

κατηγοριοποίησης καλύπτει τις πιο συχνές περιπτώσεις ενώ αυτές που δεν καλύπτει εμφανίζονται πολύ σπάνια, τότε η επάρκειά του είναι πολύ μεγαλύτερη από 83%, μην μπορώντας ωστόσο να φτάσει στο 100%.

Λάθη στη γνώση

Ο έλεγχος αξιοπιστίας προσπαθεί να εντοπίσει λάθη στη γνώση που έχει ενσωματωθεί σε ένα σύστημα γνώσης. Τα λάθη αυτά μπορεί να χωριστούν σε δύο κατηγορίες:

- *Λάθη απόφασης (errors of commission)*: Συμβαίνουν όταν το σύστημα καταλήγει σε λάθος αποτέλεσμα για συγκεκριμένες τιμές εισόδου. Τα λάθη αυτά επηρεάζουν την ακρίβεια του συστήματος και είναι σχετικά εύκολο να διαπιστωθούν, αλλά συχνά δύσκολο να εντοπιστούν και να διορθωθούν.
- *Λάθη παράλειψης (errors of omission)*: Συμβαίνουν όταν το σύστημα δεν μπορεί να καταλήξει σε αποτέλεσμα για συγκεκριμένες τιμές εισόδου. Αυτό οφείλεται στο γεγονός ότι η γνώση που είναι απαραίτητη για να λυθεί κάποιο πρόβλημα παραλήφθηκε από τη βάση γνώσης του συστήματος. Τα λάθη αυτά επηρεάζουν την επάρκεια του συστήματος και διαπιστώνονται πολύ δύσκολα γιατί η υποδειγματική περίπτωση (test case) που θα αποκαλύψει την έλλειψη δεν είναι προφανής στο μηχανικό γνώσης (αλλιώς θα είχε ήδη αντιμετωπίσει την περίπτωση).

22.5 Μεθοδολογίες Ανάπτυξης Οντολογιών

Κατά την ανάπτυξη ενός συστήματος γνώσης πολλές φορές αποτελεί βασικό βήμα η ανάπτυξη της οντολογίας (ή των πλαισίων) που θα αναπαριστά την γνώση του συστήματος. Για την ανάπτυξη των οντολογιών έχουν προταθεί ειδικές μεθοδολογίες και η *τεχνολογία οντολογιών (ontology engineering)* αποτελεί μία πολύ ενεργή ερευνητικά υπο-περιοχή της τεχνολογίας γνώσης. Η τεχνολογία οντολογιών έχει πολύ μεγάλη σχέση με την επιστημονική περιοχή της *μοντελοποίησης εννοιών (conceptual modeling)* που αποτελεί υπο-περιοχή των Βάσεων Δεδομένων αλλά και της Τεχνητής Νοημοσύνης.

Τα τελευταία χρόνια έχουν δημιουργηθεί αρκετές μεθοδολογίες ανάπτυξης οντολογιών, όπως είναι οι DOGMA, ONTOCLEAN, METHONTOLOGY, ON-TO-KNOWLEDGE, DILIGENT, NEON, πλην όμως αποτελούν ιδιαίτερα πολύπλοκες μεθοδολογικές προσεγγίσεις και η παρουσίασή τους ξεφεύγει από τους σκοπούς του βιβλίου. Παρόλα αυτά υπάρχει μία σχετικά απλή και ιδιαίτερα δημοφιλής μεθοδολογία, η *Ontology Development 101*, η οποία θα παρουσιαστεί στη συνέχεια.

Οι βασικές αρχές της μεθοδολογίας 101 είναι:

- Δεν υπάρχει ένας και μοναδικός τρόπος να μοντελοποιηθεί ένα πεδίο γνώσης, καθώς πάντα υπάρχουν εναλλακτικοί τρόποι, εξίσου ορθοί. Ο "καλύτερος" τρόπος εξαρτάται από την εφαρμογή στην οποία θα χρησιμοποιηθεί η οντολογία ή από την επεκτασιμότητα που αναμένεται στο μέλλον για την οντολογία.
- Η ανάπτυξη μιας οντολογίας είναι μια επαναληπτική διαδικασία και η χρήση της σε εφαρμογές οδηγεί συνήθως σε επανασχεδιασμό.
- Οι έννοιες μιας οντολογίας θα πρέπει να ανταποκρίνονται σε αντικείμενα (φυσικά ή λογικά) του πεδίου γνώσης και στις σχέσεις μεταξύ τους. Συνήθως, στις

προτάσεις που περιγράφουν το πεδίο της γνώσης, τα ουσιαστικά αντιπροσωπεύουν τα αντικείμενα, ενώ τα ρήματα αντιπροσωπεύουν τις σχέσεις μεταξύ τους.

Σύμφωνα με την μεθοδολογία 101, τα βασικά στάδια στην ανάπτυξη μιας οντολογίας είναι τα ακόλουθα:

- Καθορισμός εμβέλειας.
- Επαναχρησιμοποίηση.
- Απαρίθμηση όρων.
- Ορισμός κλάσεων και ταξινόμιας.
- Ορισμός ιδιοτήτων.
- Ορισμός όψεων.
- Ορισμός στιγμιότυπων.

Όπως κάθε διαδικασία ανάπτυξης συστημάτων γνώσης, η παραπάνω διαδικασία δεν είναι γραμμική στην πράξη, καθώς τα βήματα μπορεί χρειαστεί να επαναληφθούν και η επιστροφή σε προηγούμενα βήματα ενδέχεται να είναι απαραίτητη σε οποιοδήποτε σημείο της διαδικασίας. Στη συνέχεια θα παρουσιαστούν με λεπτομέρεια τα βήματα της μεθοδολογίας.

22.5.1 Καθορισμός εμβέλειας

Τα βασικά ερωτήματα που πρέπει να απαντηθούν σε αυτό το στάδιο είναι:

- Ποιο είναι το πεδίο γνώσης που θα καλύψει η οντολογία;
- Για ποιους σκοπούς ή εφαρμογές θα χρησιμοποιηθεί η οντολογία;
- Για ποιους τύπους ερωτήσεων θα πρέπει η οντολογία να παρέχει απαντήσεις;
- Ποιοι χρήστες θα χρησιμοποιούν και θα συντηρούν την οντολογία;

Στη συνέχεια, ως παράδειγμα μοντελοποίησης με οντολογία θα χρησιμοποιηθεί γνώση για κρασιά, φαγητά και συνδυασμούς τους, με σκοπό να χρησιμοποιηθεί σε εφαρμογές οι οποίες προτείνουν καλούς - κατάλληλους συνδυασμούς τους, για παράδειγμα σε ένα εστιατόριο. Για μία τέτοια οντολογία θα χρειαζόταν η αναπαράσταση των τύπων κρασιών και φαγητών, οι έννοιες καλών και κακών συνδυασμών τους, κλπ., ενώ αντίθετα δεν θα χρειαζόταν έννοιες που σχετίζονται με τη διαχείριση της αποθήκης κρασιών ή των υπαλλήλων του εστιατορίου, αν και στον πραγματικό κόσμο έχουν σχέση με τις έννοιες των κρασιών και των φαγητών.

Σε μία άλλη εφαρμογή, η οποία θα χρησιμοποιούσε την οντολογία για την επεξεργασία φυσικής γλώσσας άρθρων σε περιοδικά ή ιστοσελίδες για κρασιά, θα έπρεπε να συμπεριληφθούν συνώνυμες λέξεις με τις έννοιες της οντολογίας, πληροφορίες για τα μέρη του λόγου στα οποία ανήκουν οι έννοιες, κλπ. Αν η οντολογία επρόκειτο να χρησιμοποιηθεί για μία εφαρμογή παροχής βοήθειας σε πελάτες εστιατορίου προκειμένου να παραγγείλουν κρασί, θα έπρεπε να ενσωματωθούν και πληροφορίες για τις τιμές των κρασιών. Αν η οντολογία επρόκειτο να χρησιμοποιηθεί για παραγγελίες σε κάβες, τότε θα έπρεπε να ενσωματωθούν έννοιες όπως οι τιμές χονδρικής και η διαθεσιμότητα της κάβας αλλά και τα αποθέματα της αποθήκης, κλπ. Ενδεχομένως σε κάποιες περιπτώσεις

τώσεις να χρειαστεί πολυγλωσσικότητα, αν η οντολογία θα πρέπει να χρησιμοποιηθεί σε διαφορετική γλώσσα από αυτήν που αναπτύχθηκε.

Από όλα τα παραπάνω φαίνεται ότι η εμβέλεια της οντολογίας εξαρτάται άμεσα από την εφαρμογή ή το σύστημα γνώσης στο οποίο θα ενσωματωθεί. Εναλλακτικά, η διαπίστωση της εμβέλειας μπορεί να γίνει με την βοήθεια των *ερωτήσεων επάρκειας (competency questions)* στις οποίες θα πρέπει να είναι ικανή να "απαντήσει" η οντολογία και οι οποίες καθορίζονται από τον μηχανικό γνώσης σε συνεργασία με τους ειδικούς του τομέα. Όταν αναπτυχθεί η οντολογία, οι ερωτήσεις επάρκειας μπορούν να χρησιμοποιηθούν για την αξιολόγησή της ως προς την πληρότητα της, δηλαδή αν η οντολογία περιέχει αρκετές πληροφορίες για να τις απαντήσει και αν οι απαντήσεις έχουν ικανοποιητικό επίπεδο λεπτομέρειας.

Για παράδειγμα, στην οντολογία των κρασιών-φαγητών κάποιες ερωτήσεις επάρκειας θα μπορούσαν να είναι οι ακόλουθες:

- Ποια χαρακτηριστικά των κρασιών είναι σημαντικά στην επιλογή;
- Το κρασί Bordeaux είναι λευκό ή κόκκινο;
- Το Cabernet Sauvignon ταιριάζει με θαλασσινά;
- Ποια είναι η καλύτερη επιλογή για ψητό κρέας;
- Ποια χαρακτηριστικά των κρασιών επηρεάζουν την καταλληλότητα;
- Το άρωμα ή το σώμα ενός κρασιού αλλάζει με την χρονιά;
- Ποιες ήταν οι καλές χρονιές για το κρασί Napa Zinfandel;

Για να απαντηθούν οι παραπάνω ερωτήσεις, η οντολογία θα πρέπει να περιέχει τους τύπους κρασιών, τα χαρακτηριστικά τους, τις χρονιές εσοδείας (και πιθανό χαρακτηρισμό τους σε καλές και κακές), την ταξινόμηση φαγητών που να μπορεί να βοηθήσει στην επιλογή κρασιού, καθώς και τους προτεινόμενους συνδυασμούς κρασιού-φαγητού.

22.5.2 Επαναχρησιμοποίηση

Οι οντολογίες πλέον είναι ευρύτερα διαθέσιμες καθώς αναπτύσσεται ο Σημαιολογικός Ιστός. Συνεπώς, είναι σπανίως απαραίτητο να ξεκινήσει κάποιος από το μηδέν κατά τον ορισμό μιας οντολογίας, καθώς υπάρχει σχεδόν πάντα μια διαθέσιμη οντολογία από κάποιον τρίτο, η οποία παρέχει τουλάχιστον μια χρήσιμη αφετηρία για την οντολογία που πρέπει να αναπτυχθεί. Ιδιαίτερα όταν προβλέπεται ότι το σύστημα γνώσης θα πρέπει να μπορεί ανταλλάσσει πληροφορίες με άλλες εφαρμογές στο διαδίκτυο, επιβάλλεται να επαναχρησιμοποιηθεί κάποια κοινά αποδεκτή οντολογία, τροποποιημένη ίσως, προκειμένου να αξιοποιηθεί η διαλειτουργικότητα που προσφέρει ο Σημαιολογικός Ιστός.

22.5.3 Απαρίθμηση Όρων

Αρχικά, για τον ορισμό της οντολογίας πρέπει να γίνει καταγραφή όλων των σχετικών όρων που αναμένεται να εμφανιστούν, δηλαδή να καθοριστεί για ποιους όρους θα έχει πληροφορίες η οντολογία, ποιες ιδιότητες θα έχουν αυτοί οι όροι και ποιες άλλες πληροφορίες και σχέσεις θα έπρεπε να καταγραφούν για τους όρους αυτούς. Οι όροι αυτοί