

Το κεφάλαιο 4¹ παρουσιάζει τα διαλογικά συστήματα διακρίνοντάς τα σε δύο κατηγορίες: task-oriented (διεκπεραίωση συγκεκριμένης εργασίας) chatbot (εκτενείς συζητήσεις, για διασκέδαση και για πρακτικούς σκοπούς). Για την ανάπτυξη τους απαιτείται η μελέτη του ανθρώπινου διαλόγου και των γνωρισμάτων του: αλληλουχία εκφωνημάτων, τερματισμός διαλόγου, διαλογικές πράξεις, αμοιβαία γνώση του κόσμου, πρωτοβουλία, αρχή της συνεργασίας, συνομιλιακό υπονόημα, συνεπαγωγές και συνάφεια. Δύο είναι επίσης και οι βασικές προσεγγίσεις αρχιτεκτονικών για chatbot: rule-based και corpus-based (αλλά και συνδυασμός τους σε υβριδικές εκδοχές συστημάτων). Για τα δεύτερα απαιτείται εκπαίδευση σε πολύ μεγάλα corpora διαλογικών δεδομένων (ταινίες, συζητήσεις – πραγματικές ή από κοινωνικά δίκτυα, crowdworkers, αποκρίσεις των χρηστών στο chatbot κ.ά.). Οι απαντήσεις τέτοιων συστημάτων δίνονται στον χρήστη είτε μέσω ανάκτησης (retrieval) είτε παραγωγής (NLG).

Όσον αφορά τα task-oriented συστήματα μια πολύ εμπορική αρχιτεκτονική είναι η GUS (frame-based). Το κάθε frame είναι μια δομή γνώσης που αποτελείται από slots τα οποία το σύστημα καλείται να συμπληρώσει επιτυχώς βάσει των αποκρίσεων του χρήστη στις ερωτήσεις του. Επιπλέον από τις ερωτήσεις που απευθύνει στον χρήστη προσπαθεί να εντοπίσει το domain του task καθώς και το intent του χρήστη. Ένα τέτοιο σύστημα συμπληρώνεται από ένα στοιχείο παραγωγής των αποκρίσεων του που βασίζεται σε προκαθορισμένα templates (template-based generation). Μπορεί επίσης να περιλαμβάνει ένα κομμάτι ASR.

Η αρχιτεκτονική dialogue-state έχει εμπλουτίσει την GUS με πιο πλούσιες αναπαραστάσεις καθώς περιλαμβάνει τα εξής στοιχεία: NLU (εξαγωγή domain, intent και slots, μέσω ενός sequence μοντέλου), dialogue state tracker (προσδιορισμός τρέχουσας κατάστασης διαλόγου και πιο πρόσφατης απόκρισης του χρήστη), dialogue policy (προσδιορισμός της κάθε επόμενης λειτουργίας του συστήματος και προσδιορισμός περιεχομένου απάντησης βάσει της τρέχουσας κατάστασης του διαλόγου) και NLG (εκπαίδευση ενός μοντέλου encoder-decoder σε corpus επισημειωμένων). Επίσης, κατά περίπτωση ενσωματώνονται στα στοιχεία του ASR και του TTS.

Ως προς την αξιολόγηση των διαλογικών συστημάτων σημειώνεται ότι τα chatbots δεν είναι δυνατό να αξιολογηθούν με αυτόματο τρόπο λόγω της μεγάλης πληθώρας πιθανών αποκρίσεων για κάθε εκφώνημα. Έτσι επιλέγονται αξιολογητές που συνομιλούν με το σύστημα και κατόπιν το βαθμολογούν ως προς κάποια χαρακτηριστικά του (π.χ. ευφράδεια, αποφυγή επαναλήψεων) ή επισημειωτές που κρίνουν το κείμενο μιας ολοκληρωμένης συνομιλίας και ανάλογα βαθμολογούν είτε την κάθε απόκριση ενός συστήματος ξεχωριστά είτε ένα σύστημα συνολικά έναντι ενός άλλου (acute-eval metric). Η αξιολόγηση των task-oriented συστημάτων μπορεί να γίνει είτε με απόλυτο τρόπο (έλεγχος εάν διεκπεραιώθηκε το task) είτε μέσω ευρεστικών μεθόδων που προσομοιάζουν με την μέτρηση της ικανοποίησης του χρήστη. Η πιο σημαντική από αυτές για frame-based αρχιτεκτονικές είναι το ποσοστό σφάλματος συμπλήρωσης slot (slot error rate), που υπολογίζεται ως ο λόγος του πλήθους των inserted/deleted/substituted slots προς το σύνολο των slots μιας πρότασης.

¹ Jurafsky, D. & Martin, J. (2021). Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. 3rd Edition, September 21, 2021d raft.

Ο σχεδιασμός των διαλογικών συστημάτων γίνεται πάντα με γνώμονα τον χρήστη. Έτσι η προσέγγιση για την ανάπτυξη τους είναι ανθρωποκεντρική (value sensitive design). Συγκεκριμένα, απαιτείται ενδελεχής έρευνα και μελέτη του task και των χρηστών, ανάπτυξη συστημάτων προσομοίωσης για τον έλεγχο μιας αρχιτεκτονικής πριν την ανάπτυξη του λογισμικού (π.χ. Wizard-of-Oz systems) και ένας κύκλος περιοδικών δοκιμών του ίδιου του συστήματος για την αξιολόγηση της λειτουργικότητάς του από τους χρήστες.