

TARTU ÜLIKOOL
MATEMAATIKA-INFORMAATIKATEADUSKOND
Arvutiteaduse instituut
Informaatika eriala

Tiit Käbin

**Eestikeelsete dialoogide kogumise
veebirakendus**

Bakalaureusetöö

Juhendas: prof. M. Koit

Autor “.....” juuni 2011

Juhendaja “.....“ juuni 2011

Lubada kaitsmisele

Professor “.....“ juuni 2011

TARTU 2011

Sisukord

Sissejuhatus	3
1. Võlur Ozi tehnika	4
2. Läbiviidud eksperimendid	5
3. Kasutatud tehnoloogiad ja keeled	6
4. Veebirakenduse ülesehitus	7
4.1 Kasutaja UI.....	8
4.2 Oz UI	12
5. Edasiarendamise võimalused.....	20
Kokkuvõte	21
Summary.....	22
Kasutatud materjalid.....	23
Lisad	24

Sissejuhatus

Töö eesmärgiks on luua veebirakendus, mille abil on võimalik koguda eestikeelseid dialooge. Kogutud dialoogide abil saab üles ehitada paremaid inimesega loomulikus keeles suhtlevaid süsteeme. Rakendus peaks olema kasutatav kõigis enam levinud veebilehitsejates.

Eestis on sarnaseid süsteeme loodud ka varem, näiteks Margus Treumuthi poolt rakendus Aivo, mille abil on võimalik koguda eestikeelseid dialooge. Sarnaselt minu loodavale süsteemile kasutab ka Aivo Võlur Ozi tehnikat. Võlur Ozi tehnika baseerub hüpoteesil, et inimene suhtleb arvutiga teistmoodi kui teise inimesega.

Dialoogide kogumiseks Võlur Ozi tehnikat kasutades püütakse jätta kasutajale mulje, et ta vestleb arvutiga, mille tulemusel saame teada, millist keelt kasutab inimene, kui ta suhtleb arvutiga. Tegelikult vastab kasutaja küsimusele teine inimene, kes on Võlur Ozi rollis. Hiljem muidugi paljastatakse pettus ja selgitatakse, miks sellist dialoogide kogumise viisi kasutati, ning küsitakse kasutaja nõusolekut, kas tema abil kogutud dialoogi võib kasutada uute dialoogsüsteemide loomiseks ning olemasolevate täiendamiseks.

Rakenduse loomisel on suurimaks valupunktiks rakendust kasutama hakkavate inimeste veenmine selles, et nad suhtlevad arvutiga. Selleks tuleb loodavas süsteemis rõhku panna eesti keele õigekirja reeglitele. Lause algab suure tähega ja lõpeb lause lõpumärgiga. Samuti tuleb lähtuda põhimõttest, et arvuti (Võlur Ozi) väljastatavast tekstist poleks inimese sekkumine märgatav. Selleks peavad rakenduses olema olemas juba valmis kirjutatud lausemallid ja tüüpvastused, näiteks „Tere!“, „Palun täpsusta!“ ning „Kahjuks ma ei tunne sõna <siia tuleb tuleb kautaja poolt sisestatud mitte-eestikeelne sõna>.“.

Töö alguses annan lühikese sissejuhatuse antud temavaldkonda. Seejärel kirjeldan rakenduse loomiseks vajaminevaid tehnoloogiaid ja lõpetuseks ülevaade rakenduse ülesehitusest. Töö lisades on toodud loodud kood ja andmebaasi tabelite loomise laused.

1. Võlur Ozi tehnika

Selleks et luua rakendust, mis suhtleks inimesega loomulikus keeles, on vaja koguda arvuti ja inimese vahelisi vestlusi. Tavapärased inimeste-vahelised vestlused pole siin kohal abiks, kuna inimene kasutab arvutiga suheldes oletatavasti erinevat keelt, kui teise inimesega. Seega tuleb kasutada paremaid tehnikaid. Kui veel ei ole olemas inimesega loomulikus keeles suhtlevat dialoogsüsteemi, siis sobib dialoogide kogumiseks Võlur Ozi tehnika. Mõiste „Võlur Ozi tehnika“ võttis kasutusele 1980. aastatel psühholoogiadoktor J.F. Kelley [11].

Tehnika sobib väga hästi arvuti ja inimese vaheliste vestluste kogumiseks, sest dialoogi alguses öeldakse inimesele, et ta suhtleb arvutiprogrammiga, mis saab inimesest aru ja püüab tema esitatud küsimustele vastata. Selle tulemusel inimene kasutab eksperimendi jooksul arvatavasti lihtsamat keelt: lihtsamat sõnavara ja lihtsamaid lausekonstruktsioone. Tegelikult on inimese vestluspartneriks teine inimene, keda nimetataksegi Võlur Oziks. Võlur Ozi ülesandeks on inimese esitatud küsimustele vastata. Seejuures peab ta jätma mulje, et inimene suhtleb arvutiga: kasutama korrektset keelt ja lihtsaid lauseid. Samuti peaksid korduvate küsimuste vastused olema samad.

Eksperimendi lõpus muidugi teavitatakse inimest pettusest, et tegelikult oli tema vestluspartneriks teine inimene, mitte arvutiprogramm, nagu algselt väideti. Samuti küsitakse temalt nõusolekut, kas tema abil kogutud dialoogi võib kasutada inimesega loomulikus keeles suhtlevate rakenduste loomiseks ja täiendamiseks. Kui inimene nõusolekut ei anna, siis temaga salvestatud dialoog kustutatakse.

2. Läbiviidud eksperimendid

Kuigi maailmas on Võlur Ozi tehnika (WOZ-method) küllaltki laialdaselt levinud, siis Eestis on käesoleva ajani läbi viidud vaid kaks dialoogide kogumise eksperimenti, kus kasutati Võlur Ozi tehnikat. Nende käigus on kogutud ainult 95 dialoogi erinevatel teemadel, mida on ilmselgelt liiga vähe selleks, et saaks teha järeldusi dialoogide struktuuri ja kasutatud keele kohta. Seetõttu on pärsitud ka antud teemal uurimuste tegemine.

Esimene eksperiment, kus osales 11 inimest, viidi läbi 2001. aastal [1]. Eksperimendi käigus suudeti koguda 22 dialoogi, mille läbivaks teemaks oli reisiinfo. Dialoogide kogumiseks kasutati väga lihtsat veebilehte, mis koosnes teksti sisestus aknast ja Võluri tekstide kuvamiseks eraldatud alast. Ühe dialoogi kogumiseks kulus keskmiselt 20-25 minutit ja keskmine pikkus oli 20 dialoogiakti. Täheledatai seda, et dialoogide struktuur võrreldes suulise kõnega oli lihtsam ning inimene ei jätnud hüvasti, vaid sulges lihtsalt rakenduse.

Teine eksperimentide seeria, kus osales 35 inimest, viidi läbi 2009. aastal [2]. Eksperimendi käigus paluti kasutajatel vestelda programmiga Aivo, mis oli väidetavalt suuteline suhtlema kasutajaga kirjalikus eesti keeles. Tegelikult vastas kasutaja küsimustele interneti vahendusel teine inimene. Sarnaselt minu loodud rakendusele kasutati ka siin dialoogide kogumiseks kahte veebilehte: kasutaja liides ja Võluri liides. Esimese liidese põhi-komponentideks on teksti sisestuse ja kuvamise võimalus ning lisaks on realiseeritud ka Võluri tekstide hääleline ettelugemine. Teine liides koosneb kolmest osast teksti sisestuse ja kuvamise ning vestluste valimise võimalusest. Oluliseks erinevuseks võrreldes käesolevas töös realiseeritud veebirakendusega oli nn. asünkroonne suhtlus, mis tähendas, et Võlur ei oodanud alati ära kasutaja sisendit, vaid alustas (või jätkas) info andmist. Dialooge koguti tegelikkuses veelgi rohkem, aga uurimuskõlblikuks loeti neist 75 dialoogi erinevatel teemadel. Teemad jagunesid järgnevalt: Tartu kinokavad (62), ilmateated (7), TV-kavad (4), lennuinfo (1) ja poliitika (1).

3. Kasutatud tehnoloogiad ja keeled

Bakalaureusetöös loodud süsteem on veebirakendus, seega kasutan rakenduse kujundamiseks veebiloomise keeli ja tehnoloogiaid.

HTML ehk HyperText Markup Language on hüperteksti märgendamiskeel [6], mille abil märgendan veebilehed.

CSS ehk Cascading Style Sheets on keel [6], mida kasutan HTML lehe kujunduse loomiseks.

JavaScript on brauseripoolne objektorienteeritud skriptimise keel, mida kasutan kasutajaliidese poolsete skriptide loomisel [7]. Seal hulgas on ka AJAXi [10] päringute saatmine, nende vastuste töötlemine ja kasutajatele kuvamine. AJAX ehk Asynchronous JavaScript And Extensible Markup Language on kogum omavahel seotud tehnikaid, mis on kasutuses rakenduse kliendi poolel, et luua interaktiivseid veebirakendusi. Arendajal on võimalik sooritada päringuid serverisse, ilma et peaks kogu lehte uuesti laadima. AJAXiga edastatud andmete formaadina kasutan JSONi[12]. JSON ehk JavaScript Object Notation on andmevahetuse formaat, mida on väga lihtne töödelda, sest igal andmeüksusel on võti ja väärtus. Väärtuseks saab olla omakorda jälle list, mis peab samuti koosnema andmeüksusest, kus on võti ja väärtus. Tänu sellele on võimalik väga lihtsalt üle võrgu saata keeruka struktuuriga andmeid.

JQuery on JavaScripti raamistik, mille abil on kasutajapoolseid skripte mugavam programmeerida [8].

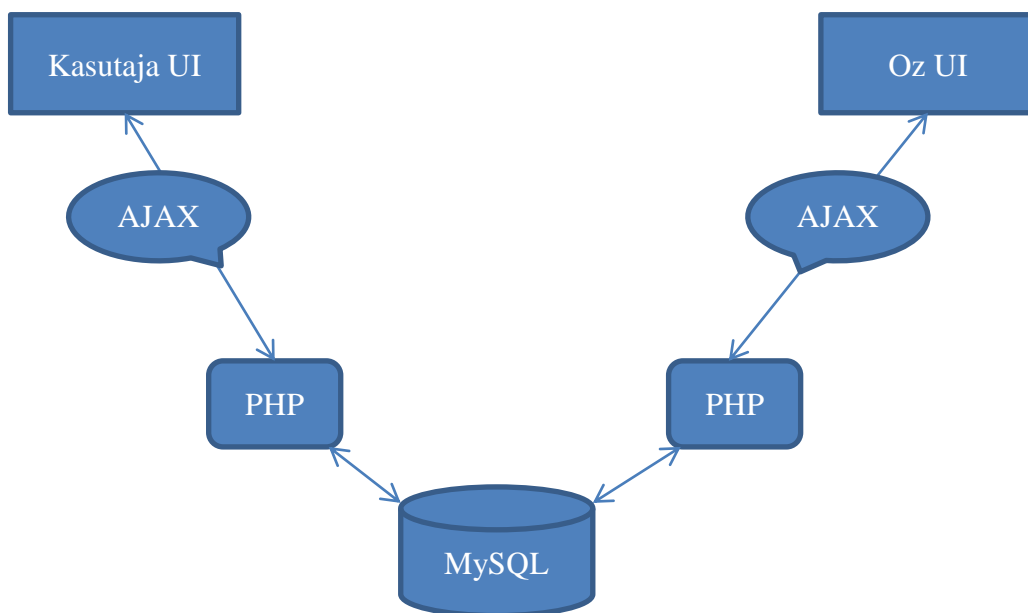
PHP on serveripoolne skriptimise keel, mille abil loon serveri poolseid lahendusi [5], sealhulgas suhtlus andmebaasiga ja saadud tulemuste AJAXile vajalikku formaati muutmine.

MySQL on avatud lähtekoodiga andmebaasi süsteem [9], mida kasutan vestluste hoidmiseks ja kasutajate info talletamiseks.

4. Veebirakenduse ülesehitus

Rakendusel on kaht tüüpi kasutajaid, kes esindavad ühte rolli kahest võimalikust. Rollid jagunevad järgmiselt: kasutaja, kes osaleb eksperimendis kui katsealune, ja Võlur, kes on Ozi rollis. Kuna need rollid on vägagi erinevad ja rakendus peab pakkuma tuge mõlemale, on seetõttu loodud kaks erinevat kasutajaliidest, kasutajapoolne liides (edaspidi Kasutaja UI [3]; UI – User Interface) ja Võlur Ozi liides (edaspidi Oz UI [4]). Liidesed mõneti küll kattuvad, aga pakuvad lisafunktsionaalsust lähtuvalt kasutaja rollist.

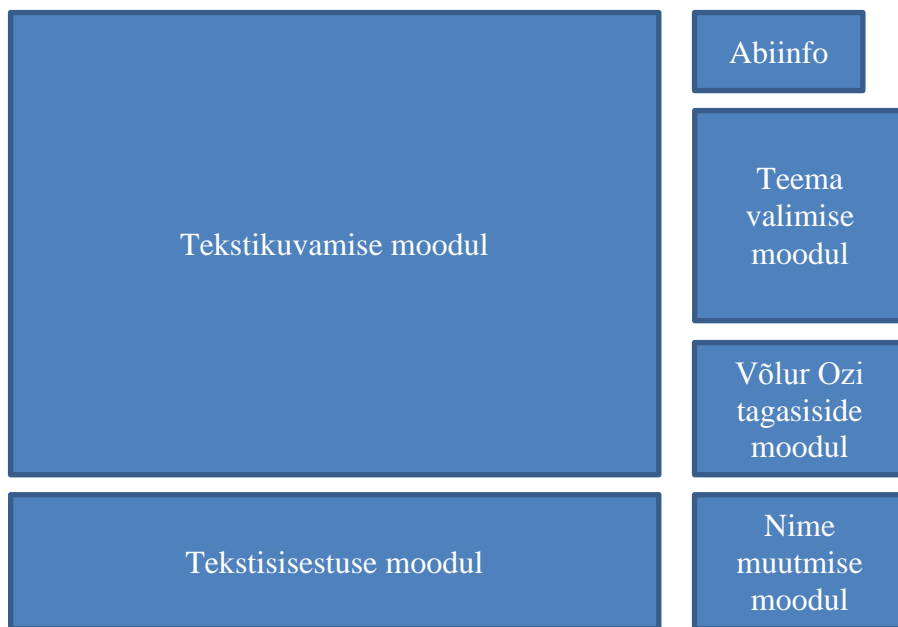
Kahe erineva kasutajaliidese (Kasutaja UI ja Oz UI) pidepunktiks on andmebaas. Siin kasutan MySQL andmebaasi. Andmebaasi tabelite loomise laused on lisas 2. Andmebaasiga suhtleb PHP, millele saavad käsud kasutajaliidestelt, kasutades tehnoloogiat AJAX. Joonisel 1 on kujutatud kahe erineva liidese omavahelise ühenduse skeem.



Joonis 1. Kahe liidese omavahelise ühenduse skeem

4.1 Kasutaja UI

Kasutaja rolli jaoks loodud liides on tavaline jututoa-laadne rakendus, mis koosneb seitsmest moodulist, milledest kuus on kasutajale nähtavad. Moodulite omavahelise paiknemise skeem asub joonisel 2.



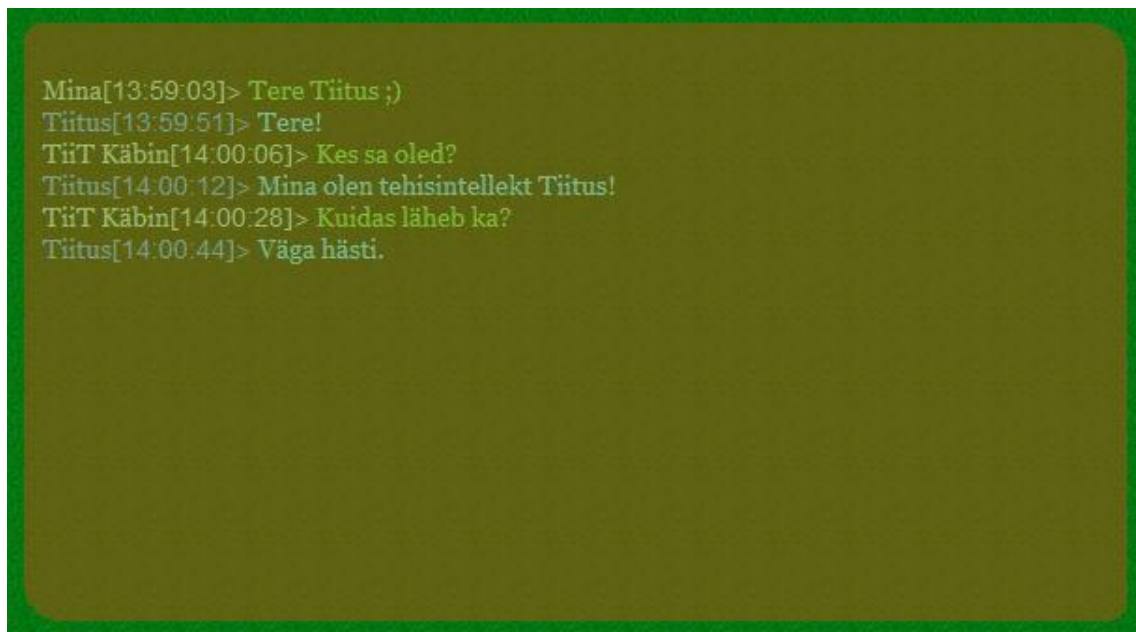
Joonis 2. Kasutaja UI moodulite paiknemise skeem

Üheks tähtsamaks osaks on tekstisisestuse moodul (vt. joonis 3). Sisestatud teksti sümbolitele pole seatud mingeid muid piiranguid peale maksimaalse sümbolite hulga, mis on 100 tähemärki. Üldjuhul on sisestatud tekstiks üks küsimus, seega ei sea 100 sümboli piirmäär otseselt kasutajatele piiranguid, sest eeldatavate küsimuste pikkus peaks jääma umbes 40-60 sümboli piiresse. Vajutades nuppu „Lisa sõnum“ sooritab JavaScript AJAX päringu serveris asuvale PHP skriptile, mis peale PHP lisab saadud andmed andmebaasi. Kui teksti sisestuse kasti pole sisestatud ühtki sümbolit, siis nupu „Lisa sõnum“ vajutust ignoreeritakse. Andmed edastatakse serverile ka siis, kui kursor on teksti sisestuse lahtris ja kasutaja vajutab klaviatuuril nuppu „Enter“, mispeale viib JavaScript läbi samad toimingud, mis nupu „Lisa sõnum“ puhul sooritatakse.



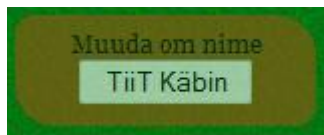
Joonis 3. Tekstisisestuse moodul

Teiseks tähtsamaks osaks on tekstikuvamise moodul. Siin kajastuvad mõlema kasutajaliidese tekstid. Igal tekstil on olemas autor, saatmise kellaeg ja tekst ise. Autorit tähistab sümbolite hulk, mis asub enne kellaega. Kõikide Võlur Ozi tekstide autoriks on märgitud „Tiitus“. Kasutaja algseks nimeks on „Mina“, aga see ei pea ilmtingimata nii olema, sest kasutajal on võimalik oma nime muuta. (Sellest täpsemalt veidike hiljem.) Samuti on mõlema rolli tekstid kuvatud erinevates värvitoonides, et lihtsustada tekstist arusaamist (vt. joonis 4). Teksti uuendamise eest hoolitseb JavaScript, mis sooritab iga kolme sekundi tagant AJAX päringu PHP-le, mispeale PHP teeb andmebaasi päringu, kus küsitakse käimasoleva vestlusega seotud uusi sõnumeid. Saadud sõnumid pannakse õigesse formaati (antud juhul JSON) ja saadetakse kliendi juurde. Päringu vastusega saabuvad sõnumid kuvab JavaScript kasutajale vastavalt sisestuse järjekorrale. Kui serverist tulev vastus on tühi ehk vahepeal uusi sõnumeid pole, siis Kasutaja UI-s mingeid toiminguid sellele ei järgne, vaid oodatakse kolm sekundit ja sooritatakse uus päring.



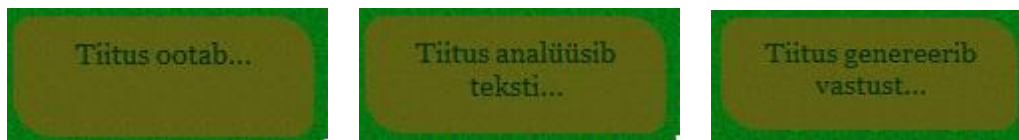
Joonis 4. Tekstikuvamise moodul

Nime muutmise moodul koosneb ainult ühest sisendlahtrist (joonis 5). Eraldi nime salvestusrežiimi pole vaja, sest nimi võetakse siis, kui kasutaja saadab uue sõnumi. Nimele pole samuti seatud muid piiranguid peale maksimaalse sümbolite hulga, mis on 15. See on piisavalt suur tähemärkide arv, sest eeldatavalt sisestab kasutaja ainult oma eesnime või siis mõne temale meelepärase hüüdnime.



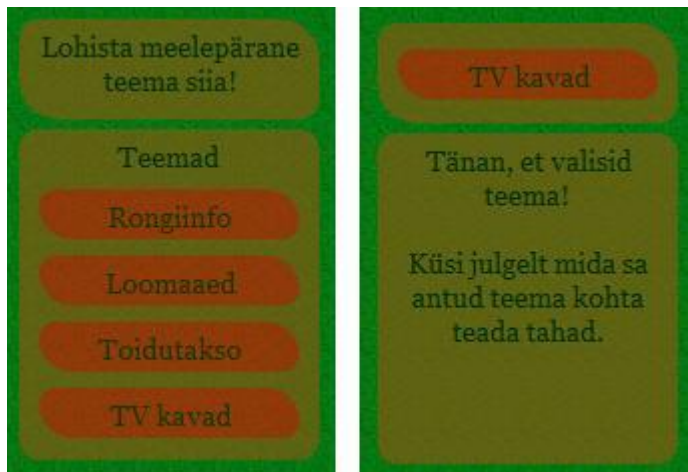
Joonis 5. Nimevahetusmoodul

Samuti on liidesesse sisse ehitatud ka Võlur Ozi tegevuste kuvamine, mis on tegelikkuses simulatsioon, sest reaalselt arvutiprogrammi, mis ise kasutaja küsimustele vastuseid leiaks, pole. Simulatsioon on vajalik, et jätta mulje Võlur Ozi olemasolust. On kolm erinevat võimalust. Esiteks, „Tiitus ootab...“ tähendab, et kasutaja pole veel teksti sisestanud, mida rakendus saaks töötlemata hakata. Teiseks, „Tiitus analüüsib teksti...“, mille kuvamise pikkuseks on juhuarv, mis sõltub sisestatud teksti pikkusest ja jääb nelja kuni kümne sekundi piiridesse. (Juhuarvude suuruse sõltuvusest sisestatud teksti sümbolite hulgaga täpsemalt lisas 3.) Kolmandaks, „Tiitus genereerib vastust...“. Viimased kaks tähistavad vastavalt seda, kas süsteem püüab sisestatud tekstist aru saada või rakendus moodustab kasutaja küsimusele vastust (vt. joonis 6). Viimane tekst vahetatakse välja esimese vastu siis, kui Võlur Ozilt on vastus saanud ja kasutajale kuvatud.



Joonis 6. Võlur Ozi tagasiside moodul

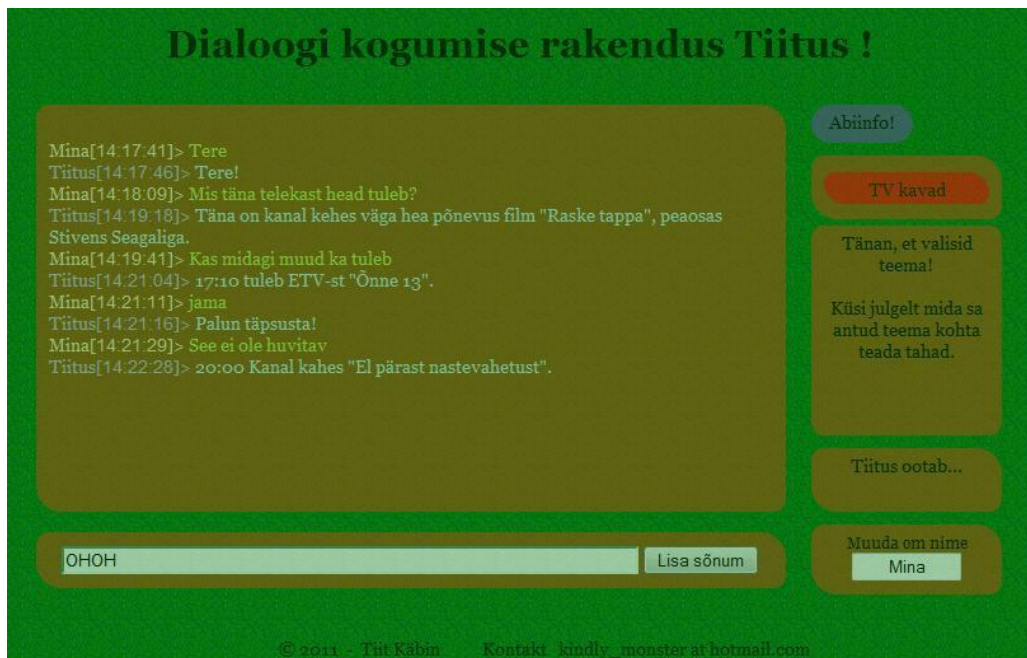
Lisaks on veel realiseeritud meelepärase teema valimine loetelust. Võimalikud valitavad teemad määrab eksperimendi läbiviija. Uusi teemasid luua ning varasemaid muuta ja kustutada saab Oz UI-s, millest täpsemalt peatükis 4.2. Maksimaalselt võib olla kuni neli teemat, mida kuvatakse kasutajale. Kui kasutaja liigutab hiire kursori teemale, siis kuvatakse talle ka antud teema lühikirjeldus. Teema valimiseks lohistab kasutaja meelepärase teema teemade nimestiku kohal paiknevasse boksi, kus asub kiri „Lohista meelepärane teema siia!“ (vt. joonis 7). Iga dialoogi ajal saab teemat valida vaid üks kord. Kui teema on valitud, siis seda muuta enam ei saa, sest kogu dialoog peab olema samal teemal. Teema valik pole kasutaja jaoks tehtud kohustuslikuks, sest väljapakutud teemade hulgas ei pruugi alati olla kasutajale meelepärast. Kui kasutaja on teema valinud, tehakse märge andmebaasi ja tulemus kuvatakse Võlur Ozile järgmise teksti uuendusega.



Joonis 7. Teemavalimise moodul

Abiinfo mooduli ülesandeks on kasutajale süsteemi kasutamist hõlbustava informatsiooni andmine, näiteks kuhu tuleb sisestada tekst ja miks tuleks valida teema. Samuti sisaldab see lühikest tutvustust, mis rakendusega on tegu. Tutvustav info ei vasta 100-protsendiliselt tõele, sest me ei saa kasutajale veel öelda, et antud rakendus ei suuda ise temaga vestelda ning tema esitatud küsimustele iseseisvalt vastuseid leida.

Kasutaja UI kujundus on toodud joonisel 8 [3].



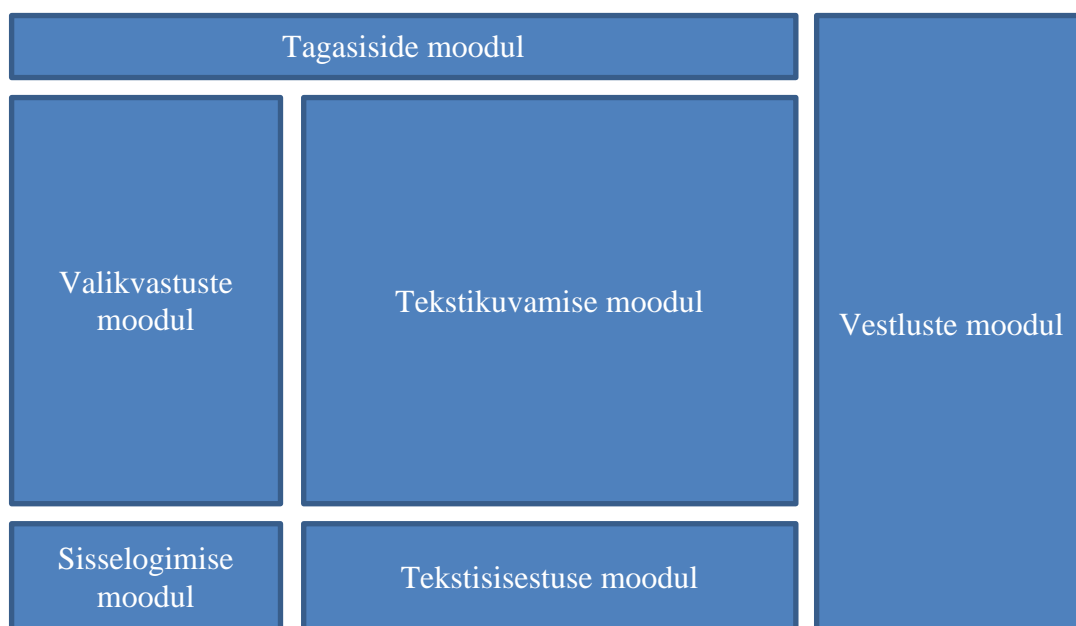
Joonis 8. Kasutaja UI

Lisaks eelmainitutele on Kasutaja UI-l veel olemas üks moodul „ozililm“. See moodul jääb kasutaja eest varjatuks, sest tegu on Võlur Ozile tagasiside andmisega tegeleva rakenduse osaga. Kui kasutajal on veebilehitseja avatud, siis iga tavapärase tekstikuvamise

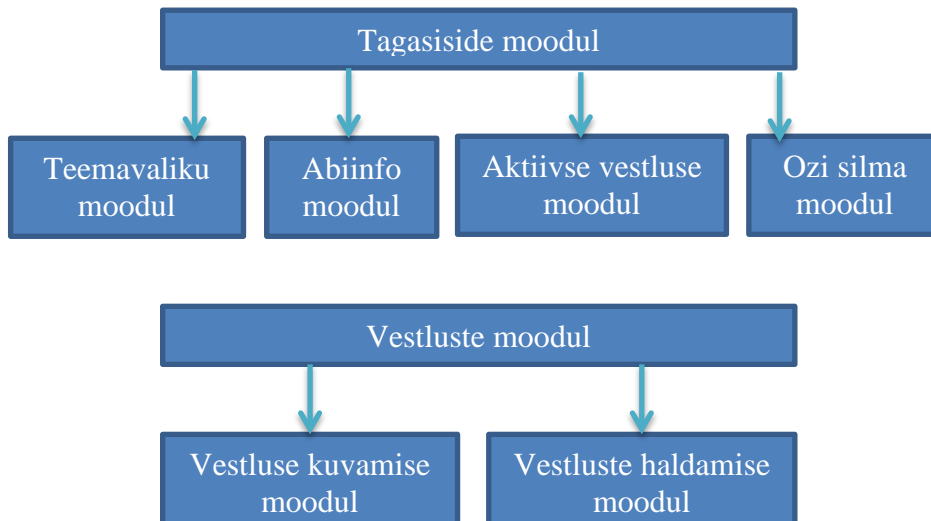
mooduli kuvamisega antakse AJAX päringuga PHP-le teada, et kasutaja pole veebilehelt lahkunud. Seepeale sooritab PHP päringu andmebaasile, kus uuendatakse kasutaja andmeid. Iga 2,5 sekundi tagant võrdleb JavaScript ka kasutaja sisestatud tekstisümbolite hulka eelmise kontrolli käigus saadud tulemusega: kui sümbolite hulk on erinev, siis antakse AJAX päringu abil PHP-le teada, et kasutaja on aktiivne. Kui aga sisestatud teksti sümbolite hulk on eelmise korraga võrreldes sama, siis antakse PHP-le teade ainult teatud intervallide tagant. (Kõigest lähemalt 4.2 Ozi silma mooduli kirjelduses.)

4.2 Oz UI

Võlur Ozi rolli jaoks loon mõnevõrra keerukama liidese, sest tegevusi, mida Võlur Oz peab saama teha, on palju rohkem. Sarnaselt eelkirjeldatud liidesele koosneb ka see liides kuuest põhimoodulist. Kaks moodulit on kasutatavad vaid sisse logitud kasutajatele. Moodulite omavahelise paiknemise skeem asub joonisel 10. Põhimoodulitest kaks jagunevad omakorda alammoduliteks, mille jagunemiste skeem on joonisel 11.

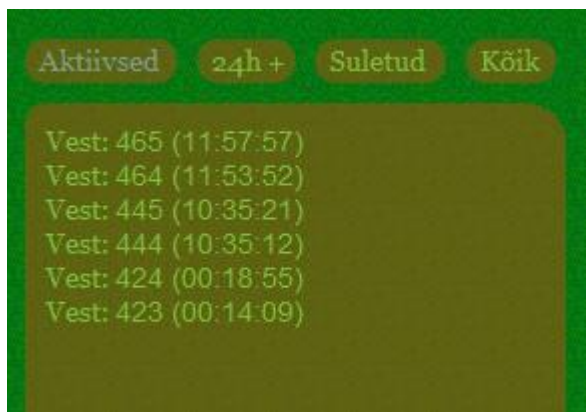


Joonis 10. Põhimoodulite omavahelise paiknemise skeem



Joonis 11. Põhimoodulite jagunemised

Kõige tähtsamaks ja ka üheks keerulisemaks mooduliks on vestluste moodul, mille ülesandeks on aktiivsete vestluste kuvamine ja andmebaasist dialoogide kättesaamise võimaldamine. Joonisel 12 on vestluste mooduli kujunduse näide.



Joonis 12. Vestluste mooduli kujundus (sisse logimata)

Vestluste kuvamise moodulis saab Võlur Oz valida vestlusi, milles ta soovib osaleda. Valitavate vestluste nimed moodustatakse sõnast „Vest:“ ja kahest andmebaasist tulevast komponendist, milledeks on vestluse *id* ning vestluse alustamise kellaeg. Kuna süsteemiga kogutavaid vestlusi koguneb palju, siis peab Võlur Oz saama olemasolevaid vestlusi sorteerida. Vestluste sorteerimiseks on olemas neli võimalust. Esiteks nupp „Aktiivsed“, millele vajutades kuvatakse vestlused, mis on loodud samal kuupäeval, mis reaalselt hetkel on, ja mille hulgast on välja arvatud need vestlused, mis on Võlur Ozi poolt

suletud. (Vestluste sulgemisest tuleb täpsemalt jutu vestluste haldamise moodulis.) Antud valik on rakenduse avamisel koheselt aktiveeritud. Teiseks võimaluseks on nupp „24h +“, millele vajutades kuvatakse vestlused, mille loomiskuupäev on erinev kehtivast kuupäevast. Ka ei kuvata siin sarnaselt eelmisega vestlusi, mis on juba suletud. Kolmandaks võimaluseks on nupp „Suletud“, millele vajutades kuvatakse vestlused, mis on Võlur Ozi poolt suletud. Neljandaks ja ühtlasi ka viimaseks sorteerimise võimaluseks on nupp „kõik“, millele vajutades kuvatakse kõik vestlused, mis asuvad andmebaasis.

Vestluste mooduli teiseks alammoduliks on vestluste haldamise moodul, mille kasutamiseks peab Võlur Oz sisselogima. (Sisse logimisest täpsemalt veidike hiljem.) Peale sisselogimist on Võlur Ozil võimalik olemasolevaid vestlusi sulgeda, kustutada ja välja võtta. Kõikide nende toimingute sooritamiseks ilmuvad peale sisselogimist, iga vestluse järele vastavad lingid, millele vajutades vastavalt kas vestlus suletakse, kustutakse, avatakse uuesti või alustatakse korpusekõlbliku variandi moodustamisega. Vestluste kustutamise ja väljavõtmise võimalust saab kasutada, kui Võlur Ozil on lahti vestluste kuvamise see osa, kus kuvatakse suletud vestlused. Siin on vestluste järel kolm linki: „Ava“, „Välja“ ja „Kustuta“. Need lingid on kättesaadavad ainult sisseloginud kasutajatele. Kui Võlur Oz vajutab linki „Kustuta“, siis küsitakse temalt üle, kas ta on kindel, et soovib vestlust tõesti kustutada. Kui Võlur Oz nõustub, siis vestlus andmebaasist kustutatakse. Kustutatud vestlusi Võlur Oz enam taastada ei saa, selleks tuleb pöörduda teenusepakkuja poole (praegusel hetkel on teenusepakkujaks planet.ee, mis kuulub firmale Zone Media OÜ[13]) ja paluda tal taastada mõni eelmistest seisudest. Selliste olukordade vähendamiseks on rakendusel olemas kahetasemeline õiguste süsteem (täpsemalt sellest sisselogimise mooduli kirjelduses). Peale lingi „Välja“ vajutamist küsitakse Võlurilt üle kaks asja. Esiteks, kui kasutaja on kasutanud mitut nime, siis küsitakse temalt, millist nime või nimesid kasutatakse. Teiseks on Võluril võimalik jätta kommentaar. Kui Võlur on need toimingud läbi viinud, siis moodustatakse HTML <div> olem, millesse kogu vestlus kuvatakse ja millest on Võlur Ozil võimalik dialoog tekstikujul kopeerida.

Vestluse valimiseks tuleb Võlur Ozil vajutada vestluse nime peale. Kui vestlus on valitud, siis aktiveerub Võluri jaoks tekstisisestuse moodul ja teksti kuvamise moodul täidetakse automaatselt valitud vestlusega. Selleks sooritab JavaScript AJAX päringu PHP. (Kõigest täpsemalt veidike hiljem.)

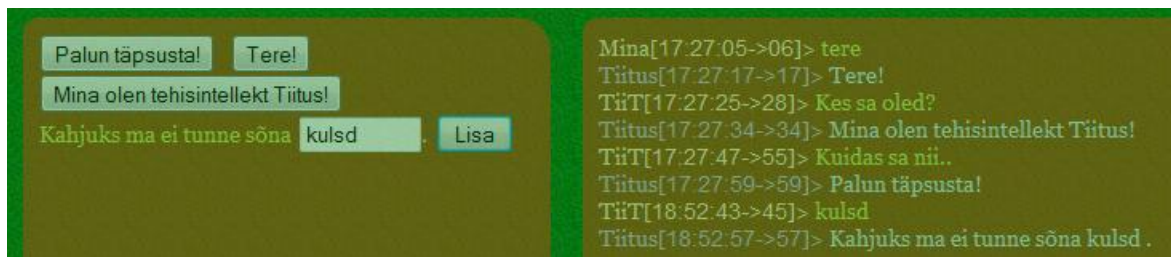
Tekstisisestuse moodulis saab Võlur sisestada Kasutaja UI-s tulnud küsimustele vastuseid. Sarnaselt esimesele liidesele on ka siin rakendatud tekstisisestusele sümbolite maksimaalse hulga piirang, milleks on 100 sümbolit. Tekstide lisamiseks saab Võlur Oz vajutada nuppu „Lisa sõnum“ või vajutada klaviatuuril nuppu „Enter“. Seepeale sooritab JavaScript AJAX päringu PHP skriptile, mis lisab Võluri lause andmebaasi. Moodulis on realiseeritud kaks õigekirjareeglit. Esiteks „lause lõpeb lauselõpumärgiga“: kui Võlur unustab lauselõpumärgi, siis lisatakse see tema eest. Kuna rakendus ei analüüsi lauseid, seega lause tüüpi määrata ei suuda, siis kasutatakse automaatselt lisatava lauselõpumärgina „punkti“. Teiseks realiseeritud õigekirjareegliks on „lause algab suure algustähega“: kui Võlur peaks selle reegli vastu eksima, siis muudetakse lause esimene täht suureks. Antud reegel rakendub vaid esimese lause puhul, see tähendab, et kui Võlur sisestab mitu lauset korraga, siis teisi lauseid ei kontrollita. Antud mooduli kujundus on sarnane Kasutaja UI-s olevaga (vt joonist 3).

Tekstikuvamise mooduli ülesandeks on mõlemast liideseast tulnud tekstide kuvamine. Igal tekstil on olemas autor, tekstisisestuse alguse ja lõpu kellaeg ning tekst ise. Selleks et mitte kuvada mõlemat kellaega terviklikult, kuvatakse Võlurile tekstisisestuse alguse kellaeg ja lõpu kellaajast ainult need numbrid, mis alguse kellaajast erinevad. Näiteks kui teksti sisestust alustati kell 17:27:47 ja lõpetati kell 17:27:55, siis kuvatakse Võlurile „17:27:47->55“ (vt joonist 13). Teksti parema loetavuse saavutamiseks on mõlema osapoole tekstid erinevat värvitooni. Valitud vestluse tekstide uuendamise eest kannab hoolt JavaScript, mis sooritab iga kolme sekundi tagant AJAX päringu PHP-le, kust saadakse vastuseks uued tekstid, mis on seotud antud vestlusega. Teksti lugemise hõlbustamiseks on erinevate kasutajate sisestatud tekstid kuvatud erinevates värvitoonides.

```
Mina[17:27:05->06]> tere
Tiitus[17:27:17->17]> Tere!
Tiit[17:27:25->28]> Kes sa oled?
Tiitus[17:27:34->34]> Mina olen tehisintellekt Tiitus!
Tiit[17:27:47->55]> Kuidas sa nii..
Tiitus[17:27:59->59]> Palun täpsusta!
```

Joonis 13. Tekstikuvamise moodul

Kuna Völur Ozil pole vastuse sisestamiseks palju aega, siis selleks on rakenduses olemas valikvastuste moodul. Siin on valmis ja pooleldi valmis vastused. Valmis vastuste aktiveerimiseks piisab Völuri ühest klikist meelepärasele vastusele, mispeale sooritab JavaScript AJAX päringu PHP-le ning vastus lisatakse kohe andmebaasi. Pooleldi valmis vastuste andmebaasi jõudmiseks kulub kaks sammu. Esmalt täidab Völur meelepärase vastuse sisendvälja. Seejärel vajutab vastusele järgnevat nuppu „Lisa“ (vt joonist 14), mispeale saadab JavaScript sisendist saadud teksti AJAX päringuga PHP-le, mis paneb vastuse kokku ja kirjutab saadud tulemuse andmebaasi.



Joonis 14. Valikvastuste moodul

Völur Ozile tagasiside andmiseks on süsteemi integreeritud tagasiside moodul, mis jaguneb omakorda neljaks alammoduliks. Esiteks, teemavaliku moodul, mis näitab ära, kas Kasutaja UI-s on kasutajal teema valitud või mitte. Kui teema on valitud, siis kuvatakse teema nimi Völur Ozile. Teema valimisest annab teada ka tekstikuvamise moodulisse ilmuv kasutajalt „RAKENDUS“ saabus sõnum, mille sisuks on teade selle kohta, et kasutaja on teema valinud ja millise. Teiseks, abiinfo moodul, mis avaneb, kui vajutada nuppu „Help!“. Siin on toodud juhised Völurile rakenduse kasutamiseks koos näidetega. Kolmandaks on aktiivse vestluse moodul, mis näitab, kas vestlus on valitud, ja kui on, siis milline vestlus on parasjagu aktiivne. Neljandaks mooduliks on „Ozi silm“, mille abil saab Völur Oz tagasisidet kasutaja liideses toimuva kohta. Tagasiside kätkeb endas infot selle kohta, kas kasutajal on veel brauser avatud, ja kui on avatud, siis kui kaua on möödas eelmisest teksti sisestusest. Selleks et vältida mooduli pidevat muutumist, kuvatakse völurile kaheksa erinevat teksti, mille tähendus on toodud tabelis 1. Tabelis esitatud ajad on uuenduste intervalliks ka Kasutaja UI Ozi silma juures. Tagasiside mooduli kujunduse näide on joonisel 15.

Tabel 1. Mooduli Ozi silm tekstide tähendused

Tekst	Seletus
Kasutaja kirjutab!	Kui eelmisest sümboli sisestuses on möödunud vähem kui 5 sekundit
Kasutaja mõtleb!	Kui eelmise sümboli sisestusest on möödunud rohkem kui 5 sekundit ja vähem kui 20 sekundit
Kasutaja on eemal!	Kui eelmise sümboli sisestusest on möödunud rohkem kui 20 sekundit ja vähem kui 1 minut
Kasutaja eemal juba 1 min!	Kui eelmise sümboli sisestusest on möödunud rohkem kui 1 minut ja vähem kui 2 minutit
Kasutaja eemal juba 2 min!	Kui eelmise sümboli sisestusest on möödunud rohkem kui 2 minutit ja vähem kui 4 minutit
Kasutaja eemal juba 4 min!	Kui eelmise sümboli sisestusest on möödunud rohkem kui 4 minutit ja vähem kui 10 minutit
Kasutaja eemal üle 10 min!	Kui eelmise sümboli sisestusest on möödunud rohkem kui 10 minutit
Kasutaja sulges brauseri!	Kui kasutaja sulges brauseri, viide on siin kuni 10 sekundit

Teema> Kasutajal on valik tegemata!

Help!

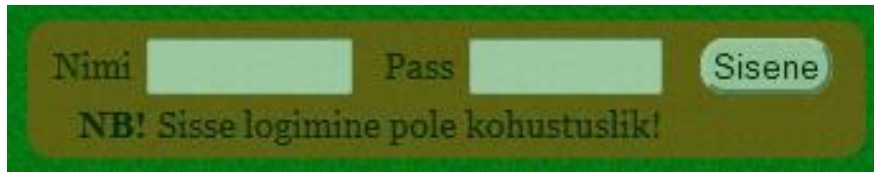
Aktiivne vestlus> Vest: 483 (17:26:48)

Silm> Kasutaja kirjutab!

Joonis 15. Tagasiside moodul

Samuti peab liides võimaldama eksperimendi läbiviijal kogutud dialooge andmebaasist kätte saada, märgendada ja vajadusel kommenteerida. Selleks on vajalik sisselogimise moodul. Peale enese autentimist saab Võlur olemasolevaid vestlusi sulgeda, kustutada, välja võtta ja uuesti avada. Rakendusse sisseloginud Võlurid on jagatud kahte gruppi. Esiteks, 1. taseme (level 1) Võlur Oz, kes saab kasutada kõiki võimalusi, välja arvatud vestluste kustutamine. Vestluste kustutamiseks peab Võlur Oz omama 2. taseme (level 2) staatust. Võlur autendib end kasutajanime ja parooliga, selleks on moodulis olemas vastavad väljad „Nimi“ ehk kasutajanimi ja „Pass“ ehk parool (vt joonist 16). Rakenduse kaitstud ossa sisenemiseks peab Võlur eelmainitud väljad täitma ja vajutama nuppu „Sisene“, mispeale kontrollitakse kasutajanime ja parooli omavahelist sobivust. Kuna puudub Võlurite registreerimise võimalus, siis uusi kasutajanimisid ja paroole saab ainult

käesoleva töö autorilt. Lisaks vestluste haldamisele on võimalik peale autentimist lisada kasutaja UI-s kuvatavaid teemasid ja neid kas aktiivseks või mitteaktiivseks muuta. Korruga võib olla ainult kuni neli aktiivset teemat, sest iga väljapakutud teemat peab Võlur Oz ka haldama ning tal peab olema juurdepääs ka info allikatele. Kuna teemad pakub välja konkreetne Võlur Oz, kes eksperimenti läbi viib, siis ei tegele rakendus dialoogide kogumise selle osaga, kust saada kasutaja küsimustele kiireid ja korrektseid vastuseid.



Joonis 16. Sisselogimise moodul

Teemade haldamiseks on nupp „Halda teemasid“, mis ilmub sisselogimise moodulisse peale autentimist. Näide teemahalduse moodulist on joonisel 17. Olemasolevate teemade juures saab muuta ainult nende aktiivsust ehk infot selle kohta kas seda teemat Kasutaja UI-s näidata või mitte. Uute teemade lisamiseks on nupp „Lisa uus teema“, mis asub teemade halduse liidese allosas. Igal teemal on olemas kolm omadust: nimi, lühikirjeldus ja aktiivsus. Kõiki kolme omadust saab uut teemat lisades määrata. Sisestatud tekstile on seatud ainult sümbolite hulga piirang, milleks on vastavalt nime puhul 20 sümbolit ja lühikirjelduse puhul 100 sümbolit.



Joonis 17. Teemahalduse moodul

Oz UI kujundus on toodud joonisel 18 [4].



Joonis 18. Oz UI kujundus

5. Edasiarendamise võimalused

Käesolevas töös loodud rakendusel on olemas ka mitmeid edasiarendamise võimalusi, mida minu bakalaureusetöö käigus ei käsitleta, nagu näiteks heli lindistus ja maha-mängimine või kasutaja veebikaamera pildi salvestus.

Samuti võiks rakendus kontrollida Võlur Ozi sisestatud teksti kõikidele eesti keele õigekirja reeglitele vastavust. Näiteks kui tehti viga, siis tõmmata vigasele kohale punane laineline joon alla, vähendamaks võimalust, et kasutaja tunneks ära teise inimese sekkumist vestluses.

Lisaks on võimalik välja töötada valikvastuste moodulisse vastuste grupid, mis sõltuksid kasutaja poolt valitud teemadest. Hetkel on valikvastuste moodulis tüüpvastused, mis kuvatakse kõikide valitud teemade puhul.

Kokkuvõte

Võlur Ozi tehnikat kasutades on eestikeelseid dialooge kogutud seni ainult 95, mida on liiga vähe selleks, et saaks teha järeldusi dialoogide struktuuri ja kasutatud keele kohta. Seetõttu on takistatud ka antud teemal uurimuste tegemine. Käesolevas töös loodud rakendus võimaldab mugavalt ja lihtsasti uusi dialooge koguda, mis peaks antud valdkonda Eestis elavdama.

Rakendusel on kaht tüüpi kasutajaid, kes esindavad ühte rolli kahest võimalikust. Kuna need rollid on vägagi erinevad ja rakendus peab pakkuma tuge mõlemale, on loodud kaks erinevat kasutajaliides: kasutajapoolne liides ja Võlur Ozi liides. Liidesed mõneti küll kattuvad, aga pakuvad lisafunktsionaalsust lähtuvalt kasutaja rollist. Mõlemad liidesed koosnevad omakorda moodulitest, millest igal on oma väga erinevad ülesanded. Kahe liidese ühenduspunktiks on andmebaas, s.t. kõik tegevused, mis liideses tehakse, salvestatakse andmebaasi ja kogu vajalik info liideste tööks saadakse andmebaasist. Andmete edastuseks ja vastuvõtuks kasutatakse tehnoloogiat AJAX, et vältida lehe pidevat uuesti laadimist.

Kuna loodud rakendus sisaldab mitmeid tehnilisi uuendusi võrreldes seni kasutatud Võlur Ozi liidestega, nagu näiteks valmis ja poolvalmis vastused, Ozi silm ning teema valik ja haldamine, siis see muudab süsteemi kasutamise mugavamaks ja kasutajasõbralikumaks.

Rakenduse edasiarenduse võimalusteks on helisalvestamise ja mahamängimise lisamine ning kasutaja veebikaamera pildi lindistus. Samuti on üheks edasiarenduse võimaluseks Võlur Ozi sisestatud tekstide õigekirjareeglitele vastavuse kontrollimine, vähendamaks võimalust, et kasutaja tunneks ära teise inimese sekkumise vestlusesse.

Estonian dialogues capture web application

Summary

Using the wizard-of-Oz technology, only 95 dialogues in the Estonian language have been accumulated so far; which is obviously too small a number to make any conclusions on the structure of the dialogues and the use of language. This is why research in this field has also been complicated. The application the current study represents allows gathering new dialogues with comfort and ease, which should enliven the sphere in question in Estonia.

The application is meant for two types of users who would play one of the two possible roles. As these roles are rather different and the application is to support both, two different user interfaces have been created: the user-end interface and the Wizard-of-Oz interface. The interfaces do overlap to a certain extent but offer additional functionalities depending on the role of the user. In their turn, both interfaces consist of modules, each fulfilling a separate task, and the latter vary considerably. The connection point of the two interfaces is the database, i.e., all the actions that are performed with the use of the interfaces are saved in the database, and all the information for the functioning of the interfaces is obtained from the database. To transfer and receive the data while preventing the page from reloading all the time, the AJAX technology is employed.

As the created application contains such numerous technological innovations, as, for example, complete and semi-complete answers, the Eye of Oz and its selection and administration, the system is more comfortable and user-friendly in comparison to the system used before; which, in its turn, is supposed to revive the interest in the use of the given system.

The opportunities for the further development of the system include voice recording and playback as well as recording the web-camera image of the user. Another opportunity for further development is the complete proofreading of the texts entered into the Wizard-of-Oz to check their compliance to rules in order to minimize the probability of the user recognizing the interference of other people in the conversation.

Kasutatud materjalid

1. M. Valdisoo, E. Vutt. "Võlur Ozi" tehnika ja eesti keeles suhtlev arvuti. A&A, 2002, nr 5
2. S. Pärkson. Võlur Ozi eksperimendid ja kogutud dialoogide analüüs, käsikiri 31.01.2011
3. Kasutaja UI <http://tiitus.planet.ee> (18.05.2011)
4. Oz UI <http://tiitus.planet.ee/ozzike> (18.05.2011)
5. PHP <http://php.net/manual/en/intro-what-is.php> (18.04.2011)
6. HTML ja CSS <http://www.w3.org/standards/webdesign/htmlcss> (18.04.2011)
7. JavaScript <http://www.devx.com/projectcool/Article/20066> (18.04.2011)
8. JQuery <http://jquery.com/> (18.04.2011)
9. MySQL <http://dev.mysql.com/doc/refman/5.0/en/what-is-mysql.html> (19.04.2011)
10. AJAX <http://derivadow.com/2007/01/05/ajax-what-is-it-its-not-dhtml/> (19.04.2011)
11. Võlur Oz <http://musicman.net/oz.html> (23.03.2011)
12. JSON <http://www.json.org/> (20.04.2011)
13. Zone Media OÜ <https://www.planet.ee/info> (23.05.2011)

Lisad

Lisa 1. Failid

Rakenduse kood, kõik loodud failid asuvad aadressil

<https://code.google.com/p/minu-elli/source/browse/#svn%2Felli>

Lisa 2. Andmebaasi loomise kirjed

Rakenduse kasutamiseks vajaminevate tabelite loomise MySQL päringud.

```
CREATE TABLE IF NOT EXISTS `dialoog_dialoog` (  
  `id` int(11) NOT NULL AUTO_INCREMENT,  
  `tekst` varchar(120) CHARACTER SET utf8 NOT NULL,  
  `kasutaja_id` int(6) NOT NULL,  
  `kes` int(1) NOT NULL,  
  `algus` timestamp NOT NULL DEFAULT '0000-00-00 00:00:00',  
  `lopp` timestamp NOT NULL DEFAULT '0000-00-00 00:00:00',  
  `nimi` varchar(20) CHARACTER SET utf8 DEFAULT NULL,  
  PRIMARY KEY (`id`)  
) ENGINE=MyISAM DEFAULT CHARSET=latin1 AUTO_INCREMENT=1 ;
```

```
CREATE TABLE IF NOT EXISTS `dialoog_kasutaja` (  
  `id` int(6) NOT NULL AUTO_INCREMENT,  
  `ip` varchar(20) NOT NULL,  
  `host` varchar(200) NOT NULL,  
  `saabus` timestamp NOT NULL DEFAULT '0000-00-00 00:00:00',  
  `status` int(1) NOT NULL DEFAULT '1',  
  `type` int(1) NOT NULL DEFAULT '0',  
  `oz_eye_time` timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP,  
  `last_mess_count` int(6) NOT NULL DEFAULT '0',  
  `teema_id` int(6) DEFAULT NULL,  
  PRIMARY KEY (`id`)  
) ENGINE=MyISAM DEFAULT CHARSET=latin1 AUTO_INCREMENT=1 ;
```

```
CREATE TABLE IF NOT EXISTS `dialoog_reguser` (  
  `id` int(6) NOT NULL AUTO_INCREMENT,  
  `name` varchar(25) CHARACTER SET utf8 NOT NULL,  
  `pass` varchar(25) CHARACTER SET utf8 NOT NULL,  
  `level` int(2) NOT NULL DEFAULT '1',  
  `email` varchar(200) CHARACTER SET utf8 NOT NULL,  
  PRIMARY KEY (`id`)  
) ENGINE=MyISAM DEFAULT CHARSET=latin1 AUTO_INCREMENT=1 ;
```

```

CREATE TABLE IF NOT EXISTS `dialog_teema` (
  `id` int(6) NOT NULL AUTO_INCREMENT,
  `status` int(2) NOT NULL,
  `name` varchar(20) CHARACTER SET utf8 NOT NULL,
  `description` varchar(100) CHARACTER SET utf8 NOT NULL,
  PRIMARY KEY (`id`)
) ENGINE=MyISAM DEFAULT CHARSET=latin1 AUTO_INCREMENT=1 ;

```

Lisa 3. Juhuarvude moodustamise tabel

Juhuarvu suuruse sõltuvus sisestatud teksti sümbolite hulgast.

Sisestatud teksti sümbolite hulk	Juhuarvu moodustamise vahemik
1 kuni 15 sümbolit	4 kuni 6 sekundit
16 kuni 30 sümbolit	5 kuni 7 sekundit
31 kuni 60 sümbolit	6 kuni 9 sekundit
61 kuni 100 sümbolit	8 kuni 10 sekundit