



# «Ανάπτυξη μηχανής παραγωγής φυσικής γλώσσας για οντολογίες OWL»

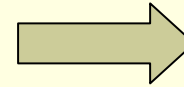
Διπλωματική εργασία ΜΠΣ «Επιστήμη Υπολογιστών»  
Γαλάνης Δημήτριος

Επιβλέπων: Ι. Ανδρουτσόπουλος  
Δεύτερος Αξιολογητής: Π. Κωνσταντόπουλος

# Παραγωγή κειμένων από οντολογίες OWL

## Οντότητα

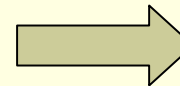
```
<Chianti rdf:ID="ChiantiClassico">  
  <hasBody rdf:resource="#Medium" />  
  <hasMaker rdf:resource="#McGuinnesso" />  
</Chianti>
```



Chianti Classico is a Chianti. It has medium body and it is made by Mc Guinnesso.

## Τάξη

```
<owl:Class rdf:ID="Burgundy">  
<owl:intersectionOf rdf:parseType="Collection">  
  <owl:Class rdf:about="#Wine" />  
  <owl:Restriction>  
    <owl:onProperty rdf:resource="#locatedIn" />  
    <owl:hasValue rdf:resource="#BourgogneRegion"/>  
  </owl:Restriction>  
</owl:intersectionOf>  
<rdfs:subClassOf>  
  <owl:Restriction>  
    <owl:onProperty rdf:resource="#hasSugar" />  
    <owl:hasValue rdf:resource="#Dry" />  
  </owl:Restriction>  
</rdfs:subClassOf>  
</owl:Class>
```



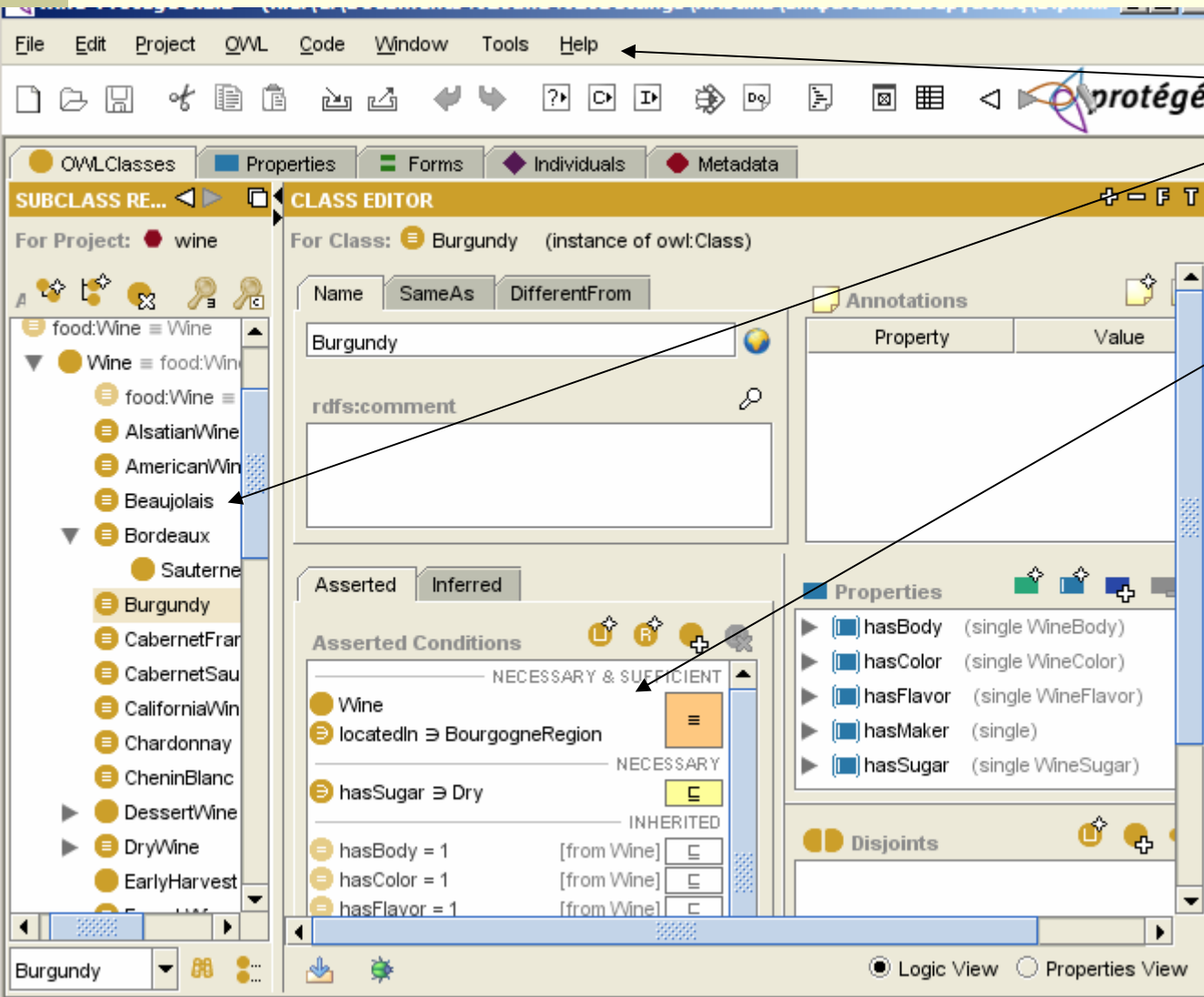
Burgundy is a wine. It is dry and it is produced in Bourgogne.

# Γιατί κείμενα από οντολογίες;

---

- Η απόδοση σε φυσική γλώσσα μπορεί να βοηθήσει στον εντοπισμό προβλημάτων.
  - Αν έχουμε τους ορισμούς των τάξεων και τις περιγραφές των οντοτήτων σε ΦΓ, μπορούμε να τους δείξουμε σε ειδικούς της γνωστικής περιοχής που δεν ξέρουν OWL.
- Εφαρμογές στο Σημασιολογικό Ιστό:
  - Οι ιστότοποι του ΣΙ (π.χ. καταστημάτων, μουσείων) μπορούν να αποθηκεύουν το περιεχόμενό τους σε OWL και να το παρουσιάζουν σε διάφορες φυσικές γλώσσες (σε ανθρώπους) ή σε λογική μορφή (σε τεχνητούς πράκτορες).

# Ορισμός τάξεων OWL στο Protégé

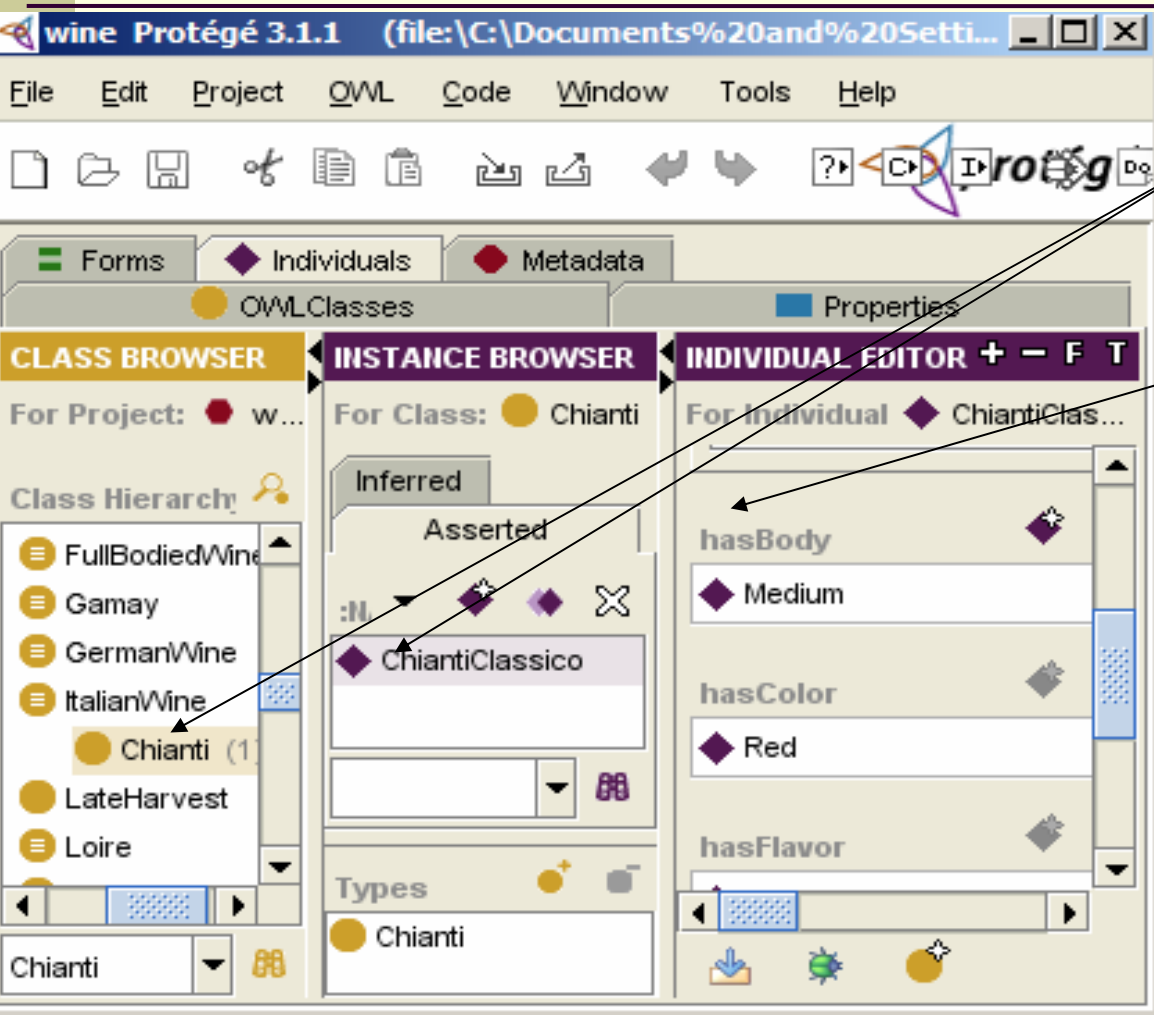


Protégé

Ορισμός  
ιεραρχιών  
τάξεων.

Ορισμός  
περιορισμών  
για τις τάξεις.

# Ορισμός οντοτήτων OWL στο Protégé



- Οντότητες:
  - Στιγμιότυπα τάξεων
  - Ορισμός τιμών για τις ιδιότητες των οντοτήτων

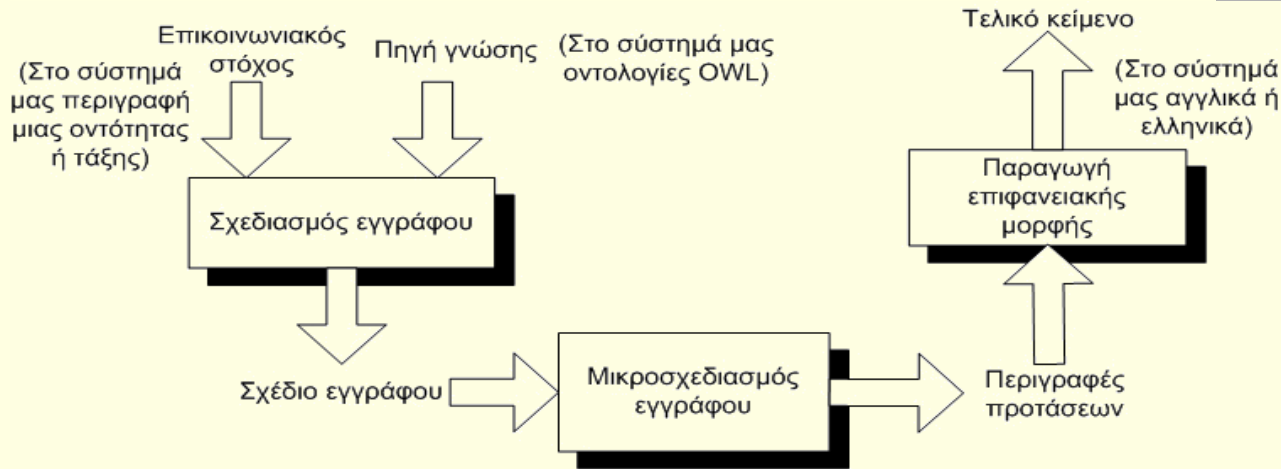
Το σύστημα της εργασίας συνεργάζεται και επεκτείνει το Protégé, όπως θα εξηγηθεί στη συνέχεια.

# Διάρθρωση της παρουσίασης

---

- Συνήθης αρχιτεκτονική συστημάτων ΠΦΓ.
- Η γλώσσα OWL.
- Το σύστημα παραγωγής φυσικής γλώσσας της εργασίας.
- Το «εργαλείο συγγραφής» και η σχέση με το Protégé.

# Αρχιτεκτονική συστημάτων ΠΦΓ.



- **Σχεδιασμός εγγράφου**
  - Επιλογή περιεχομένου (τι θα πούμε)
  - Καθορισμός δομής (με τι σειρά, οργάνωση σε παραγράφους, ενότητες κλπ )
- **Μικροσχεδιασμός εγγράφου**
  - Παραγωγή προδιαγραφών προτάσεων (π.χ. [ref(nominative, BancroftChardonnay), verb(active, present, «has»), ref(accusative, Moderate), «flavor»] )
  - Ομαδοποίηση προτάσεων (π.χ January was drier than average. January was colder than average =>January was drier and colder than average. )
  - Παραγωγή αναφορικών εκφράσεων (π.χ. «αυτό», «αυτό το έκθεμα», «αυτό το άγαλμα», «το προηγούμενο άγαλμα», «ο Δορυφόρος» )
- **Παραγωγή επιφανειακής μορφής**
  - Παραγωγή τελικού κειμένου (π.χ. BancroftChardonnay has moderate flavor)
  - Προσθήκη ετικετών HTML

# Η γλώσσα οντολογιών OWL (Web Ontology Language)

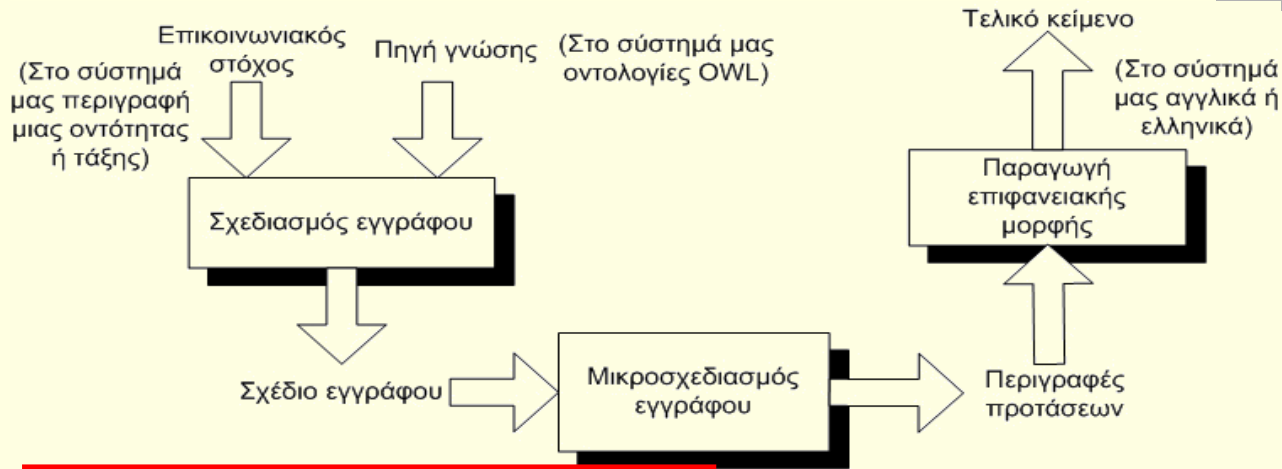
- Γλώσσα σήμανσης για τη δημιουργία και τη διανομή οντολογιών στο διαδίκτυο.
- Αναπτύχθηκε από το W3C.
- Αναπτύχθηκε ως επέκταση του λεξιλογίου της RDF.
- Ενσωμάτωσε εμπειρία από τη γλώσσα DAML+OIL.
- Παρέχει τρεις υπογλώσσες:
  - OWL Lite, OWL DL, OWL Full
  - $OWL\ Lite \subset OWL\ DL \subset OWL\ Full$
  - OWL Lite: ουσιαστικά απλές ταξινομίες
- **OWL DL**: αντιστοιχεί σε Descriptions Logics
  - υπολογιστικά πλήρης
  - αποφασίσιμη

Αυτήν  
υποστηρίζει  
η εργασία.

- IST M-PIRO: περιγραφές αντικειμένων από οντολογίες. Όμως ελλιπής υποστήριξη OWL.
- Αυτή η εργασία: παρόμοιο αλλά απλούστερο σύστημα που υποστηρίζει όμως εγγενώς την OWL.



# Αρχιτεκτονική συστημάτων ΠΦΓ.



- **Σχεδιασμός εγγράφου**
  - Επιλογή περιεχομένου (τι θα πούμε)
  - Καθορισμός δομής (με τι σειρά, οργάνωση σε παραγράφους, ενότητες κλπ )
- **Μικροσχεδιασμός εγγράφου**
  - Παραγωγή προδιαγραφών προτάσεων (π.χ. [ref(nominative, BancroftChardonnay), verb(active, present, «has»), ref(accusative, Moderate), «flavor»] )
  - Ομαδοποίηση προτάσεων (π.χ January was drier than average. January was colder than average =>January was drier and colder than average. )
  - Παραγωγή αναφορικών εκφράσεων (π.χ. «αυτό», «αυτό το έκθεμα», «αυτό το άγαλμα», «το προηγούμενο άγαλμα», «ο Δορυφόρος» )
- **Παραγωγή επιφανειακής μορφής**
  - Παραγωγή τελικού κειμένου (π.χ. BancroftChardonnay has moderate flavor)
  - Προσθήκη ετικετών HTML

# Επιλογή περιεχομένου

- Για οντότητες: επιλογή όλων των δηλώσεων  $X P Y$  και μια δήλωση που καθορίζει σε ποια τάξη ανήκει η  $X$ .

π.χ.

**BancroftChardonnay** locatedIn **NapaRegion**

...

**BancroftChardonnay** hasMaker **Bancroft**

**BancroftChardonnay** owl:type **Chardonnay**

Μηνύματα

Τιμή ιδιότητας

Ιδιότητα

Η οντότητα που περιγράφουμε

- Για τάξεις: επιλογή των δηλώσεων *subClassOf*, *equivalentClass*.

π.χ.

- PinotNoir equivalentClass (intersectionOf (Wine, madeFromGrape hasValue PinotNoirGrape, madeFromGrape maxCardinality 1))
- PinotNoir subClassOf (hasColor hasValue Red)

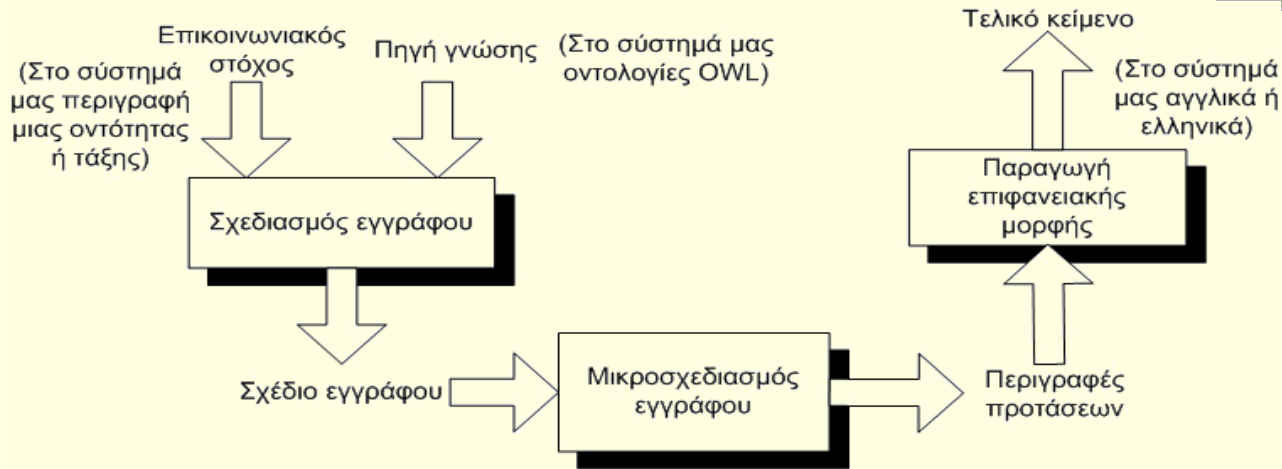
Υπο-  
μηνύματα

# Καθορισμός δομής

- Στην περίπτωση του συστήματος της εργασίας, απλά ταξινόμηση μηνυμάτων.
- Οντότητες:
  - Για κάθε ιδιότητα ορίζεται μια σειρά, δηλαδή ένας θετικός ακέραιος.
  - Το μήνυμα που δηλώνει σε ποια τάξη ανήκει η οντότητα έχει ως σειρά τον αριθμό 0.
- Τάξεις:
  - Τα μηνύματα αποτελούνται από υπο-μηνύματα.
  - Εξαγωγή των υπο-μηνυμάτων και ταξινόμηση όπως στην περίπτωση των οντοτήτων.

```
<?xml version="1.0" encoding="UTF-8"?>
<Ordering>
  <Order property="hasMaker">1</Order>
  <Order
    property="madeFromGrape">1</Order>
  .....
  <Order
    property="hasSugar">2</Order>
  <Order property="hasColor">3</Order>
</Ordering>
```

# Αρχιτεκτονική συστημάτων ΠΦΓ.



- Σχεδιασμός εγγράφου
  - Επιλογή περιεχομένου (τι θα πούμε)
  - Καθορισμός δομής (με τι σειρά, οργάνωση σε παραγράφους, ενότητες κλπ )
- Μικροσχεδιασμός εγγράφου
  - Παραγωγή προδιαγραφών προτάσεων (π.χ. [ref(nominative, BancroftChardonnay), verb(active, present, «has»), ref(accusative, Moderate), «flavor»] )
  - Ομαδοποίηση προτάσεων (π.χ January was drier than average. January was colder than average =>January was drier and colder than average. )
  - Παραγωγή αναφορικών εκφράσεων (π.χ. «αυτό», «αυτό το έκθεμα», «αυτό το άγαλμα», «το προηγούμενο άγαλμα», «ο Δορυφόρος» )
- Παραγωγή επιφανειακής μορφής
  - Παραγωγή τελικού κειμένου (π.χ. BancroftChardonnay has moderate flavor)
  - Προσθήκη ετικετών HTML

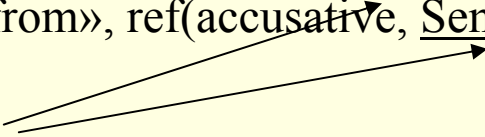
# Παραγωγή περιγραφών προτάσεων

- Για κάθε ιδιότητα  $P$  και κάθε γλώσσα, ορίζουμε τουλάχιστον ένα σχεδιάγραμμα που δείχνει πώς να εκφράσουμε μηνύματα  $X P Y$ .
- Σχεδιάγραμμα: μια σειρά από θέσεις (slots) που οι τιμές τους απλά συνδέονται για να σχηματίσουν μια πρόταση.
- Παράδειγμα σχεδιοτύπου για την ιδιότητα `madeFromGrape`:
  - Θέση 1: Αναφορική έκφραση που αναφέρεται στην οντότητα  $X$ , στην ονομαστική.
  - Θέση 2: Ο παθητικός ενεστωτικός ρηματικός τύπος «`is made`».
  - Θέση 3: Η συμβολοσειρά «`from`».
  - Θέση 4: Αναφορική έκφραση που αναφέρεται στην οντότητα  $Y$ , στην αιτιατική.
  - Θέση 5: Αναφορά στην τάξη `Grape`.

# Παραγωγή περιγραφών προτάσεων

- Τα σχεδιοτύπα για μηνύματα  $X P Y$  μπορούν σε κάθε θέση να περιέχουν:
  - Μια αναφορική έκφραση που αναφέρεται στην οντότητα  $X$ .
  - Την τιμή  $Y$  της ιδιότητας  $P$  στο μήνυμα ή μια αναφορική έκφραση στην οντότητα  $Y$ .
  - Ένα ρηματικό τύπο φυσικής γλώσσας του σχεδιοτύπου, συνοδευόμενο από πληροφορίες για τη φωνή και το χρόνο του.
  - Οποιαδήποτε άλλη συμβολοσειρά.
  - Αναφορά σε μια τάξη.
- Το `ChateauDYchemSauterne madeFromGrape SemillonGrape` με τη χρήση του παραπάνω σχεδιοτύπου γίνεται.

[ref(nominative, ChateauDYchemSauterne), verb(passive, present, «is made»), «from», ref(accusative, SemillonGrape), ref(Grape)]



ονόματα οντοτήτων που θα αντικατασταθούν από κατάλληλες αναφορικές προτάσεις

# Παραγωγή περιγραφών προτάσεων

- Μήνυμα X P Y της οντότητας ChateauDYchemSauterne :  
ChateauDYchemSauterne madeFromGrape SemillonGrape=>  
«Chateau DYchem Sauterne is made from Semillon grape»
- Τα ίδια σχεδίοτυπα χρησιμοποιούνται και για τα υπο-μηνύματα περιγραφών τάξεων.
- Περιορισμός hasValue της τάξης PinotNoir:  
madeFromGrape hasValue PinotNoirGrape => «Pinot Noir is made from Pinot Noir grape».
- Περιορισμός πληθαιρίθμου της τάξης PinotNoir :  
madeFromGrape maxCardinality 1 => «Pinot Noir is made from at most 1 grape»
- Περιορισμός allValuesFrom της τάξης giraffe:  
eats allValuesFrom leaf => «A giraffe eats only leaves»
- Ομοίως για τους περιορισμούς minCardinality, Cardinality, someValuesFrom.
- Αντίστοιχα κείμενα παράγονται στα ελληνικά.

# Λεξικό του συστήματος

- Κάθε τάξη ή οντότητα συσχετίζεται με ένα ουσιαστικό ή ονοματική φράση.
- Οι πληροφορίες που αποθηκεύονται για κάθε ουσιαστικό ή ονοματική φράση για τα ελληνικά είναι:
  - Ονομαστική, γενική και αιτιατική ενικού και πληθυντικού.
  - Το γένος.
  - Ο αριθμός που θα χρησιμοποιήσουμε για να αναφερθούμε στην οντότητα ή στην τάξη.
  - Μια μεταβλητή που δηλώνει αν το ουσιαστικό είναι μετρήσιμο.
- Αντίστοιχες πληροφορίες αποθηκεύονται για κάθε ουσιαστικό ή ονοματική φράση για τα αγγλικά.
  - Αποθηκεύουμε όμως μόνο την ονομαστική του ενικού και πληθυντικού.

```
<OntClass name="Wine">
  <NP gender="nonpersonal" num="singular">
    <singular>
      <nominative>κρασί</nominative>
      <genitive>κρασιού</genitive>
      <accusative>κρασί</accusative>
    </singular>
    <plural>
      ...
    </plural>
  </NP>
</OntClass>
```



# Ομαδοποίηση

- 1ος κανόνας ομαδοποίησης («περιγραφή τάξης και ιδιότητας»).
- Ομαδοποιεί δύο προτάσεις, η μια ορίζει την τάξη μιας οντότητας και η άλλη παρέχει την τιμή μιας ιδιότητας της οντότητας στην παθητική φωνή.

Π.χ

Bancroft Chardonnay is a Chardonnay. It is made in Bancroft .

=>

Bancroft Chardonnay is a Chardonnay made in Bancroft.

To Bancroft Chardonnay είναι ένα Chardonnay. Φτιάχνεται στο Bancroft.

=>

Το Bancroft Chardonnay είναι ένα Chardonnay που φτιάχνεται στο Bancroft.

ή

Το Bancroft Chardonnay είναι ένα Chardonnay το οποίο φτιάχνεται στο Bancroft.

# Ομαδοποίηση

- 2ος κανόνας ομαδοποίησης («απλή σύζευξη»): απλή σύζευξη δύο ή περισσότερων προτάσεων, όπου η κάθε μια παρέχει την τιμή κάποιας ιδιότητας της ίδιας οντότητας

Π.χ

It is dry. It has medium body.

=>

It is dry and has medium body.

- 3ος κανόνας ομαδοποίησης («κοινό υποκείμενο και ρήμα»):

Π.χ

It has medium body. It has moderate flavor.

=>

It has medium body and moderate flavor.

- Ο 3<sup>ος</sup> κανόνας έχει υψηλότερη προτεραιότητα από τον 2<sup>ο</sup>.
- Το σύστημα έχει συνολικά 5 κανόνες ομαδοποίησης, που προέρχονται από τη μεταπτυχιακή εργασία του Α. Μελέγκογλου (Εδιμβούργο).

# Παραγωγή αναφορικών εκφράσεων

- Στο σύστημα της εργασίας, αρκεί ο εξής απλοϊκός αλγόριθμος:
- Αν αναφερόμαστε σε μια οντότητα για πρώτη φορά, χρησιμοποιούμε το όνομα της ή μια δεικτική αντωνυμία
- Τις επόμενες φορές που αναφερόμαστε στην οντότητα, χρησιμοποιούμε μια κατάλληλη αντωνυμία στα αγγλικά ή κενό υποκείμενο στα ελληνικά.

π.χ

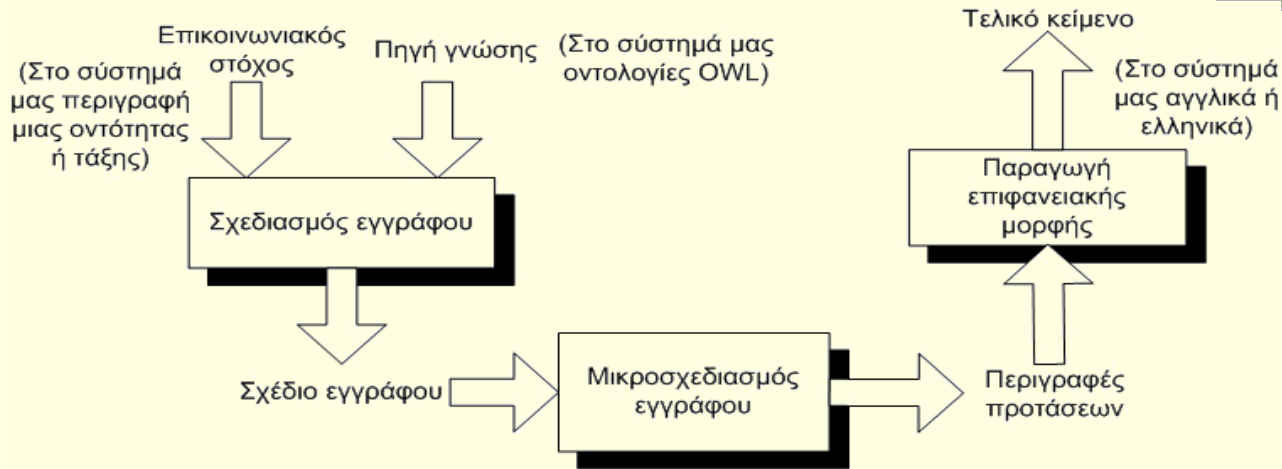
Bancroft Chardonnay is a Chardonnay made in Bancroft . It is dry and has medium body. It has moderate flavor and is produced in the Napa region.

ή

This is a Chardonnay made in Bancroft....

To Cabernet Sauvignon είναι ένα κρασί. Παράγεται από το πολύ 1 σταφύλι, το σταφύλι Cabernet Sauvignon. Είναι ξηρό και έχει κόκκινο χρώμα. Έχει μέτριο ή ισχυρό άρωμα και μέτρια ή πλήρη υφή.

# Αρχιτεκτονική συστημάτων ΠΦΓ.



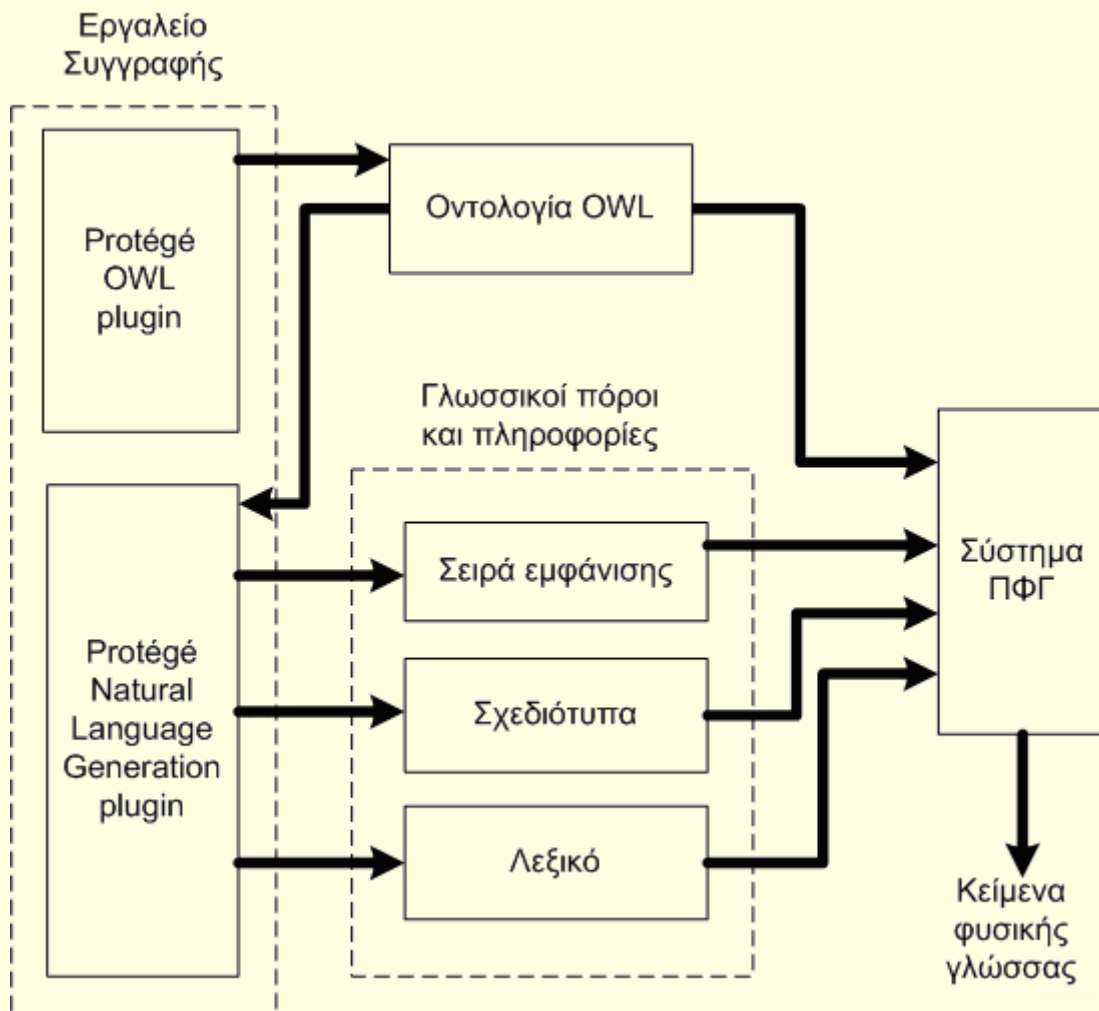
- Σχεδιασμός εγγράφου
  - Επιλογή περιεχομένου (τι θα πούμε)
  - Καθορισμός δομής (με τι σειρά, οργάνωση σε παραγράφους, ενότητες κλπ )
- Μικροσχεδιασμός εγγράφου
  - Παραγωγή προδιαγραφών προτάσεων (π.χ. [ref(nominative, BancroftChardonnay), verb(active, present, «has»), ref(accusative, Moderate), «flavor» ] )
  - Ομαδοποίηση προτάσεων (π.χ January was drier than average. January was colder than average =>January was drier and colder than average. )
  - Παραγωγή αναφορικών εκφράσεων (π.χ. «αυτό», «αυτό το έκθεμα», «αυτό το άγαλμα», «το προηγούμενο άγαλμα», «ο Δορυφόρος» )
- Παραγωγή επιφανειακής μορφής
  - Παραγωγή τελικού κειμένου (π.χ. BancroftChardonnay has moderate flavor)
  - Προσθήκη ετικετών HTML

# Παραγωγή επιφανειακής μορφής

---

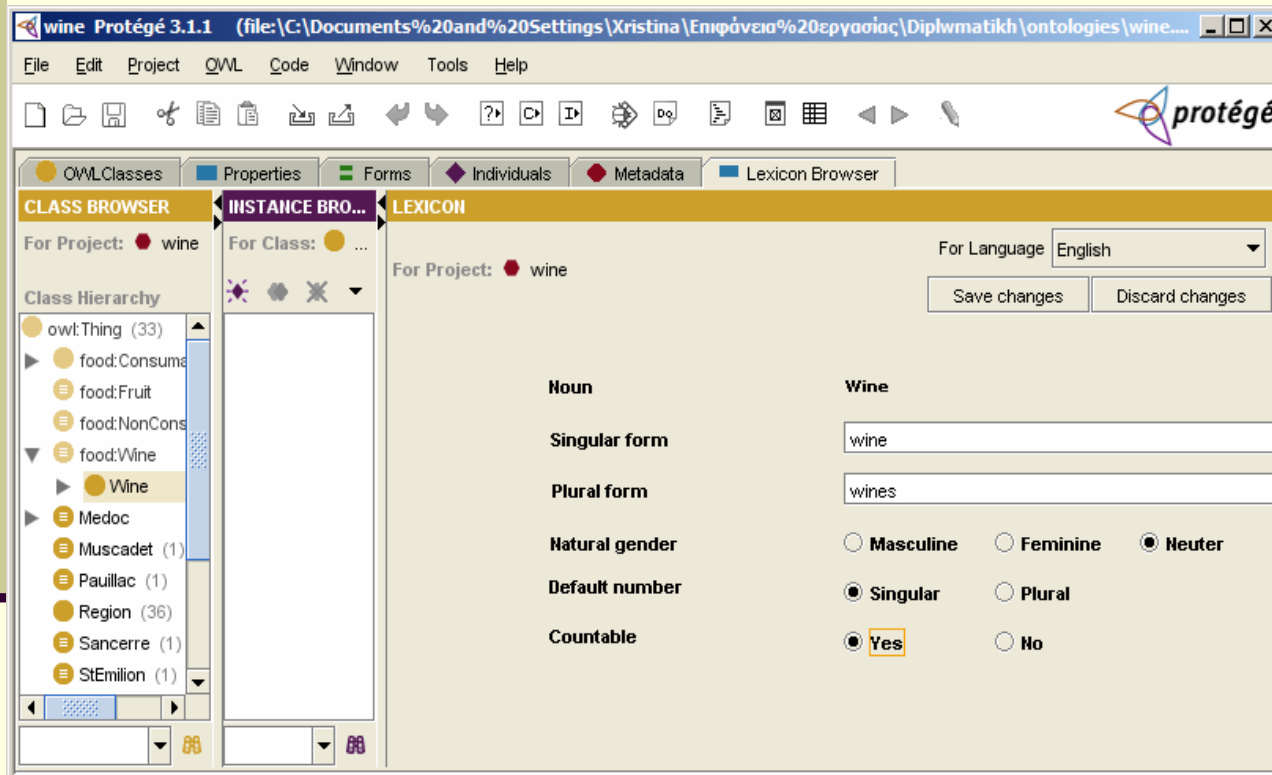
- Προσθήκη των κατάλληλων σημείων στίξης.
- Μετατροπή σε κεφαλαίο του πρώτου γράμματος κάθε πρότασης.
- Προσθήκη ετικετών HTML αν το κείμενο πρόκειται να παρουσιαστεί σε μια ιστοσελίδα.

# Εργαλείο συγγραφής



- Αντίστοιχο του «εργαλείου συγγραφής» του M-PIRO
- Το «εργαλείο συγγραφής» που δημιουργήθηκε βοηθά τον χρήστη να δημιουργούν τους γλωσσικούς πόρους που απαιτούνται (λεξικό, σχεδιάγραμμα, σειρά εμφάνισης).
- Το «εργαλείο συγγραφής» προστέθηκε ως επέκταση στο πρόγραμμα Protégé.

# Εργαλείο συγγραφής



- Φόρμα εισαγωγής πληροφοριών στο αγγλικό λεξικό

# Εργαλείο συγγραφής

The screenshot shows the Protégé software interface. The top menu bar includes File, Edit, Project, OWL, Code, Window, Tools, and Help. The toolbar contains various icons for file operations and editing. The main workspace is divided into several panels:

- PROPERTY BROWSER:** Shows a list of properties for the project 'wine', including adjacentRegion, food.course, food.hasDrink, food.hasFood, food.madeFromFruit, hasMaker, hasVintageYear, hasWineDescriptor, locatedIn, madeIntoWine, producesWine, and yearValue.
- ORDERING:** A table showing the order of properties for the project 'wine'. The table has two columns: 'property' and 'Order'. The properties and their orders are: hasMaker (5), madeFromGrape (3), hasSugar (2), hasColor (3), hasBody (5), hasFlavor (3), and locatedIn (4). Below the table are 'save' and 'sort' buttons.
- TEMPLATES:** A panel for defining templates. It shows 'For Property' set to 'adjacentRegion'. There are options for 'used' (unchecked), 'Language' (English), and 'Template' (1). A 'Save Template' button is present. Below this, there are radio buttons for 'property owner' (selected), 'property value', 'string', 'verb', and 'class descriptor'. A 'Case' dropdown menu is set to 'Nominative'. At the bottom, there are buttons for 'insert slot before', 'insert slot after', and 'delete slot'.

■ Ορισμός σειράς.

■ Ορισμός σχεδιοτύπων.



# Συμπεράσματα και μελλοντικές κατευθύνσεις

- Δημιουργήθηκε ένα σύστημα ΠΦΓ που παράγει κείμενα από συμβολικές πληροφορίες μια οντολογίας OWL.
- Παρόμοιο με του M-PIRO. Πιο απλά κείμενα, αλλά πλήρης υποστήριξη OWL DL.
  - Επίσης τα κείμενα περιγράφουν όχι μόνο οντότητες (όπως στο M-PIRO), αλλά και τάξεις.
- Συνοδεύεται από «εργαλείο συγγραφής», που υλοποιήθηκε ως επέκταση του Protégé.
- Πιθανές μελλοντικές επεκτάσεις:
  - Μοντέλα χρηστών
    - Επιλογή μηνυμάτων ανάλογα με τον τύπο αναγνώστη (π.χ. πιο τεχνικές πληροφορίες στους ειδικούς).
    - Επιλογή σχεδιότυπων ανάλογα με τον τύπο του αναγνώστη (π.χ. πιο απλές εκφράσεις σε παιδιά)
  - Ιστορικό αλληλεπίδρασης
    - Παραγωγή συγκρίσεων (π.χ. «Ο φορητός υπολογιστής που βλέπετε έχει τη μεγαλύτερη επεξεργαστική ισχύ από όλους που έχουν παρουσιαστεί έως τώρα»). Ξεκινά νέα διπλωματική εργασία.
    - Απορριπτή επαναλήψεων (πληροφορίες που έχουμε ήδη πει)