσσας NaturalOWL, το οποίο παράγει στα ελληνικά και τα αγγλικά εξατομικευμένες περιγραφές οντοτήτων ή τάξεων οντοτήτων μιας οντολογίας OWL [2, 12] Το σύστημα έχει αναπτυχθεί εξολοκλήρου σε Java, βασίζεται σε ιδέες του ευρωπαϊκού ερευνητικού έργου M-PIRO [3] και υποστηρίζεται από το «εργαλείο συγγραφής» του M-PIRO, το οποίο βοηθά χρήστες που δεν είναι ειδικοί της επεξεργασίας φυσικής γλώσσας να προσαρμόσουν το σύστημα σε νέες οντολογίες [10]. Το NaturalOWL υποστηρίζει εγγενώς οντολογίες OWL και έχει ενσωματωθεί στο περιβάλλον ανάπτυξης οντολογιών και βάσεων γνώσεων Protégé [4], το οποίο μπορεί πλέον να χρησιμοποιηθεί αντί του εργαλείου συγγραφής του M-PIRO. Επίσης, μια πιο πρόσφατη διπλωματική εργασία [5] πρόσθεσε στο NaturalOWL βελτιωμένες, σε σχέση με το M-PIRO, δυνατότητες παραγωγής συγκρίσεων με προηγούμενες οντότητες για τις οποίες έχουν παραχθεί κείμενα.

Κύριος σκοπός της παρούσας εργασίας ήταν η βελτίωση του NaturalOWL είτε με την προσθήκη νέων λειτουργιών είτε με τη βελτίωση των υπαρχουσών. Επιπρόσθετα, χρησιμοποιήθηκε η οντολογία SWRC – Semantic Web for Research Communities [6], ώστε να διερευνηθεί η συμπεριφορά του συστήματος κατά την παραγωγή κειμένων από μια νέα οντολογία.

1.2 Διάρθρωση της εργασίας

Στο δεύτερο κεφάλαιο παρουσιάζονται οι βελτιώσεις του NaturalOWL που έγιναν στη διάρκεια της εργασίας. Πιο συγκεκριμένα, εξετάζονται τα προς βελτίωση σημείακαι ο τρόπος με τον οποίο αντιμετωπίστηκε κάθε ένα από αυτά.

Στο τρίτο κεφάλαιο παρουσιάζεται η οντολογία SWRC – Semantic Web for Research Communities και παραδείγματα παραγωγής κειμένων φυσικής γλώσσας μέσω του συστήματος NaturalOWL από αυτήν.

Πριν τα παρακάτω κεφάλαια, ο αναγνώστης θα πρέπει να έχει διαβάσει τις προηγούμενες εργασίες που αναφέρθηκαν στην ενότητα 1.1. Προτείνεται, επίσης, η ανάγνωση του βιβλίου των Reiter και Dale [8].

1.3 Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή της εργασίας κ. Ίωνα Ανδρουτσόπουλο για την συνεχή καθοδήγησή του και τη λεπτομερή ανατροφοδότηση που μου παρείχε σε κάθε βήμα της. Επίσης θα ήθελα να ευχαριστήσω τον υποψήφιο διδάκτορα του τμήματος Δημήτρη Γαλάνη, που με ενημέρωσε για το υπάρχον σύστημα και τις λεπτομέρειες υλοποίησης του, στο οποίο βασίστηκε η εργασία, καθώς και για τη βοήθεια που μου παρείχε καθ' όλη τη διάρκεια της υλοποίησης της εργασίας. Ευχαριστώ, επίσης, τον υποψήφιο διδάκτορα του τμήματος Γεράσιμο Λάμπουρα, που διάβασε και σχολίασε μια προηγούμενη μορφή του κειμένου της εργασίας.

2 Βελτιώσεις του συστήματος NaturalOWL

Σε αυτό το κεφάλαιο παρουσιάζονται αναλυτικά τα σημεία του συστήματος που χρειάζονταν βελτίωση, η συμπεριφορά του συστήματος πριν και μετά τις βελτιώσεις, καθώς και ο τρόπος υλοποίησης των βελτιώσεων.

2.1 Απαλοιφή περιττών μηνυμάτων στην επιφανειακή μορφή

Το σύστημα περιλαμβάνει ένα λεξικό το οποίο περιέχει πληροφορίες σχετικές με τα ουσιαστικά ή τις ονοματικές φράσεις που αντιστοιχούν σε οντότητες ή τάξεις της οντολογίας. Στο αρχικό σύστημα κατά τη δημιουργία μιας οντότητας (instance) μιας τάξης δεν ήταν εφικτό να δημιουργηθεί αυτόματα η εγγραφή της στο λεξικό. Ως εκ τούτου, στα παραγόμενα κείμενα παρεμβαλλόταν ένα μήνυμα λάθους της ακόλουθης μορφής, μέχρι να προστεθεί η αναγκαία εγγραφή στο λεξικό:

"Today it is exhibited in **NOT FOUND FILLER[null]**. *It dates from between 470 and 460 B.C."*

Επίσης, στην περίπτωση που αρχικά είχε δημιουργηθεί η αναγκαία εγγραφή στο λεξικό και στη συνέχεια διαγράφηκε, το παραγόμενο κείμενο περιείχε μήνυμα λάθους της μορφής:

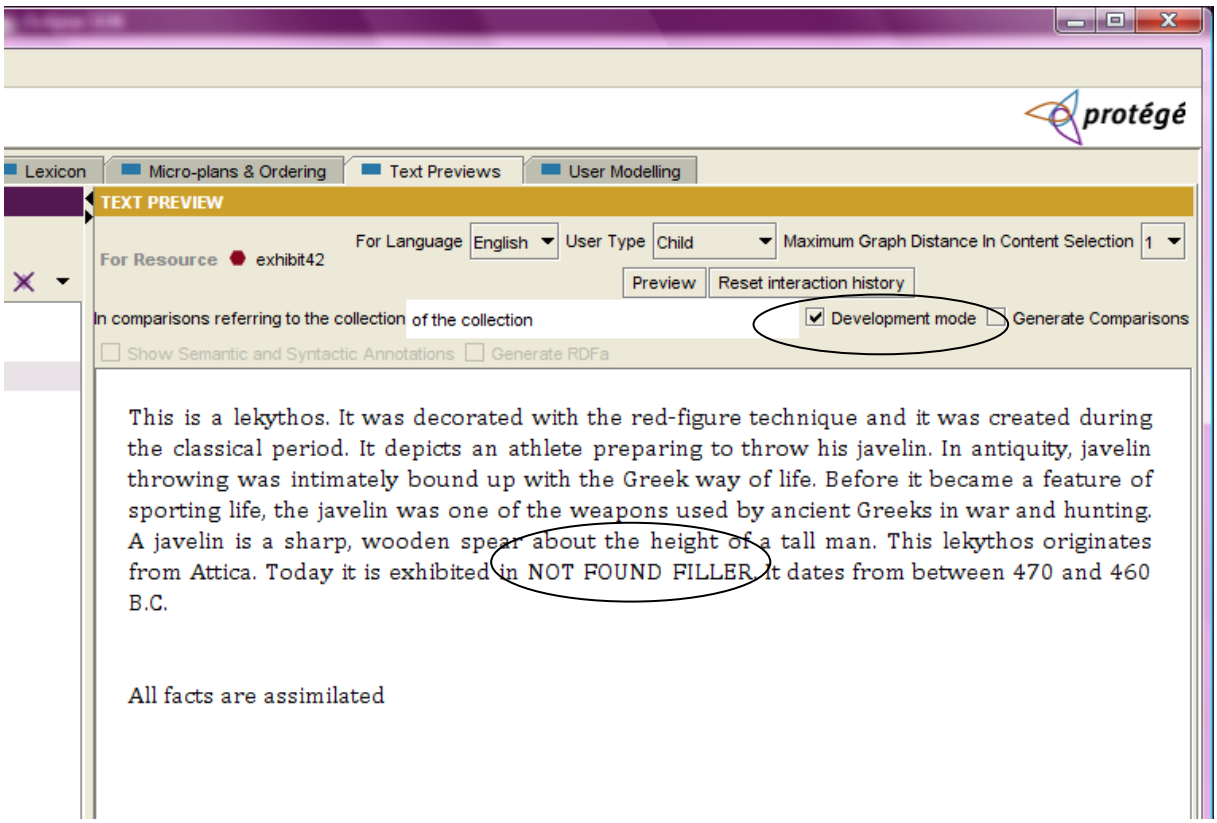
" Today it is exhibited in ***NOT FOUND FILLER****. It dates from between 470 and 460 B.C."*

Επιπλέον προβλήματα ανέκυπταν και στις περιπτώσεις των μικροσχεδίων (microplans), δηλαδή αν δεν είχε δηλωθεί κανένα μικροσχέδιο για κάποια ιδιότητα (property) της οντολογίας ή αν είχε δηλωθεί και στη συνέχεια είχαν διαγραφεί όλες οι θέσεις (slots) του. Στις περιπτώσεις αυτές εμφανιζόταν στο παραγόμενο κείμενο ένα μήνυμα λάθους της μορφής:

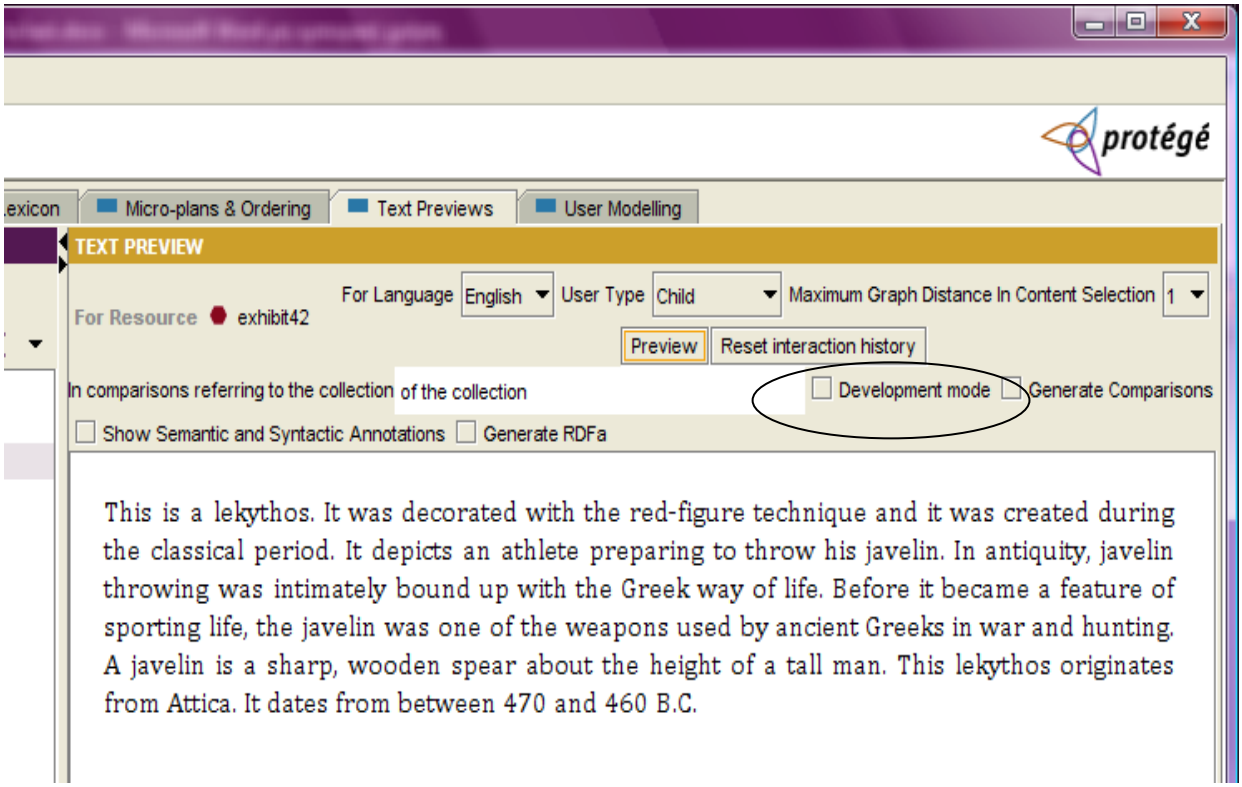
MICROPLAN NOT FOUND ή **Owner [property name] Filler**.

Κατά τη «συγγραφή», όσο δηλαδή εισάγονται στο σύστημα πληροφορίες για μια οντολογία, τις τάξεις της, τις οντότητές της, το λεξικό της, τα σχεδιότυπά της, τους τύπους χρηστών της κλπ., είναι χρήσιμο να εμφανίζονται τα παραπάνω μηνύματα, διότι δίνεται η δυνατότητα στο χρήστη-συγγραφέα να παρατηρήσει τις ελλείψεις στα παραπάνω δεδομένα. Όταν όμως το σύστημα χρησιμοποιείται από τελικούς χρήστες (π.χ. για την περιγραφή εκθεμάτων μουσείων σε επισκέπτες), είναι προτιμότερο να μην εμφανίζονται μηνύματα λάθους.

Εφόσον το σύστημα μπορεί να χρησιμοποιηθεί από δύο είδη χρηστών (συγγραφείς και τελικούς), προστέθηκε η δυνατότητα παραγωγής κειμένου σε δύο διαφορετικές μορφές. Κατά συνέπεια, προστέθηκε ένα πεδίο επιλογής κατάστασης στην καρτέλα «Text Previews». Όταν επιλέγεται το πεδίο «Development mode» που φαίνεται στην εικόνα 1, το σύστημα παράγει κείμενα όπως έκανε μέχρι σήμερα, εξυπηρετώντας κυρίως τους χρήστες-συγγραφείς. Όταν δεν είναι επιλεγμένο το πεδίο αυτό, τα κείμενα που παράγονται είναι απαλλαγμένα από μηνύματα λάθους.



Εικόνα 1: Παραγόμενο κείμενο για χρήστες-συγγραφείς.

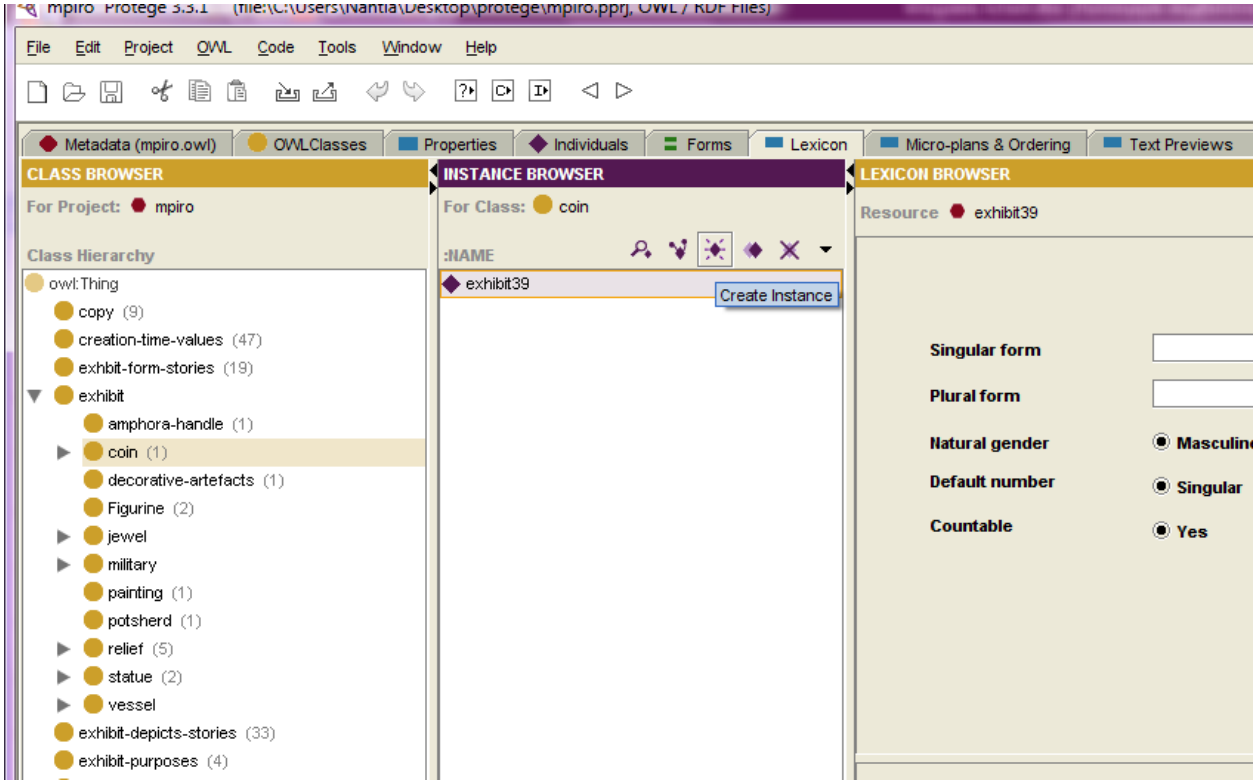


Εικόνα 2: Παραγωγή κειμένου για τελικούς χρήστες.

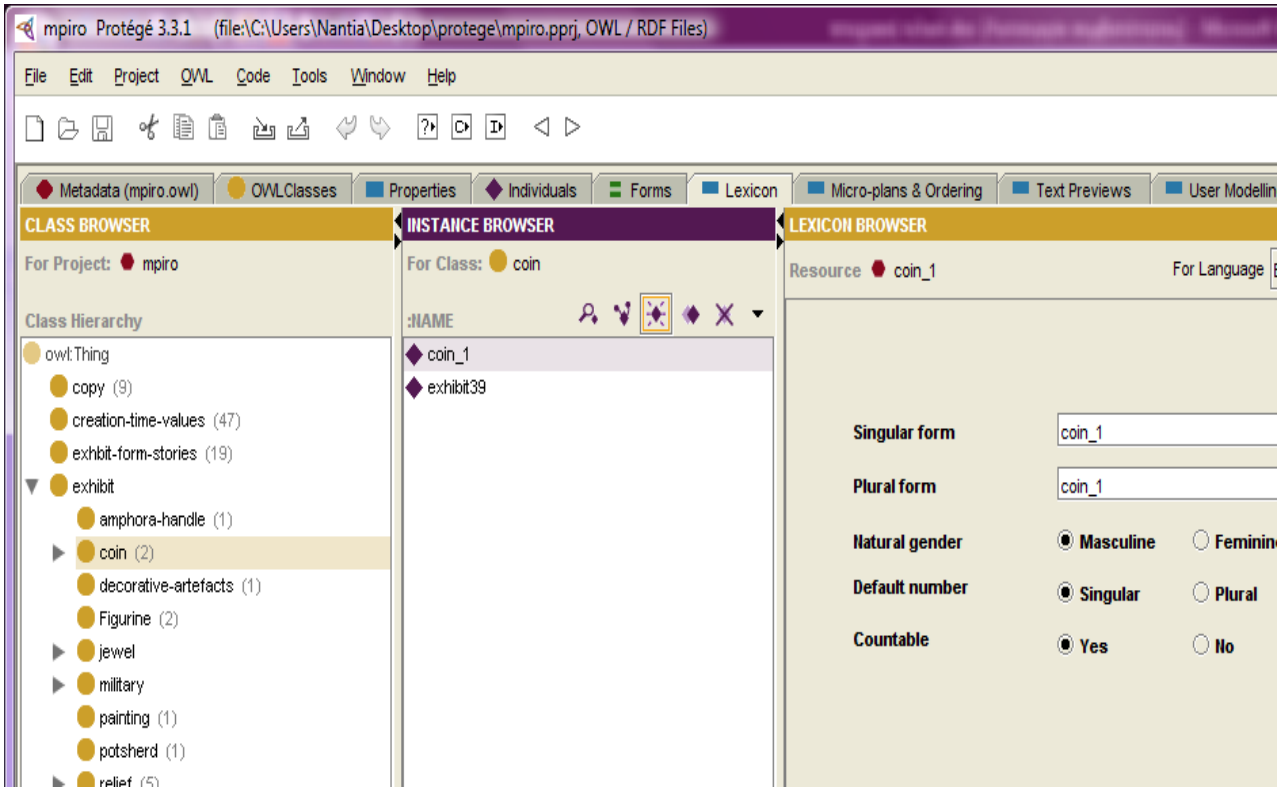
2.2 Αυτόματη εισαγωγή εγγραφών στο λεξικό

Στο αρχικό σύστημα όταν ο χρήστης δημιουργούσε μια οντότητα (instance) μιας τάξης ή μια τάξη, δε δημιουργούταν αυτόματα η αντίστοιχη εγγραφή στο λεξικό που συνοδεύει την οντολογία. Αυτό είχε ως αποτέλεσμα να προκύπτουν μηνύματα της μορφής, που αναφέρθηκαν στην ενότητα 2.1 .

Για την αντιμετώπιση αυτού του προβλήματος, σε κάθε περίπτωση που ο χρήστης δημιουργεί μια οντότητα (instance) μιας τάξης ή μια τάξη, εισάγεται αυτόματα μια καινούρια εγγραφή στο λεξικό. Η εγγραφή αυτή είναι η ονομασία της οντότητας, που δόθηκε αυτόματα από το σύστημα, και είναι ίδια για τον ενικό και πληθυντικό αριθμό για τα αγγλικά καθώς και για όλες τις πτώσεις των δύο αριθμών αν πρόκειται για τα ελληνικά. Ο χρήστης μπορεί στη συνέχεια να τροποποιήσει την εγγραφή του λεξικού.



Εικόνα 3: Δημιουργία νέας οντότητας της τάξης coin.



Εικόνα 4: Αυτόματη δημιουργία της εγγραφής του λεξικού για τη νέα οντότητα.

2.3 Χρήση λίστας κουκκίδων σε ιδιότητα με πολλαπλές τιμές

Στη διάρκεια αυτής της εργασίας, προστέθηκε η δυνατότητα χρήσης λίστας κουκκίδων (bullet list), όταν υπάρχουν πολλαπλές τιμές σε κάποια ιδιότητα οντότητας. Σε αυτή την περίπτωση, στα παραγόμενα κείμενα του αρχικού συστήματος οι πολλαπλές τιμές της ιδιότητας ενώνονται είτε με κόμμα, είτε με το «και» ή και με τα δύο στοιχεία, όπως φαίνεται στο παράδειγμα που ακολουθεί.

«Ion Androutsopoulos is an assistant professor. He is the head of the Natural Language Processing group. He supervises: Lampouras Gerasimos, Malakasiotis Makis, Galanis Dimitris, Kosmopoulos Aris. He cooperates with Vasilis Vassalos. He works on the INDIGO project. His publications are : An Open-Source Natural Language Generator for OWL Ontologies and its Use in Protege and Second Life, Generating Multilingual Descriptions from Linguistically Annotated OWL Ontologies: the NaturalOWL System, Learning Textual Entailment using SVMs and String Similarity Measures, Adaptive Spam Filtering Using Only Naive Bayes Text Classifiers. He co-organized the 12th Conference of the European Chapter of the Association for Computational Linguistics. »

Στο βελτιωμένο σύστημα, επιλέγοντας τη χρήση λίστας με κουκκίδες, παράγεται το παρακάτω κείμενο:

«Ion Androutsopoulos is an assistant professor. He is the head of Natural Language Processing group. He supervises :

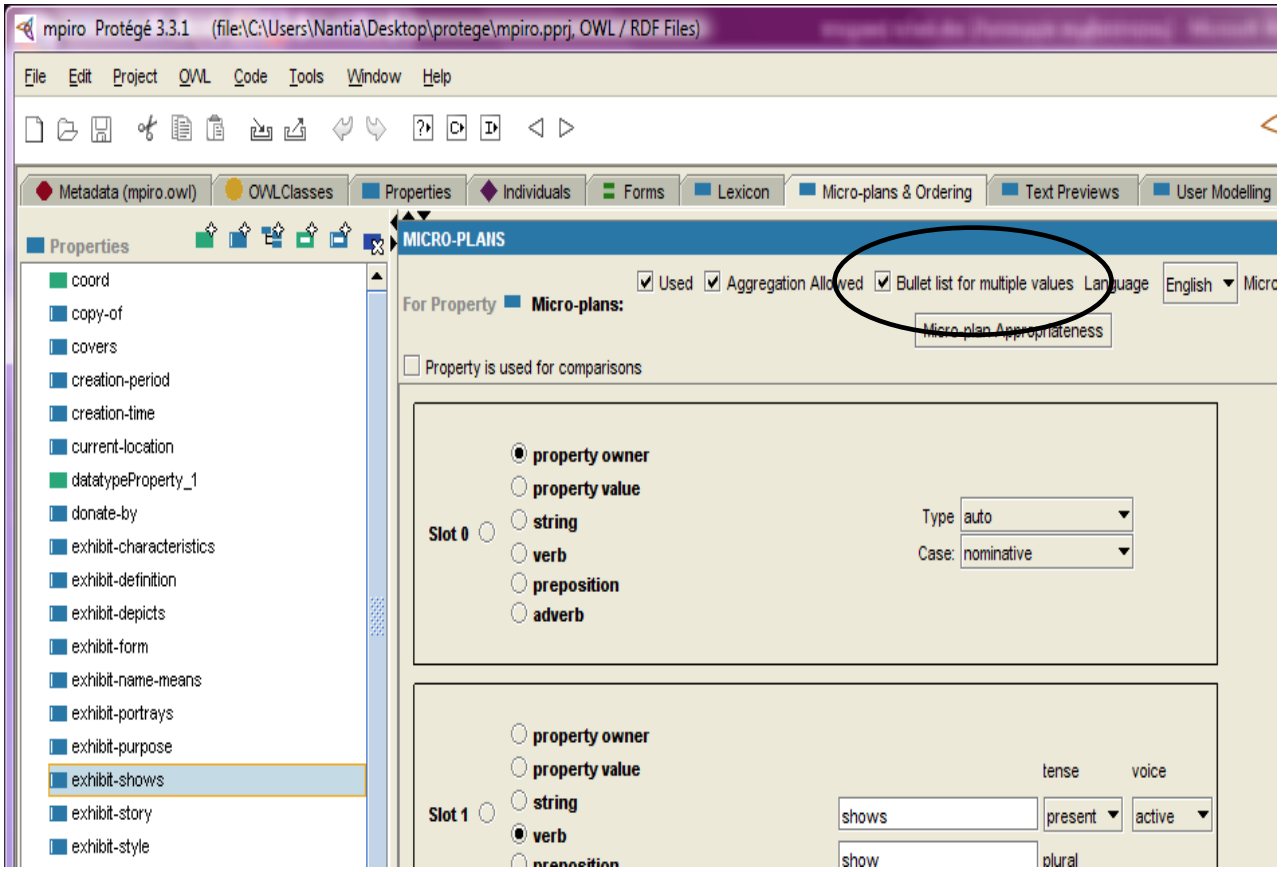
- *Lampouras Gerasimos,*
- *Malakasiotis Makis,*
- *Galanis Dimitris,*
- *Kosmopoulos Aris.*

He cooperates with Vasilis Vassalos . He works on the INDIGO project. His publications are :

- *An Open-Source Natural Language Generator for OWL Ontologies and its Use in Protege and Second Life,*
- *Generating Multilingual Descriptions from Linguistically Annotated OWL Ontologies: the NaturalOWL System,*
- *Learning Textual Entailment using SVMs and String Similarity Measures,*
- *Adaptive Spam Filtering Using Only Naive Bayes Text Classifiers.*

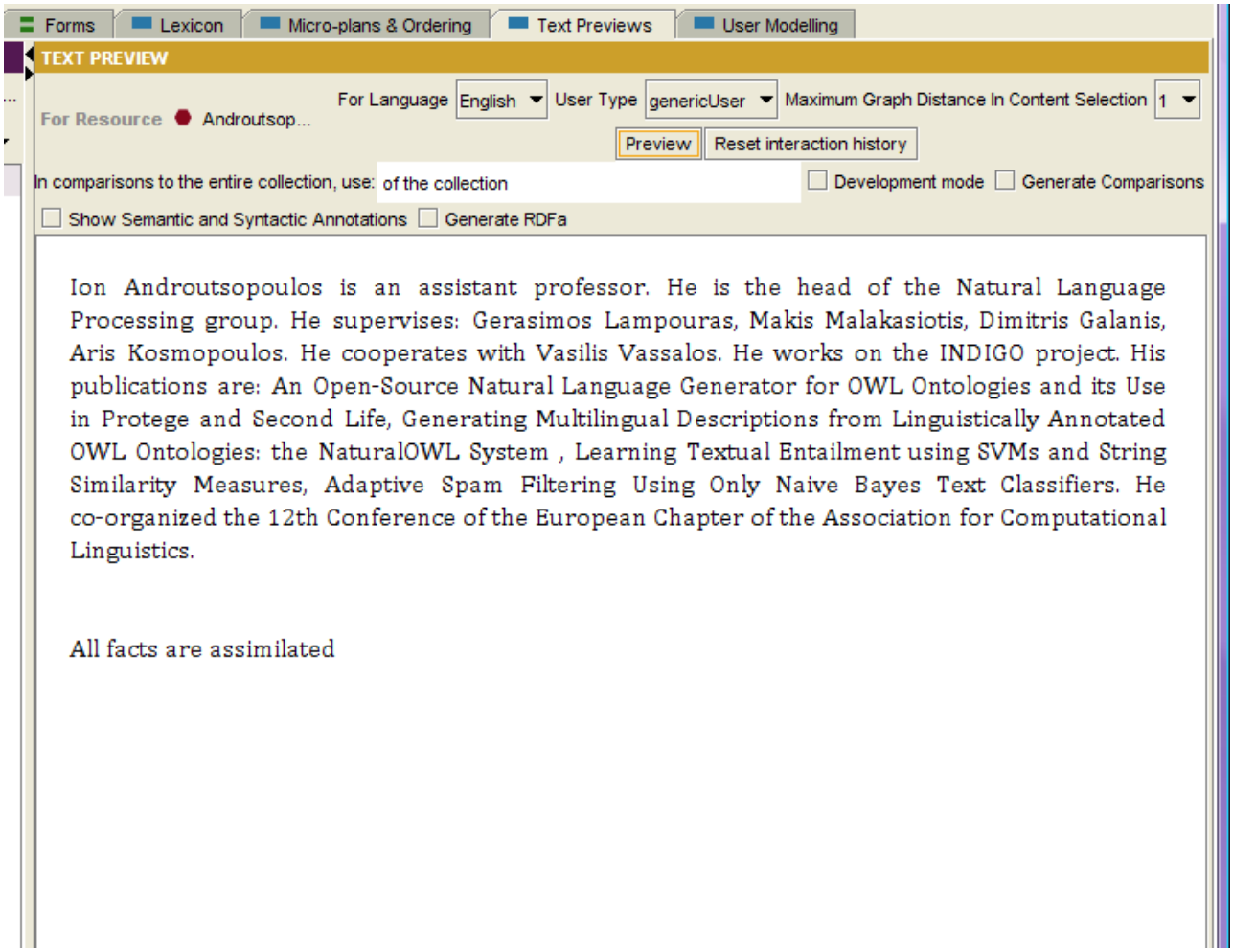
He co-organized the 12th Conference of the European Chapter of the Association for Computational Linguistics.»

Η δυνατότητα αυτή παρέχεται μέσα από την καρτέλα «Microplans & Ordering», όπως φαίνεται στην εικόνα 5.

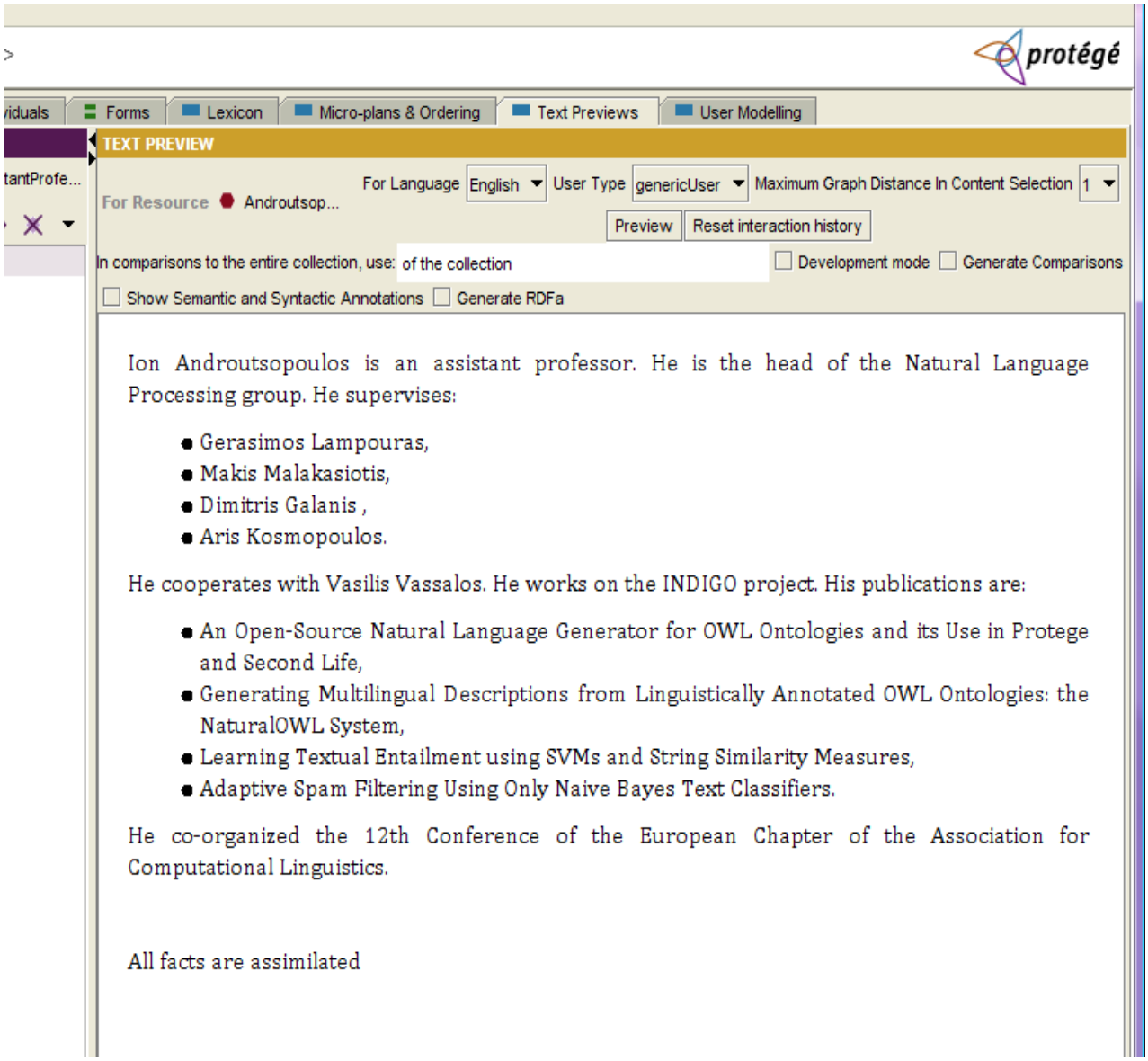


Εικόνα 5: Επιλογή χρήσης λίστας με κουκκίδες για την ιδιότητα "exhibit-shows".

Στις εικόνες 6 και 7 φαίνεται η χρήση λίστας με κουκκίδες στην περίπτωση πολλαπλών τιμών μιας ιδιότητας.



Εικόνα 6: Κείμενο χωρίς χρήση λίστας με κουκκίδες.



Εικόνα 7: Παραγωγή κειμένου με χρήση λίστας κουκκίδων.

2.4 Χρήση κατάλληλης φράσης στην παραγωγή κειμένων για οντότητες που είναι μοναδικές σε μια τάξη της οντολογίας

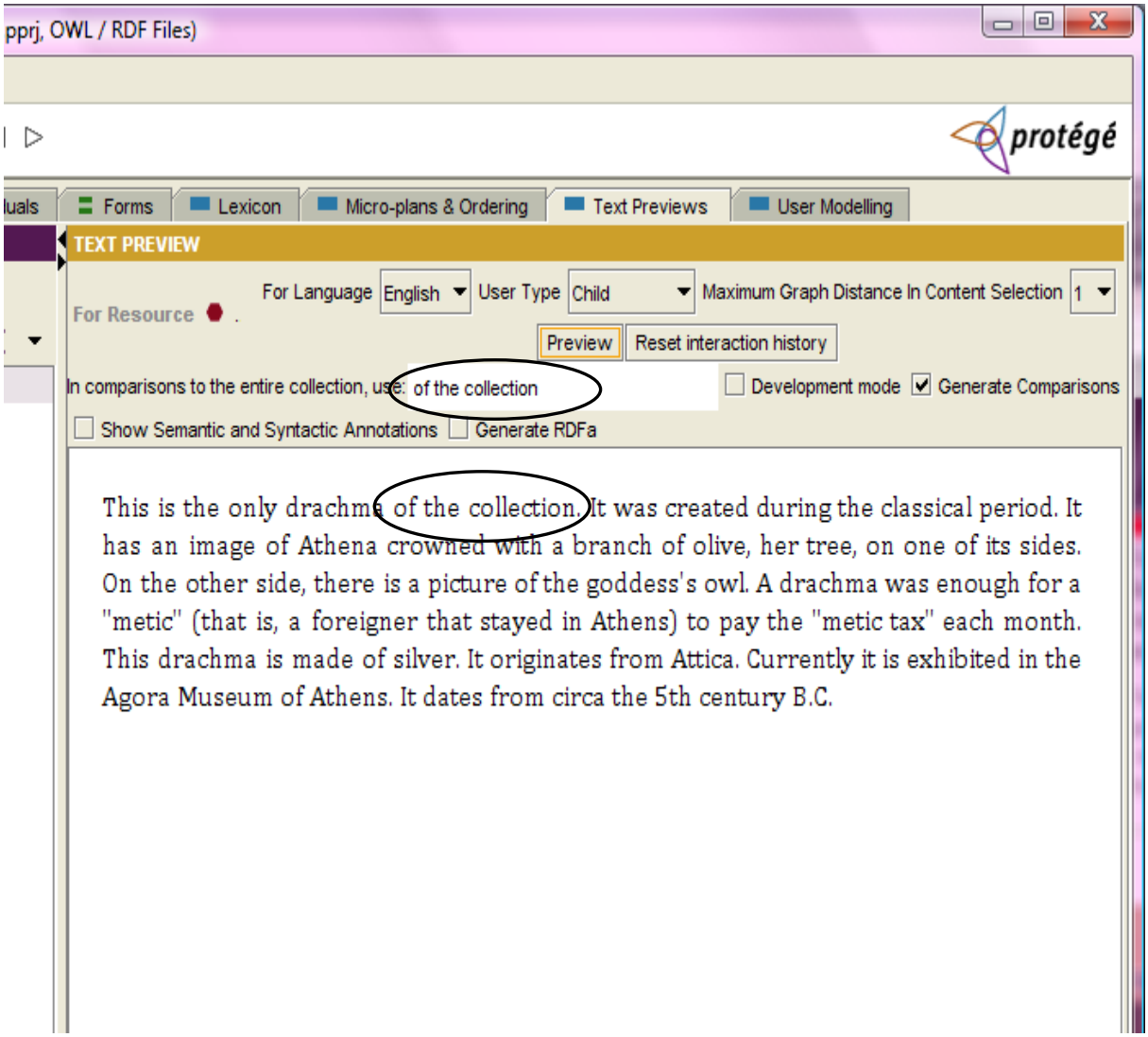
Για να αναφέρει το σύστημα ότι μόνο μία οντότητα έχει μια συγκεκριμένη ιδιότητα στην οντολογία χρησιμοποιεί μια φράση της μορφής:

"This is the only amphora *of the collection*" για τα αγγλικά κείμενα

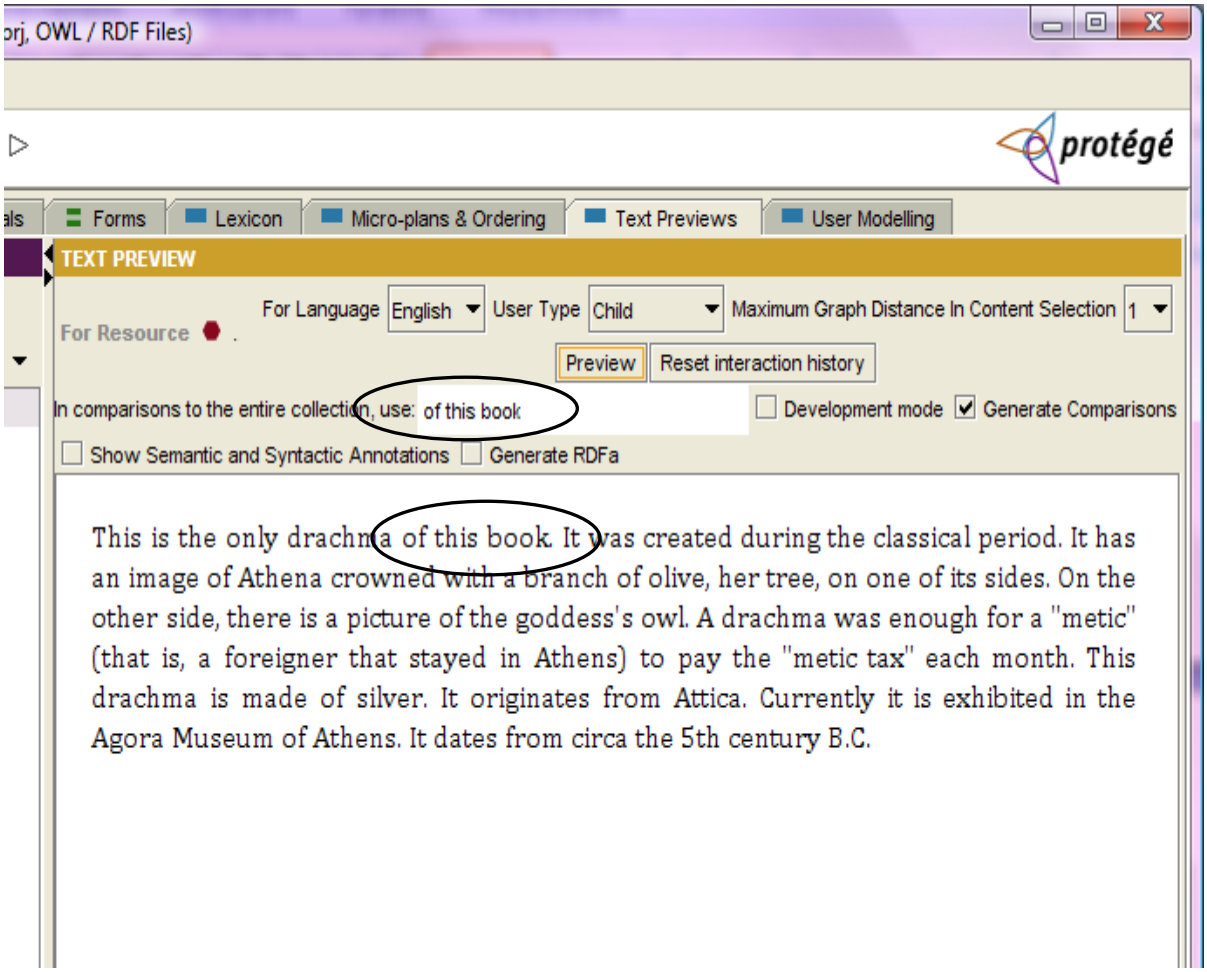
και

«Αυτός είναι ο μοναδικός αμφορέας *της αίθουσας*» για τα ελληνικά.

Η φράση «of the collection/της αίθουσας» ήταν προκαθορισμένη στο υπάρχον σύστημα και δεν υπήρχε τρόπος να την αλλάξει εύκολα ο χρήστης-συγγραφέας, ενώ αυτό χρειαζόταν σε διαφορετικές οντολογίες που δεν περιγράφουν εκθέματα αιθουσών. Στη διάρκεια της εργασίας, προστέθηκε στην καρτέλα Text Previews, ένα νέο πεδίο εισαγωγής κειμένου, όπου ο χρήστης μπορεί να επιλέξει την κατάλληλη φράση, ανάλογα με την οντολογία, όπως φαίνεται παρακάτω.



Εικόνα 8: Παραμετροποίηση της προεπιλεγμένης φράσης «of the collection».



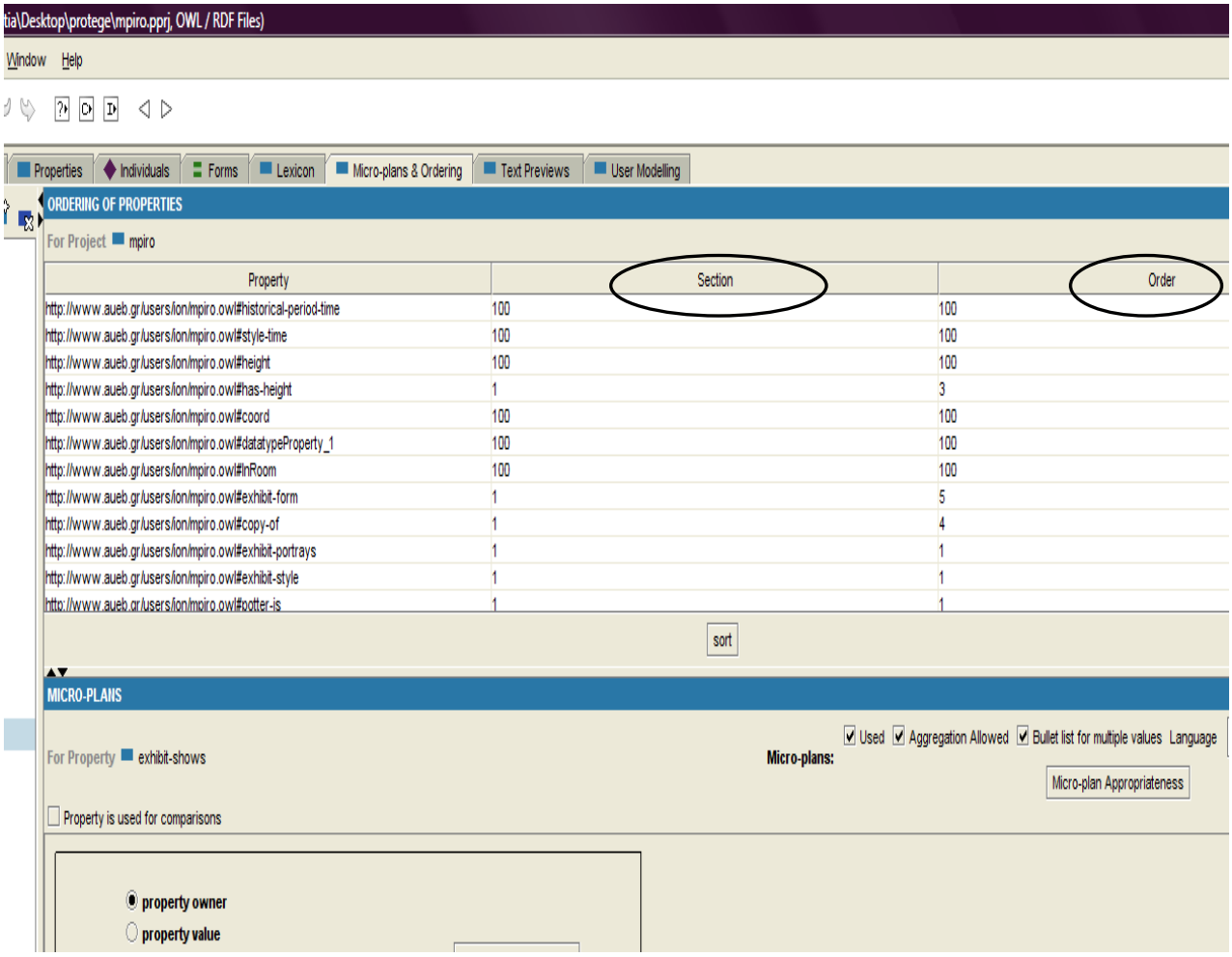
Εικόνα 9: Αλλαγή της φράσης από «of the collection» σε «of this book».

Αν ο χρήστης αφήσει το πεδίο κενό, τότε το σύστημα θα χρησιμοποιήσει την προεπιλεγμένη έκφραση.

2.5 Χωρισμός ιδιοτήτων σε ενότητες και ορισμός προτεραιότητας των ενοτήτων

Το αρχικό σύστημα μας δίνει τη δυνατότητα να ορίσουμε μια τιμή-σειράς (order) για κάθε ιδιότητα (property). Οι τιμές αυτές ορίζουν μια μερική διάταξη (partial order) με την οποία θα πρέπει να εμφανίζονται στο παραγόμενο κείμενο οι προτάσεις που αντιστοιχούν στις ιδιότητες· προτάσεις που αντιστοιχούν σε ιδιότητες με την ίδια τιμή-σειράς επιτρέπεται να εμφανιστούν με οποιαδήποτε διάταξη στο κείμενο. Γειτονικές προτάσεις είναι δυνατόν να συνενωθούν σε μεγαλύτερες (sentence aggregation). Συχνά προκύπτουν έτσι, όμως, κείμενα στα οποία συνενώνονται προτάσεις ιδιοτήτων που δεν έχουν νοηματική συνάφεια, δηλαδή

από διαφορετικές νοηματικές ενότητες (π.χ. τρόπος χρήσης κατασκευής ενός εκθέματος με τη σημερινή του θέση σε ένα μουσείο) με αποτέλεσμα τα κείμενα να φαίνονται αφύσικα. Προκειμένου να αντιμετωπιστεί αυτό το πρόβλημα, στη διάρκεια της εργασίας προστέθηκε ένα καινούριο πεδίο που χαρακτηρίζει κάθε ιδιότητα. Το πεδίο αυτό ονομάζεται ενότητα (section) και φανερώνει τη νοηματική ενότητα στην οποία ανήκει η ιδιότητα, ενώ το υπάρχον πεδίο σειρά (order) αντιστοιχεί πλέον στη μερική διάταξη της ιδιότητας μέσα στην ενότητά (section) της. Κατά την παραγωγή κειμένου, οι ιδιότητες αρχικά ομαδοποιούνται κατά ενότητες· η ομαδοποίηση αυτή γίνεται κατά αύξοντα αριθμό ενότητας. Στη συνέχεια, οι ιδιότητες κάθε ενότητας διατάσσονται, ώστε να ικανοποιείται η μερική διάταξη.



Εικόνα 10: Πεδίο ενότητας («section») για κάθε ιδιότητα.

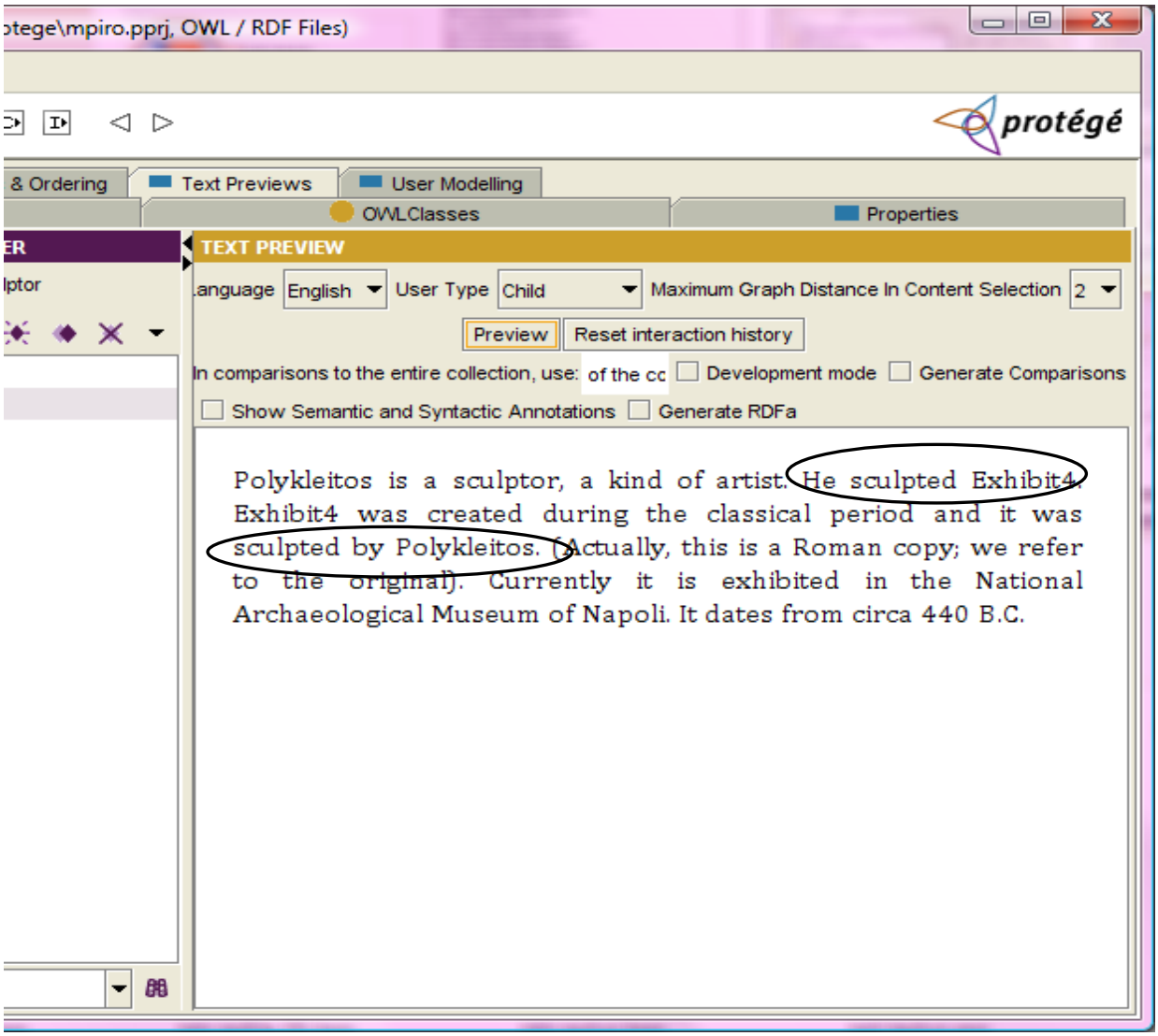
2.6 Χειρισμός περιττών αντίστροφων γεγονότων

Κατά την παραγωγή κειμένου για ένα στιγμιότυπο μιας τάξης, κι εφόσον προκύψουν κατά το στάδιο της επιλογής περιεχομένου ιδιότητες οι οποίες είναι αντίστροφες μεταξύ τους, το

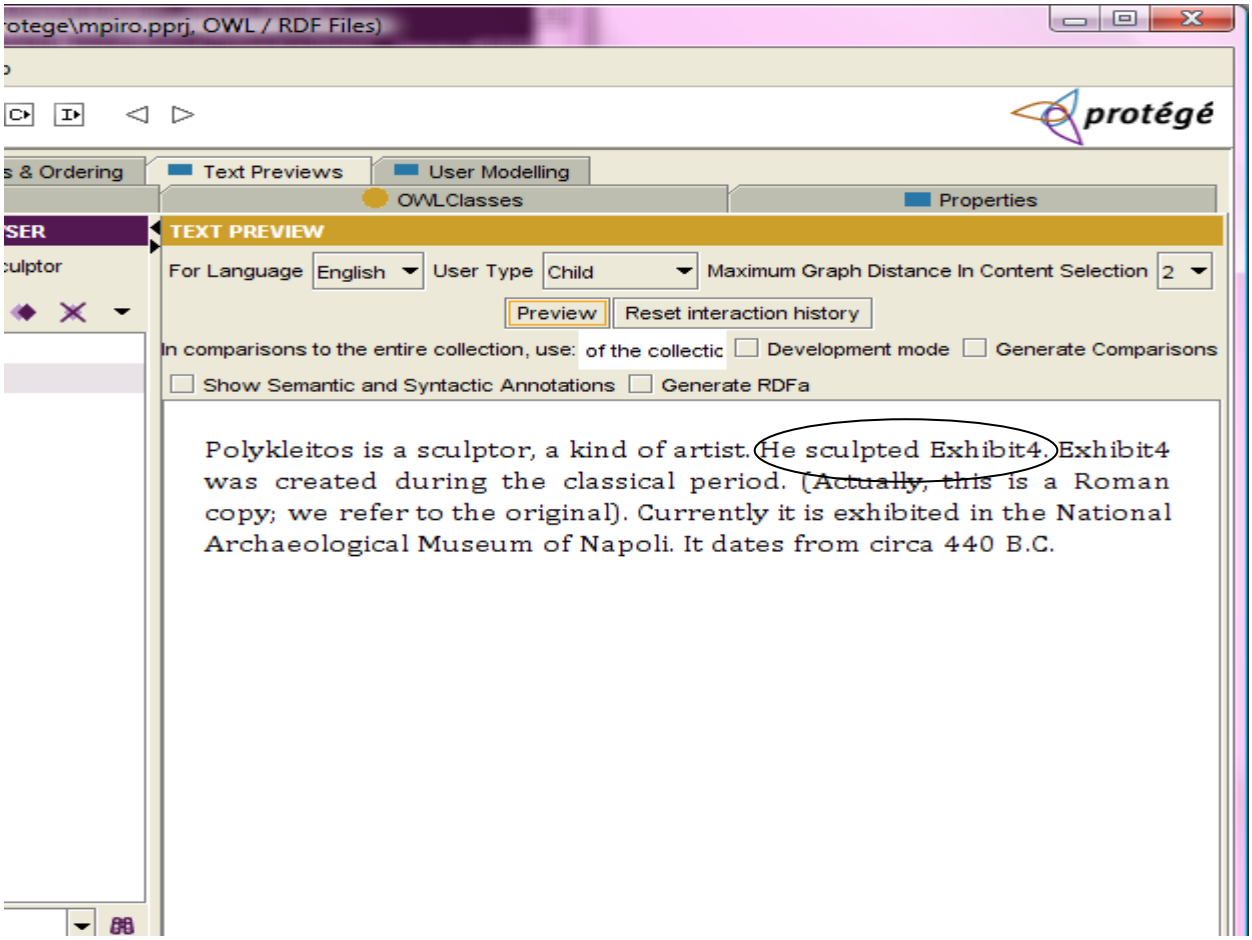
παραγόμενο κείμενο θα περιέχει την ίδια πληροφορία δύο φορές, όπως φαίνεται στο παράδειγμα που ακολουθεί και στην εικόνα 11. Το έκθεμα “exhibit4” δημιουργήθηκε μόνο για τις ανάγκες του παραδείγματος.

“Polykleitos is a sculptor, a kind of artist. He sculpted exhibit4. Exhibit4 was created during the classical period and it was sculpted by Polykeitos. (Actually, this is a Roman copy; we refer to the original). Currently it is exhibited in the National Archaeological Museum of Napoli. It dates from circa 440 B.C. “

Το κείμενο αναφέρει δύο ισοδύναμα γεγονότα (facts), στα οποία συμμετέχουν δύο αντίστροφες ιδιότητες, σχηματικά: <Πολύκλειτος, sculpted, exhibit4> και <exhibit4, sculpted-by, Πολύκλειτος>. Μόνο το ένα από τα δύο γεγονότα πρέπει να αναφερθεί, αλλά στο μοντέλο χρήστη (user model) πρέπει να προστεθεί η πληροφορία ότι ο χρήστης έχει ενημερωθεί και για τα δύο γεγονότα, αφού είναι ισοδύναμα.



Εικόνα 11: Ισοδύναμα γεγονότα με αντίστροφες ιδιότητες σε ένα παραγόμενο κείμενο.

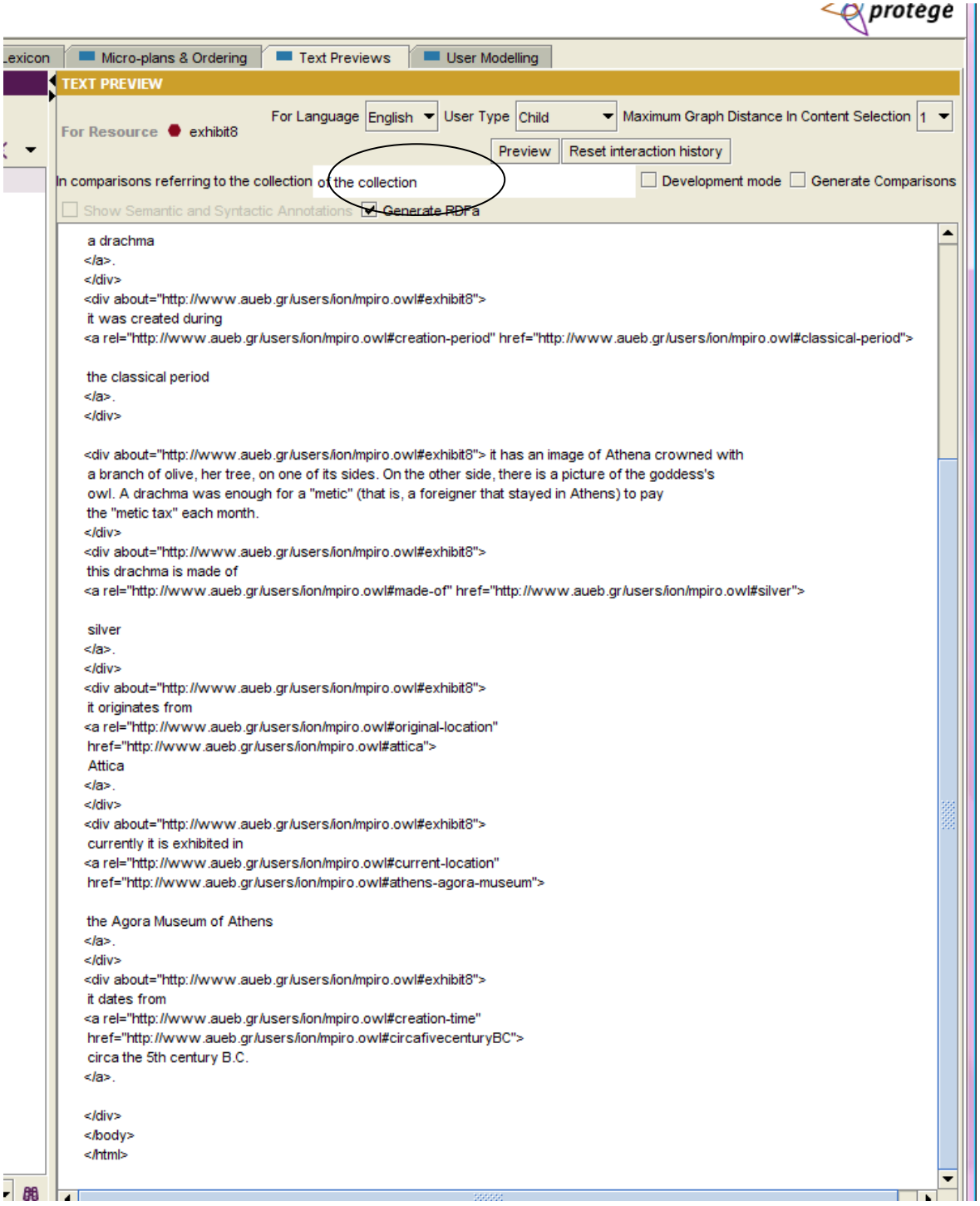


Εικόνα 12: Παραγωγή κειμένου που περιέχει μόνο το ένα από τα δύο ισοδύναμα γεγονότα.

2.7 Δυνατότητα παραγωγής κειμένων με επισημειώσεις RDFa

Στην καρτέλα Text Previews προστέθηκε ένα πεδίο επιλογής παραγωγής κειμένων επισημειωμένων με RDFa [7], που ονομάζεται «Generate RDFa». Όταν επιλέγεται αυτό το πεδίο, απενεργοποιείται η χρήση του πεδίου «Show Semantic and Syntactic Annotations», ώστε να μην παράγεται ταυτόχρονα κείμενο με δυο διαφορετικά είδη

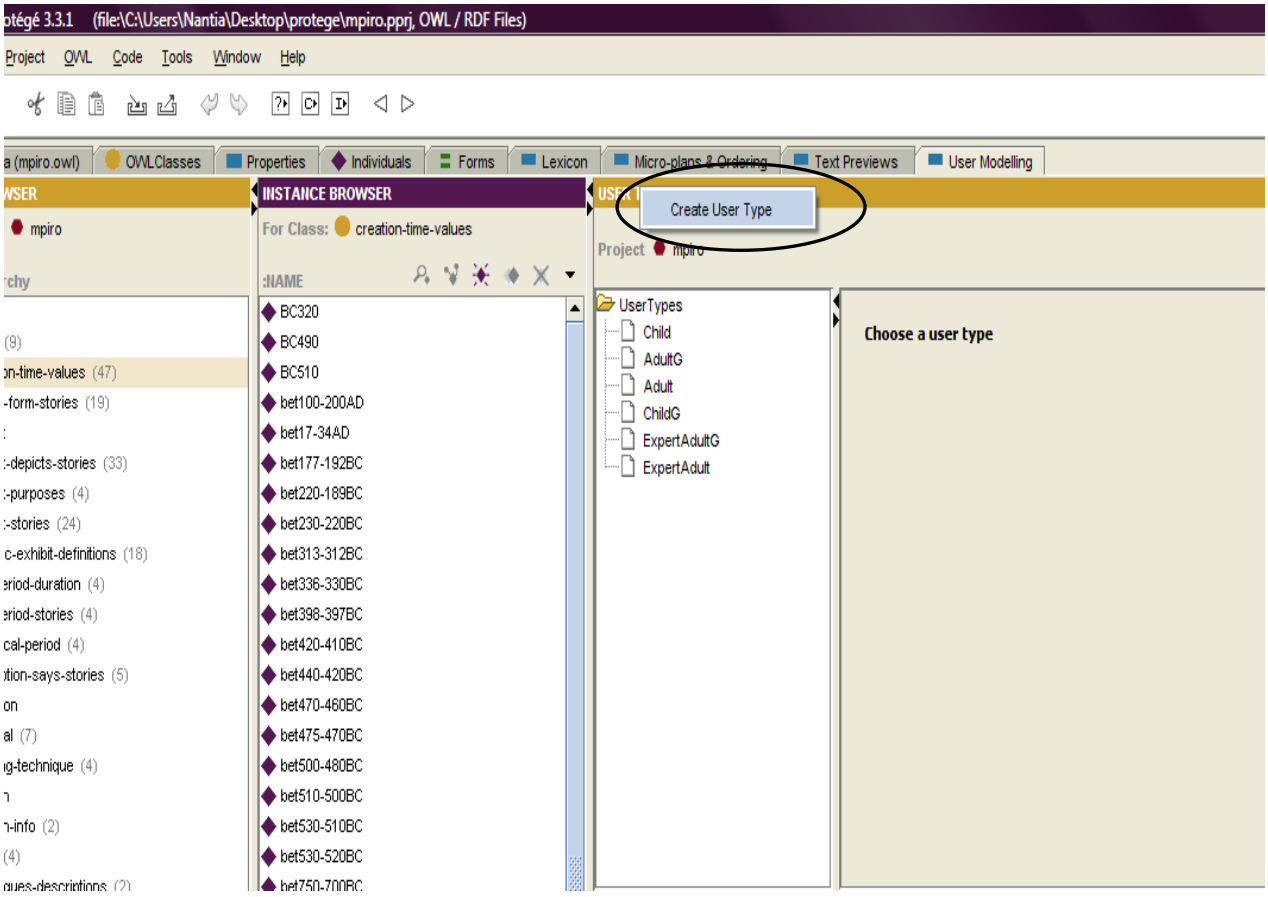
επισημειώσεων.



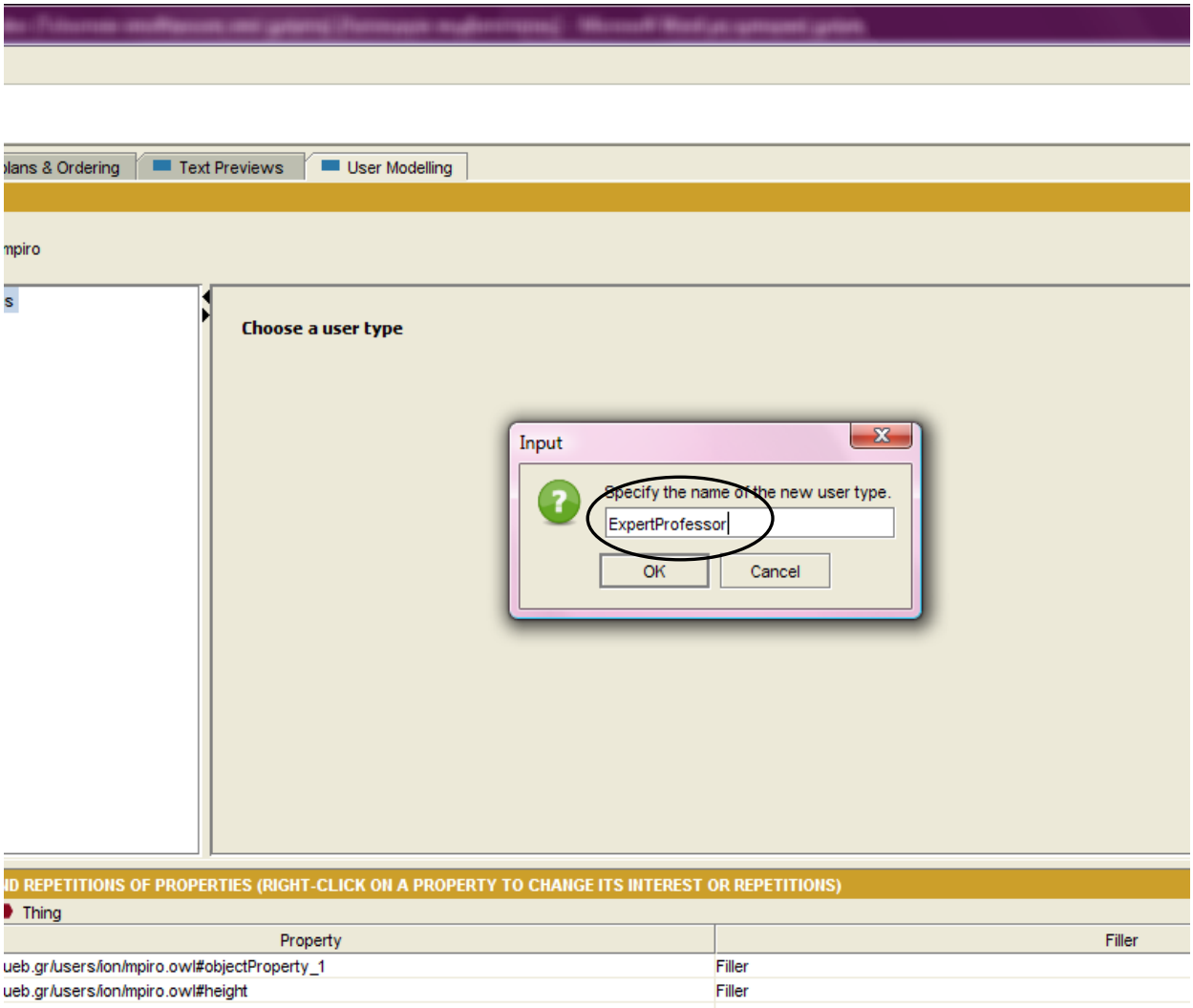
Εικόνα 13: Παραγωγή κειμένου με RDFa markup.

2.8 Αυτόματη εισαγωγή πληροφοριών για νέο τύπο χρηστών

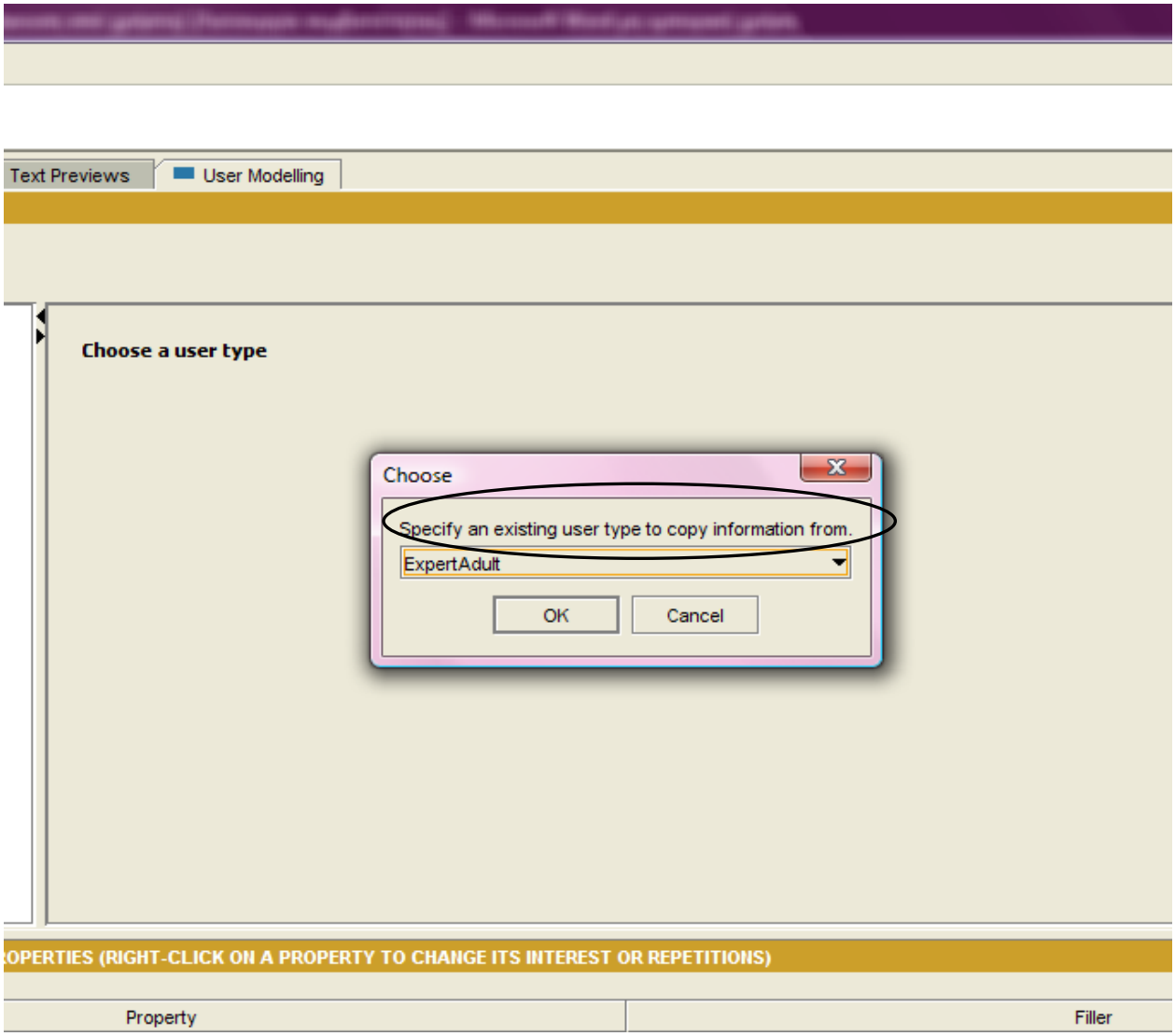
Στην προηγούμενη μορφή του συστήματος, όποτε δημιουργείτο ένας νέος τύπος χρηστών, όλες οι πληροφορίες για το νέο τύπο έπρεπε να εισαχθούν από την αρχή. Στο βελτιωμένο σύστημα, κατά τη δημιουργία ενός νέο τύπου χρήστη, είναι δυνατόν να επιλεγεί κάποιος από τους ήδη υπάρχοντες τύπους και να αντιγραφούν από εκείνον οι πληροφορίες στο νέο τύπο. Οι πληροφορίες είναι δυνατόν στη συνέχεια να τροποποιηθούν. Αυτό είναι χρήσιμο όταν ένας τύπος χρηστών έχει μικρές μόνο διαφορές από υπάρχοντες τύπους.



Εικόνα 14: Δημιουργία νέου τύπου τελικού χρήστη του συστήματος.



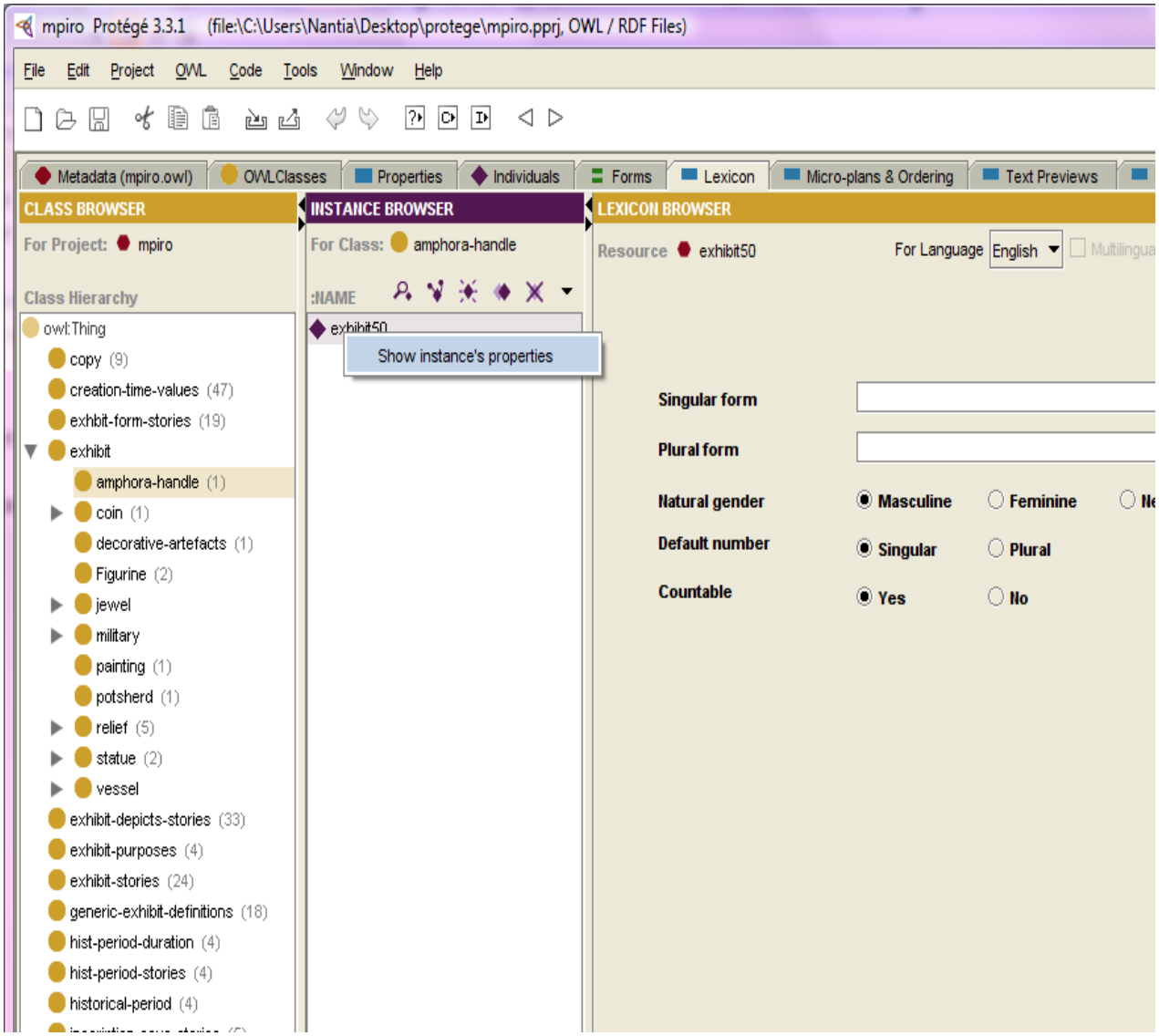
Εικόνα 15: Ονομασία νέου τύπου τελικού χρήστη.



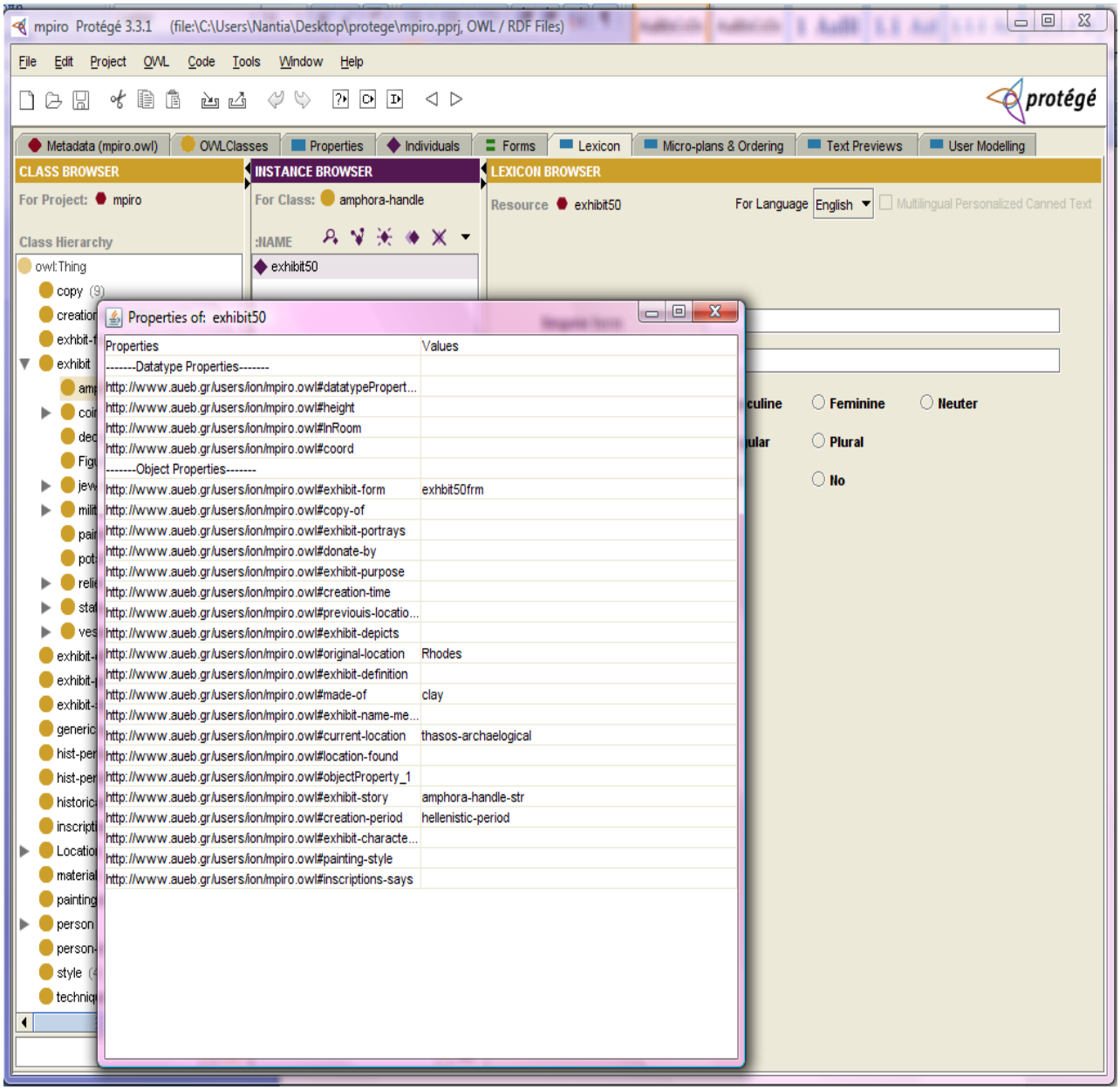
Εικόνα 16: Επιλογή υπάρχοντος τύπου χρήστη, αντίγραφο του οποίου θα είναι αρχικά ο νέος τύπος χρήστη.

2.9 Προβολή λεπτομερειών εγγραφής του λεξικού σε αναδυόμενο παράθυρο

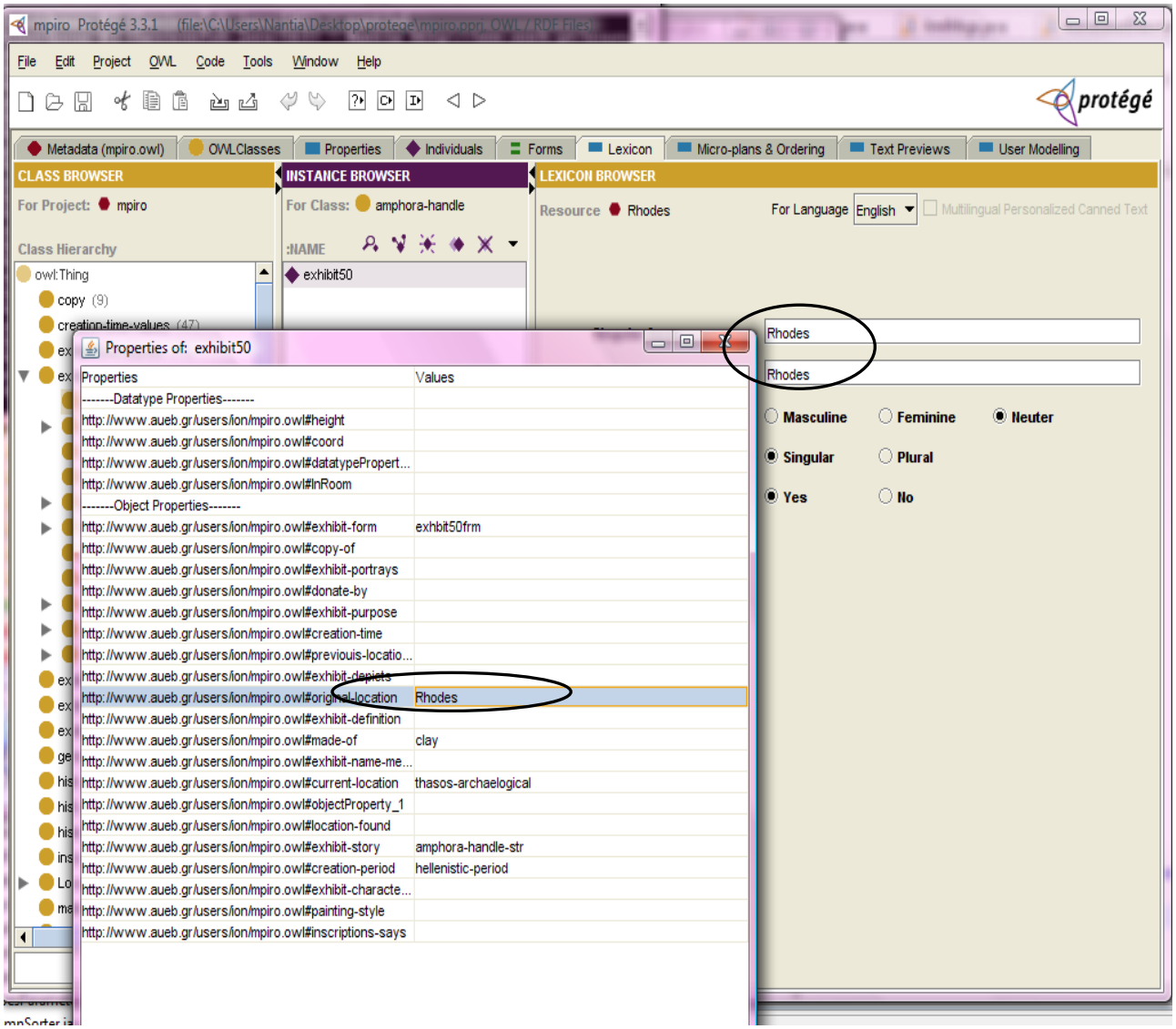
Ο χρήστης έχει τη δυνατότητα, όταν βρίσκεται στην καρτέλα «Lexicon» να κάνει δεξί κλικ πάνω σε μια οντότητα του Instance Browser και να επιλέξει «Show lexicon entry's details» στο αναδυόμενο μενού που εμφανίζεται. Στη συνέχεια, εμφανίζεται ένα παράθυρο με όλες τις ιδιότητες της συγκεκριμένης οντότητας και τις τιμές τους. Ο χρήστης μπορεί να επιλέξει την τιμή κάποιας ιδιότητας, που ενδέχεται να είναι μια άλλη οντότητα. Στην περίπτωση αυτή, εμφανίζεται στο Lexicon Browser η εγγραφή του λεξικού που αντιστοιχεί στην οντότητα-τιμή και ο χρήστης έχει τη δυνατότητα να τροποποιήσει την εγγραφή.



Εικόνα 17: Επιλογή οντότητας, στην περίπτωση που θέλουμε να τροποιήσουμε τις εγγραφές λεξικού που αντιστοιχούν στις οντότητες-τιμές της.



Εικόνα 18: Αναδυόμενο παράθυρο με όλες τις ιδιότητες της οντότητας που επιλέχθηκε και τις τιμές τους.

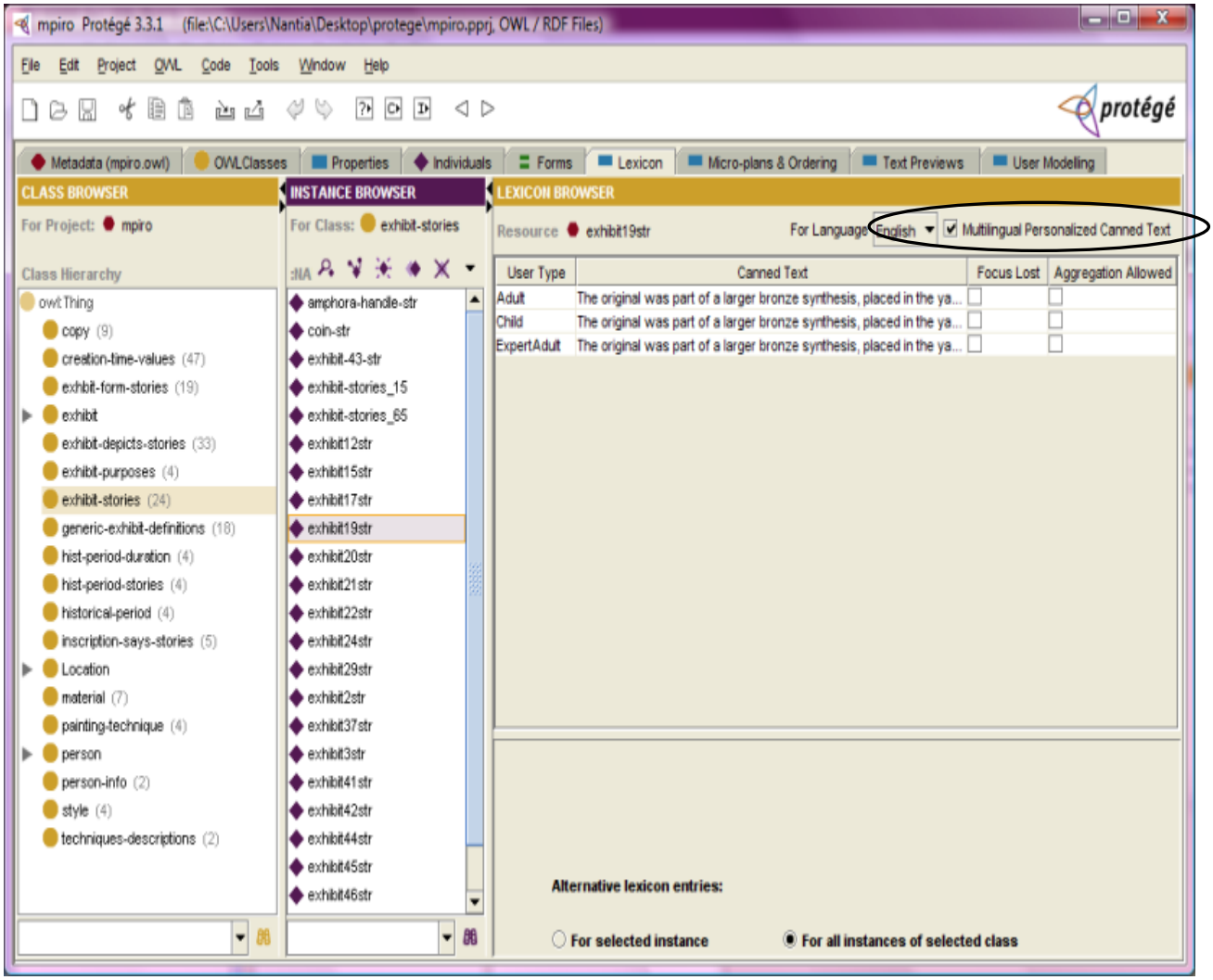


Εικόνα 19: Επιλέγοντας την τιμή-οντότητα μιας ιδιότητας (object property), εμφανίζεται στο lexicon browser η εγγραφή του λεξικού που αντιστοιχεί στην τιμή-οντότητα, ώστε να είναι εύκολη η τροποποίηση της εγγραφής του λεξικού.

2.10 Βελτίωση προβολής των Multilingual Personalized Canned Texts

Στην προηγούμενη μορφή του συστήματος, όταν ο χρήστης βρισκόταν στην καρτέλα «Lexicon» κι επέλεγε το tick box «Multilingual Personalized Canned Texts», τότε το panel διαμορφωνόταν κατάλληλα για να εισαχθούν οι συμβολοσειρές (strings) που αντιστοιχούν στην οντότητα, όπως φαίνεται στην επόμενη εικόνα. Στη συνέχεια, όμως, το panel παρέμενε στην ίδια μορφή, ακόμη κι αν ο χρήστης επέλεγε οντότητα που δεν ήταν Multilingual Personalized Canned Text

(MPCT). Στη νέα μορφή του συστήματος, το panel παίρνει αυτόματα την κατάλληλη μορφή, ανάλογα με το αν η επιλεγόμενη οντότητα έχει δηλωθεί (μέσω του tick box) ως MPCT ή όχι.



Εικόνα 20: Προβολή λεπτομερειών του επιλεγμένου Multilingual Personalized Canned Text

Για να επιστρέψει ο χρήστης της εφαρμογής στο panel που είναι διαμορφωμένο για τις εγγραφές του λεξικού έπρεπε να αλλάξει την επιλογή του στο check box που φαίνεται στην παραπάνω εικόνα και στη συνέχεια να επιλέξει μια τάξη ή οντότητα τάξης που αντιστοιχεί σε ένα ουσιαστικό. Αν επέλεγε μια οντότητα ,που δεν ήταν Multilingual Personalized Canned Text χωρίς να αλλάξει την επιλογή του στο check box τότε το panel δεν άλλαζε και δεν μπορούσε να προβληθεί η σχετική πληροφορία.

Στο βελτιωμένο σύστημα ο χρήστης δε χρειάζεται να αλλάξει την επιλογή του στο check box, αρκεί να επιλέξει μια τάξη ή οντότητα τάξης και το σύστημα ελέγχει ποιο είναι το κατάλληλο panel για την προβολή της πληροφορίας που είναι αποθηκευμένη στο λεξικό.

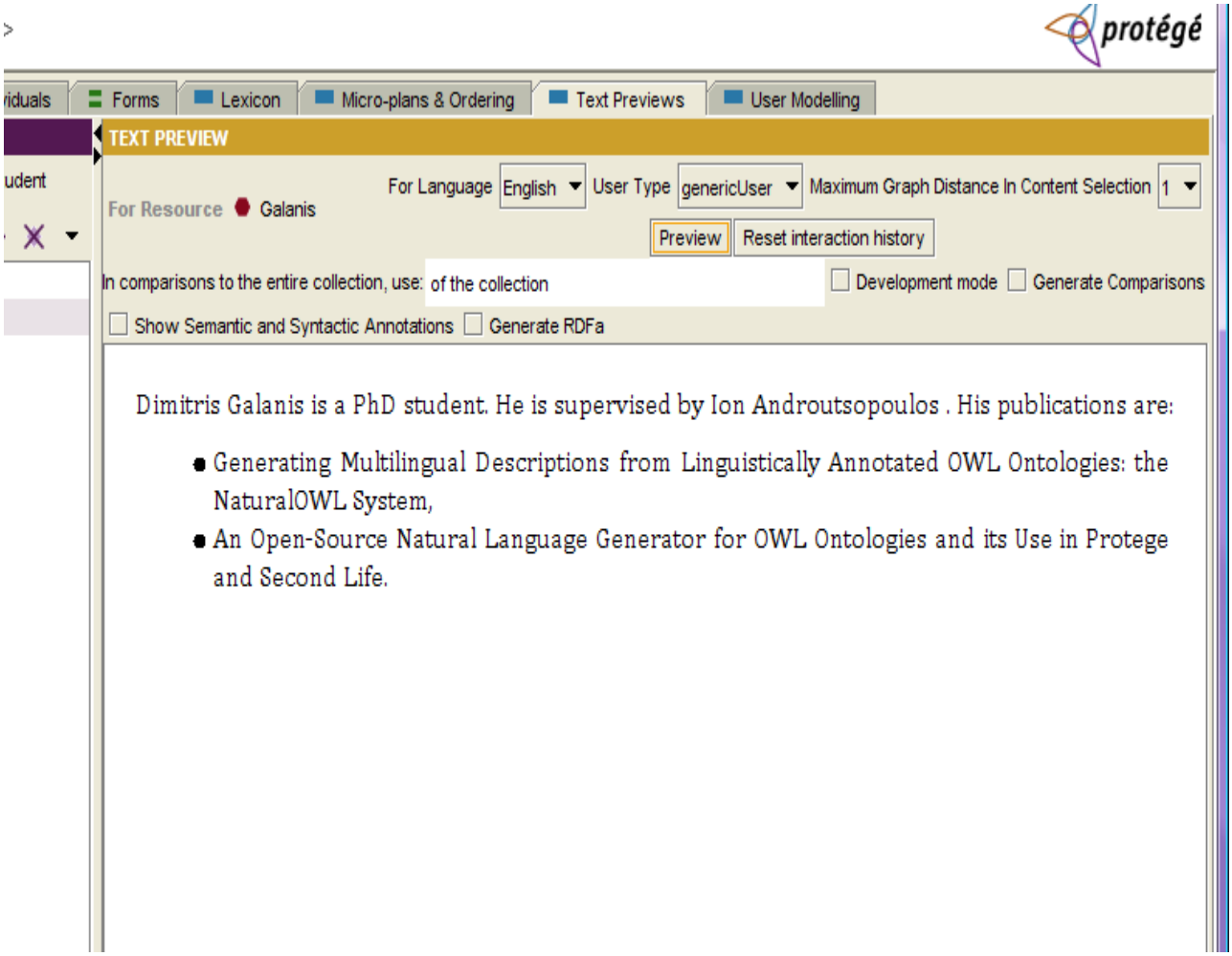
3 Δοκιμές με την οντολογία SWRC

Όπως προαναφέρθηκε, στη διάρκεια της εργασίας χρησιμοποιήθηκε οντολογία «Semantic Web for Research Communities» (SWRC), προκειμένου να ελεγχθεί η συμπεριφορά του συστήματος με μια οντολογία διαφορετική από εκείνες των εκθεμάτων μουσείων που είχαν κυρίως χρησιμοποιηθεί στο παρελθόν. Η SWRC δημιουργήθηκε από την ομάδα OntoWare του «Institute of Applied Informatics and Formal Description Methods» του πανεπιστημίου Karlsruhe της Γερμανίας και χρησιμεύει στη μοντελοποίηση οντοτήτων ερευνητικών κοινοτήτων, όπως φοιτητές, καθηγητές, δημοσιεύσεις και των σχέσεων που τους συνδέουν.
Η δομή των τάξεων της οντολογίας είναι η ακόλουθη:

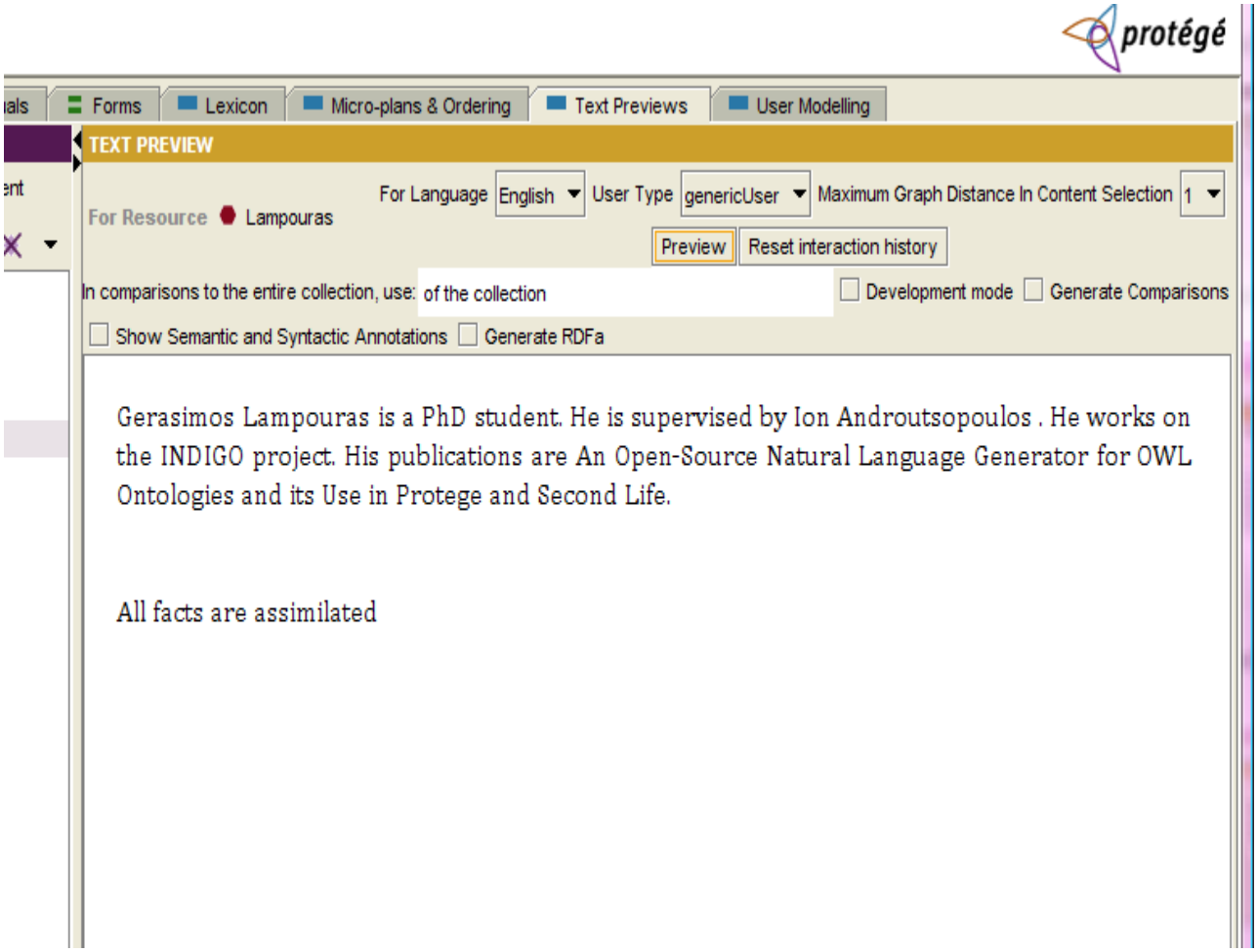
```
owl:Thing
        p1:Event
                p1:Conference
                p1:Exhibition
                p1:Lecture
                p1:Meeting
                        p1:ProjectMeeting
                p1:Workshop
        p1:Organization
                p1:Association
                p1:Department
                p1:Enterprise
                p1:Institute
                p1:ResearchGroup
                p1:University
        p1:Person
                p1:Employee
                        p1:AcademicStaff
                                p1:FacultyMember
                                        p1:AssistantProfessor
                                        p1:AssociateProfessor
                                        p1:FullProfessor
                                p1:Lecturer
                        p1:AdministrativeStaff
                        p1:Manager
                        p1:TechnicalStaff
                p1:Student
                        p1:Graduate
```

p1:PhDStudent
p1:Undergraduate
p1:Product
p1:SoftwareComponent
p1:Project
p1:DevelopmentProject
p1:SoftwareProject
p1:ResearchProject
p1:Publication
p1:Article
p1:Book
p1:Booklet
p1:InBook
p1:InCollection
p1:InProceedings
p1:Manual
p1:Misc
p1:Proceedings
p1:Report
p1:ProjectReport
p1:TechnicalReport
p1:Thesis
p1:MasterThesis
p1:PhDThesis
p1:Unpublished
p1:Topic
p1:ResearchTopic

Στις εικόνες που ακολουθούν φαίνονται ενδεικτικά παραδείγματα παραγόμενων κειμένων από την οντολογία που αναφέρθηκε. Δυστυχώς δεν υπήρχε χρόνος για μια πιο συστηματική αξιολόγηση των κειμένων που παράγονται από την SWRC (π.χ. με ανθρώπους-αξιολογητές).



Εικόνα 21: Παράδειγμα παραγόμενου κειμένου από την οντολογία SWRC.



Εικόνα 22: Παράδειγμα παραγόμενου κειμένου από την οντολογία SWRC.

4 Συμπεράσματα

Σκοπός αυτής της εργασίας ήταν η βελτίωση του υπάρχοντος συστήματος παραγωγής φυσικής γλώσσας NaturalOWL. Οι κυριότερες βελτιώσεις επιτρέπουν:

- να μην εμφανίζονται στους τελικούς χρήστες μηνύματα λαθών που απευθύνονται στους χρήστες-συγγραφείς,
- να δημιουργούνται αυτόματα εγγραφές λεξικού κατά την εισαγωγή οντοτήτων τάξεων,
- να παράγονται λίστες κουκκίδων, όταν οι ιδιότητες έχουν πολλαπλές τιμές,
- να χωρίζονται οι ιδιότητες σε νοηματικές ενότητες, ώστε να αποφεύγονται αφύσικες συνενώσεις προτάσεων,
- να μην επαναλαμβάνονται ισοδύναμα γεγονότα, όταν εμπλέκονται αντίστροφες ιδιότητες,
- να παράγονται κείμενα με επισημειώσεις RDFa,
- να προστίθενται ευκολότερα νέοι τύποι χρηστών, αρχικά ως αντίγραφα υπαρχόντων τύπων,
- να εμφανίζονται ορθότερα οντότητες που αντιστοιχούν σε Multilingual Personalized Canned Texts.

Επίσης, εκτελέστηκαν δοκιμές με την οντολογία SWRC, με ικανοποιητικά αποτελέσματα, αν και δεν στάθηκε δυνατόν τα παραγόμενα κείμενα να αξιολογηθούν συστηματικά.

Υπάρχουν ακόμα πολλά περιθώρια βελτίωσης του συστήματος. Για παράδειγμα, θα ήταν πολύ χρήσιμο, αν το σύστημα παρήγαγε αυτόματα τις διάφορες πτώσεις των ουσιαστικών στις εγγραφές του λεξικού, εισάγοντάς του μόνο την ονομαστική του ενικού αριθμού. Υπάρχουν, επίσης, περιπτώσεις όπου τα ονόματα φυσικής γλώσσας των οντοτήτων περιέχονται στην ίδια την οντολογία ως τιμές ιδιοτήτων· θα ήταν χρήσιμο να παράγονταν σε αυτή την περίπτωση αυτόματα οι εγγραφές του λεξικού για αυτά τα ονόματα. Επίσης, θα ήταν ενδιαφέρον να δοκιμαστεί πιο συστηματικά το σύστημα και με άλλες οντολογίες.

Αναφορές

[1] Δ. Γαλάνης, "*Ανάπτυξη μηχανής παραγωγής φυσικής γλώσσας για οντολογίες OWL*", μεταπτυχιακή διπλωματική εργασία, Τμήμα Πληροφορικής, Οικονομικό Πανεπιστήμιο Αθηνών, 2006.

[2] Ontology Web Language (OWL) (www.w3.org/2004/OWL/)

[3] Isard, J. Oberlander, I. Ανδρουτσόπουλος και C. Matheson, "Speaking the Users' Languages". *IEEE Intelligent Systems*, 18(1):40-45, 2003.

[4] The Protégé Ontology Editor and Knowledge Acquisition System (http://protege.stanford.edu/)

[5]Γ. Καρακατσιώτης, "Αυτόματη παραγωγή συγκρίσεων σε ένα σύστημα παραγωγής κειμένων φυσικής γλώσσας", μεταπτυχιακή διπλωματική εργασία, Τμήμα Πληροφορικής, Οικονομικό Πανεπιστήμιο Αθηνών, 2007.

[6] http://ontoware.org/projects/swrc/ , The SWRC (Semantic Web for Research Communities) Ontology, Institute AIFB, University of Karlsruhe.

[7] Resource Description Framework - in – attributes(RDFa) (http://www.w3.org/TR/xhtml-rdfa-primer/)

[8]E. Reiter and R. Dale, *Building Natural Language Generation Systems*. Cambridge University Press, 2000.

[9]J. Oberlander, Γ. Καρακατσιώτης, A. Isard και I. Ανδρουτσόπουλος, "Building an Adaptive Museum Gallery in Second Life" (pre-print). Πρακτικά του *Museums and the Web*, Montreal, Quebec, Καναδάς, 2008.

[10]I. Ανδρουτσόπουλος, J. Oberlander και B. Καρκαλέτσης, "Source Authoring for Multilingual Generation of Personalised Object Descriptions". *Natural Language Engineering*, 13(3):191-233, Cambridge University Press, 2007.

[11] Grigoris Antoniou, Frank van Harmelen , "A Semantic Web Primer, 2nd Edition" (http://www.amazon.com/Semantic-Primer-Cooperative-Information-Systems/dp/0262012421/ref=sr_1_1?ie=UTF8&s=books&qid=1254745936&sr=8-1)