



GÖTEBORGS UNIVERSITET

“Can’t you read I’m frustrated?”

En sentimentanalytisk studie i människa-chatbotinteraktion

“Can’t you read I’m frustrated?”

A study in human-chatbot interaction

**Evelina Larsson
Stina Ljungkvist**

Kandidatuppsats i kognitionsvetenskap

Rapportnummer: 2017:128

“CAN’T YOU READ I’M FRUSTRATED?”

Förord

Denna uppsats är skriven vid Göteborgs universitet på Institutionen för tillämpad informationsteknologi, under vårterminen 2017. Stina har ansvarat för allt som relaterar till emotioner, och Evelina för allt som relaterar till chatbotar. Detta genomsyrar hela uppsatsen, och gäller allt från bakgrund till metod, i vilken Stina genomförde sentimentanalysen och Evelina simulerade chatbotinteraktionen. Med detta sagt tar båda fullt ansvar för hela uppsatsen, eftersom det i slutändan varit en gemensam iterativ process.

“CAN’T YOU READ I’M FRUSTRATED?”

Sammanfattning

För att en serviceorienterad chatbot ska kunna bidra till kognitiv avlastning för en användare är det viktigt att relationen mellan de två präglas av en viss nivå av tillit. Ett sätt att åstadkomma detta är genom emotionsdetektering i den text användaren lämnar ifrån sig vid interaktion. En passande respons på användarens emotionella tillstånd har visat sig förbättra relationen mellan användare och gränssnitt, särskilt i situationer då interaktionen inte är framgångsrik. Denna uppsats inleder med att redogöra för vad som orsakar negativa emotioner, dels generellt och dels specifikt inom människa-datorinteraktion. Genom att utföra sentiment- och ansiktsuttrycksanalys vid interaktion mellan användare och en prototyp av en chatbot undersöks därefter om negativa emotionella tillstånd är möjliga att utrona. Resultaten tyder på att negativa emotioner går att detektera, samt att det med fördel sker via sentimentanalys snarare än ansiktsuttrycksanalys.

Nyckelord: Sentimentanalys, chatbot, emotioner, människa-datorinteraktion.

Abstract

In order for a service-oriented chatbot to contribute to cognitive relief for a user, it is important that the relationship between the two is characterized by a certain level of trust. One way of achieving this is through emotion detection in the text the user delivers during interaction. An appropriate response to the user's emotional state has been shown to enhance the relationship between user and interface, especially in situations where the interaction is not successful. This essay begins with explaining what causes negative emotions, both generally and specifically in human-computer interaction. By performing sentiment analysis and facial expression analysis on user interaction with a prototype of a chatbot, it is then investigated whether negative emotional states are possible to detect. The results indicate that negative emotions can be detected, and that this is preferably done via sentiment analysis rather than facial expression analysis.

Keywords: Sentiment analysis, chatbot, emotions, human-computer interaction.

Innehållsförteckning

1 Inledning	s. 5
1.1 Syfte	s. 6
1.2 Avgränsningar	s. 6
2 Bakgrund	s. 7
2.1 Emotioner	s. 7
2.1.1 Negativa emotioner.....	s. 7
2.1.2 Negativa emotioner vid människa-datorinteraktion	s. 8
2.2 Chatt som kommunikationsmedium	s. 8
2.2.1 Skillnader i kommunikation mellan människa-människa och människa-chatbot	s. 9
2.2.2 Emotionell investering vid chatbotinteraktion	s. 10
2.3 Vikten av tillit.....	s. 11
2.3.1 Den sociala roboten	s. 11
2.4 Metodologisk bakgrund	s. 12
2.4.1 Sentimentanalys	s.12
2.4.2 Ansiktsuttrycksanalys	s. 14
3 Metod	s. 15
3.1 Material	s. 15
3.2 Experimentets utformning	s. 15
3.2.1 Procedur	s. 16
3.3 Analys av text	s. 18
3.4 Analys av video	s. 19
3.5 Analys av frågeformulär	s. 20
3.6 Pilotstudier	s.20
4. Resultat	s. 22
4.1 Resultat från textanalys	s. 22
4.2 Resultat från videoanalys	s. 22
4.3 Resultat från frågeformulär	s. 23
5 Diskussion	s. 24
5.1 Analys av text	s. 24
5.1.1 Resultat från sentimentanalys	s. 24
5.2 Resultat från ansiktsuttrycksanalys	s. 25
5.3 Deltagare	s. 26
5.4 Botparadoxen	s. 26
5.5 Chatbotens profilbild	s. 26
5.6 Interaktion med chatboten	s. 27
5.6.1 Manipulation av interaktioner	s. 28
5.7 Brister i studien	s. 28

“CAN’T YOU READ I’M FRUSTRATED?”

5.7.1 Miljö och procedur	s. 28
5.7.2 Analysmetod	s. 29
5.7.3 Mänskliga faktorn	s. 29
5.8 Realiserbarhet	s.30
5.9 Framtida studier	s. 30
6 Slutsats	s. 32
Referenser	s. 33
Bilaga A	s. 36
Bilaga B	s. 37
Bilaga C	s. 38
Bilaga D	s. 42
Bilaga E	s. 43

1 Inledning

För att användaren av ett webbaserat rekommendationssystem inte ska drabbas av informationsöverflöd bör gränssnittet i så stor utsträckning som möjligt bidra till kognitiv avlastning för denne (Klingberg, 2009). Om detta sker framgångsrikt underlättas beslutsprocessen för användaren, samtidigt som tilliten till gränssnittet ökar (Pu & Chen, 2007). Rekommendationssystem återfinns i dagsläget i flertalet internetjänster, bland annat inom näthandel, sökmotorer och för bokning av hotell eller resor, och definieras som “Ett system som använder olika informationskällor för att förse användare med prognoser och rekommendationer.” (s.109) (egen översättning) (Bobadilla, Ortega, Hernando & Gutiérrez, 2013). Det är rimligt att anta att användningen av dem kommer att öka i och med att internet blivit huvudkällan till information (Klingberg, 2009). Enligt Statistiska Centralbyrån använde 55% av Sveriges befolkning mellan 16-74 år internet för att köpa/boka reserelaterade tjänster under det första kvartalet av 2016 (Statistiska Centralbyrån, 2016).

I nuläget finns det ingen tjänst som via ett och samma webbaserade gränssnitt erbjuder användaren att jämföra alternativ, boka, samt ha en direkt överblick över alla de komponenter som är av relevans vid en konferensbokning (inbjudningar, bokning av hotell, bokning av catering, och liknande). Av denna anledning har Mohamed Abdelnaby och Claudia Bertone vid masterprogrammet “Interaction Design and Technologies” på Chalmers Universitet tagit sig an att utforma en sådan tjänst, vilket ligger till grund för denna studies syfte - att undersöka en delkomponent i deras framtagna prototyp. Tanken är att gränssnittet i fråga ska ha en chattfunktion där direkt interaktion ska kunna ske mellan användaren och ett automatiserat system, en så kallad ‘chatbot’ (Shawar & Atwell, 2005), vars syfte är att vara behjälplig för användaren vid interaktion med tjänsten. Chatboten ska kunna användas då frågor kring gränssnittet uppstår, men målet är att den också, i egenskap av interaktivt element, ska kunna fungera som en typ av automatiserad assistent kring utformningen av det event användaren avser att skapa.

Eftersom målet med att använda det gränssnitt som här är under utveckling är att spara tid, samt få kognitiv avlastning genom att ha allt gällande bokningar och dylikt samlat på ett och samma ställe, är det viktigt att chattfunktionen ytterligare bidrar till att underlätta för användaren. I det fall den inte gör det bör systemet avbryta vad det håller på med, och istället hänvisa användaren till kundtjänst eller presentera ett annat lämpligt tillvägagångssätt för denne att lösa sitt problem. För att kunna känna av när ett avbrott hade varit lägligt gäller det att systemet har möjlighet att identifiera tecken på att interaktionen inte går enligt användarens plan. Liksom Derks, Fischer och Bos (2008) påpekar i sin artikel “The role of emotion in computer-mediated communication” är det inte alltid vi väljer att kommunicera vårt emotionella tillstånd explicit, varpå det finns en poäng i att fokusera på att söka efter ord med implicit emotionell laddning. Sådana tecken går att utröna genom att analysera naturligt språk i vad som kallas sentimentanalys (se till exempel Clavel & Callejas, 2016). Genom att utföra en sådan analys blir det möjligt att identifiera en texts polaritet, det vill säga, om texten är positivt, negativt, eller objektivt laddad. I ett fall då en användare interagerar med en chatbot, och textpolariteten i den input användaren frambringar skulle göra en negativ

“CAN’T YOU READ I’M FRUSTRATED?”

skiftning, kan det ses som ett tecken på att chatboten inte är fullt så hjälpsam som den borde vara (se till exempel Fitriane, Wiggers & Rothkrantz, 2003).

En utmaning med att realisera den typ av chatbot vi förespråkar här är att sentimentanalysen behöver ske i realtid. I större delen av den litteratur vi undersökt där målet varit att detektera emotioner i text har författarna haft tillgång till gammal data i form av tweets, bloggar och chattloggar. De metoder som förespråkar en utredning bortom sentimentanalys för att avgöra vilken emotionell laddning en text har, kännetecknas ofta just av att de behöver tillgång till stora mängder autentisk data (data hämtad från bloggar/chattloggar online eller loggad interaktion med faktisk chatbot, se till exempel Ghazi, Inkpen, & Szpakowicz, 2014; Hill, Randolph Ford & Ferreras, 2015; Smith, Lee, Barnden & Hancox, 2011). Av uppenbara skäl leder detta till att analysmetoderna blir betydligt mer känsliga för språkliga nyanser och kontextuella faktorer, varpå de med större precision kan upptäcka tecken på emotioner - implicita som explicita.

Oavsett om målet med att undersöka denna typ av analysmetoder på lång sikt är att utveckla dem så att de ska bli möjliga att implementera i realtid eller ej, är vi i denna studie härvid begränsade med anledning av två faktorer. Den första faktorn har att göra med vilken analytisk mjukvara som finns att tillgå, eftersom den vi planerar att implementera möjliggör analys på befintlig text och därmed inte i realtid. Den andra faktorn relaterar till datainsamlingen, eftersom vi arbetar med en avgränsad mängd data.

1.1 Syfte

Syftet med denna studie är att undersöka huruvida ett automatiserat chattsystem har möjlighet att avgöra om interaktionen är framgångsrik eller ej utifrån det som kommuniceras av en användare. Således lyder forskningsfrågan: Är den text som genereras vid interaktion mellan en användare och en chatbot tillräcklig för att användarens negativa emotionella tillstånd ska kunna utrönas vid en sentimentanalys?

1.2 Avgränsningar

Denna studie tar inte hänsyn till vare sig chatbotens eller experimentdeltagarnas personlighetsdrag och hur dessa påverkar interaktion av det slag vi undersöker. Utöver det bortser vi helt från att undersöka positiva emotioner, och vi tar inte heller i beaktande att stress/tidspress kan påverka hur användare uttrycker sig. Även semantiska analyser och reflektioner kring den språkliga diskrepans som möjligen uppstår när svenskspråkiga experimentdeltagare får delta i ett experiment på engelska hamnar utanför ramarna. Där hamnar också frågan kring hur ett automatiserat system bör bemöta negativa emotioner, samt antagandet att negativa emotioner ibland uttrycks via ironi/sarkasm.

2 Bakgrund

Nedan redogör vi för en kognitivt präglad teori kring emotioners uppkomst och natur, som ligger till grund för vårt antagande att negativa emotioner uppstår hos användaren när denne inte lyckas med att genomföra en formulerad plan. Därefter presenteras teorier kring negativa emotioner inom specifikt människa-dator-interaktion, varpå vi går vidare med att undersöka forskningsfrågan.

2.1 Emotioner

Denna studie utgår från Oatley och Johnson-Lairds (1987) teori som menar på att emotioner fyller en viktig funktion för det kognitiva bearbetandet. Detta eftersom de utgör en avgörande delkomponent i hur människor bestämmer, organiserar och utför olika planer som hör till mål vi sätter upp. Mål är abstrakta representationer av ett scenario som vi strävar efter att uppnå. Planer består i sin tur av en länk mellan vårt nuvarande tillstånd och måltillståndet - en möjlig väg mot målet som vi använder som underlag för att agera. Ett mål kan vara att cykla till jobbet, att lyckas lämna in ett projektarbete i tid, eller att köpa ett hus. De kan alltså vara triviala eller livsavgörande och mer eller mindre medvetna eller uttalade. Lyckas vi med att uppnå målet väcks en positiv emotion, men när vi stöter på hinder som kan försvåra vår väg mot målet upplevs istället en negativ sådan. Emotionssignalerna hör enligt teorin till ett eget kommunikationssystem vars delkomponenter inte behöver utvärderas på ett högre plan för att emotionen ska få fäste, vilket innebär att vi kan hinna uppleva den innan det står klart huruvida den var befogad eller ej (Oatley & Johnson-Laird, 1987).

En premis för denna teori är att det finns fem universala grundemotioner. Dessa är glädje, oro, ilska, sorg och avsmak. Förutom dessa basemotioner vill andra teorier ibland även inkludera förakt och överraskning (till exempel Ekman, 1992), men enligt Oatley och Johnson-Laird (1987) är dessa snarare kombinationer av andra emotioner. Ett krav för att en emotion ska räknas som en basemotion är att ansiktsuttrycket som följer av emotionen ska vara universalt, alltså igenkännbart för individer från kulturer runt om i världen (Oatley & Johnson-Laird, 1987).

2.1.1 Negativa emotioner

Med negativa emotioner menas i denna studie de emotionella tillstånd som faller in under Oatley och Johnson-Lairds (1987) kategori ‘dysforiska’, det vill säga ‘sorg’, ‘oro’, ‘ilska’ och ‘avsky’. Negativa emotioner uppstår när planen som ska leda till vårt mål avbryts, särskilt i situationer vi inte är bekanta med sedan innan. Om planen avbryts och vi inte vet hur vi nu ska ta oss mot målet förskjuts vi in i ett ‘dysforiskt tillstånd’ tills vi antingen bytt mål eller funnit en ny möjlig väg mot målet (Oatley & Johnson-Laird, 1987). I en vanlig konversation med en annan person har vi oftast inga svårigheter att anpassa vår plan i realtid och därmed uteblir generellt en stark emotionell reaktion om någonting skulle hamna utanför den formulerade planen. Om vi däremot befinner oss i en obekant situation, till exempel om vi försöker boka en konferens via en hemsida vi aldrig förut besökt, kan negativa emotioner uppstå när planen avbryts och vi inte längre vet hur vi bör agera för att komma vidare (Oatley & Johnson-Laird, 1987).

“CAN’T YOU READ I’M FRUSTRATED?”

Den här studien eftersträvar att detektera och uppmärksamma negativa emotioner snarare än att ställa positiva mot negativa, och en av de emotioner som ligger i fokus är frustration. Frustration beskrivs av Averill (1982) som “avbrottet av en pågående eller planerad aktivitet” (s. 173) (egen översättning), och samma författare beskriver också hur frustration i många fall leder till ilska, men inte alltid. Detta torde berättiga en tolkning av frustration som en lätt variant av (eller ett förstadium till) ilska. Frustration beskrivs även av Freud (1921) som en emotion som uppkommer när någonting konkurrerar med, hotar, eller avbryter förverkligandet av ett mål eller en pågående plan. Tidigare studier visar att denna emotion är den mest frekvent förekommande negativa emotionen vid människa-datorinteraktion, och relaterar förutom till själva avbrottet i en plan också till användarens förväntningar av interaktionen mer generellt (Bessi re, Newhagen, Robinson & Shneiderman, 2004). Dessa beskrivningar g r i linje med med Oatley och Johnson-Lairds (1987) teori om emotioner.

2.1.2 Negativa emotioner vid m nniska-datorinteraktion

En interaktion mellan anv ndare och gr nssnitt som upprepade g nger  r framg ngsrik och d r anv ndarens emotioner tas h nsyn till pr glas av tillit och  kad lojalitet, j mf rt med om emotioner inte tas h nsyn till  verhuvudtaget (Clavel & Callejas, 2016). Med det sagt har en viss, relativt l g, grad av frustration hos anv ndaren av ett datorbaserat gr nssnitt samtidigt visat sig f rh ja dennes kognitiva kapacitet. Detta s rskilt om kostnaden f r problemet inte uppfattas som st rre  n bel ningen som kommer n r problemet l ses, eller om upplevelsen av problemet inte  r alltf r negativ (Bessi re et al., 2004). Om problemet skulle uppfattas som ol sbart, alltf r l ngvarigt, eller s  pass komplext att det resulterar i informations verfl d v cks ist llet starkare k nslor av frustration, alternativt irritation eller ilska (Bessi re et al., 2004).

K nslor av frustration vid m nniska-datorinteraktion springer ur interna eller externa hinder. De interna blockeringarna kan best a av bristf llig kunskap eller f rm ga hos individen som interagerar med gr nssnittet, medan de externa handlar om brister i det system som interageras med och/eller i den fysiska verkligheten (Bessi re et al., 2004). D rmed blir det i denna studie efterstr vansv rt att skapa externa blockeringar, f r att p  s  vis undvika upplevda k nslor av inkompetens eller skuld hos anv ndaren.

2.2 Chatt som kommunikationsmedium

Att framg ngsrikt f ra en konversation i textformat  r en process som ut ver formulering och tolkning av b da deltagare kr ver ett kollaborativt engagemang. N r tv  parter interagerar med varandra har ingen av dem full insyn i den andres medvetande, och de kan s ledes inte vara s kra p  att den andre tolkar det den skriver p  det s tt den avsedde det. Vi tycks dock vara duktiga p  att koordinera konversationer genom att systematiskt s ka och bidra med bevis p  att det som sagts blivit f rst tt - vad som brukar kallas f r ‘*grounding*’ (Clark & Brennan, 1991). Negativ liksom positiv feedback tycks b da vara n dv ndiga f r att h lla konversationen p  r tt sp r, och n r n gonting g r fel kr vs det att  tminstone en av

“CAN’T YOU READ I’M FRUSTRATED?”

deltagarna gör den ansträngning som behövs för att reda ut missförståndet (Brennan & Hulteen, 1995).

När det gäller människa-datorinteraktion, och särskilt människa-chatbotinteraktion, är det endast individen som sitter framför skärmen som har möjlighet att korrigera sitt meddelande om dennes input inte får förväntad respons. Meredith och Stokoe (2014) undersökte i studien “Repair: Comparing Facebook ‘chat’ with spoken interaction” hur reparation och korrigering vanligtvis går till när användare kommunicerar via chatt. De fann att en form av korrigering, vad de kallar ‘*message construction repair*’, återfinns i chatt men inte i talad konversation. Detta är aktuellt för oss eftersom det gränssnitt vi avser undersöka ‘inbjuder till’ en särskild typ av interaktion, nämligen skriven sådan, där användarens respons hela tiden är avhängande av chatbotens respons och tvärtom. Med ‘inbjuder till’ syftas här på begreppet ‘*affordance*’. *Affordance* innebär de användningsmöjligheter som en användare uppfattar hos en artefakt eller ett gränssnitt (Norman, 2013).

Message construction repair är vad som sker när användaren korrigerar meddelandet innan det skickas iväg. Möjligheten att göra detta återfinns bara i de tekniska kommunikationsmedium som inbjuder till sådant, och för med sig att en användare har möjlighet att omformulera eller helt radera det meddelande som påbörjats. Detta innebär att det till skillnad från vid kommunikation som sker ansikte mot ansikte saknas vissa interaktionella möjligheter för samtalsparterna. Således går det att säga att chattbaserade konversationer präglas av ‘reducerad spontanitet’ (Meredith & Stokoe, 2014). Detta förstärks dessutom av att det saknas möjlighet att plocka upp icke-verbala ledtrådar, såsom kroppsspråk, ansiktsuttryck och tonläge, varpå individen har större kontroll över hur denne väljer att uttrycka sina emotioner (Derks et al., 2008).

2.2.1 Skillnader i kommunikation mellan människa-människa och människa-chatbot

Hill et al. visar i sin studie från 2015 på hur sättet vi uttrycker oss i skrift i viss mån skiljer sig beroende på om det är en människa eller ett automatiserat system vi kommunicerar med. Studien gick ut på att jämföra gamla chattkonversationer mellan två individer med sådana som förts mellan individ och chatboten Cleverbot¹. I enlighet med vad Derks med kollegors sammanfattning “The role of emotion in computer-mediated communication” från 2008 visar, utgår Hill och kollegor (2015) från antagandet att vi uttrycker emotioner lika bra eller bättre vid textmässig kommunikation som vid *face-to-face*-kommunikation. Att det faktiskt fanns olikheter i den data de jämförde kan dock komma att få en viss betydelse i den studie vi har för avseende att genomföra här. Till exempel fann de att interaktionerna med Cleverbot innehöll fler meningsutbyten, även om dessa bestod av färre ord. En annan faktor som skilde sig åt var huruvida orden var positivt eller negativt laddade, eftersom det visade sig att konversationerna med Cleverbot innehöll betydligt fler svordomsord eller ord som hade sexuella anspelningar. De föreslår att en anledning till att dialogerna med Cleverbot faktiskt var längre men med kortare meningar skulle kunna vara att vi anpassar språket efter den/det vi talar med, snarare än att vi på något vis skulle famla oss fram i samtalet. Varför dessa

¹ www.cleverbot.com

“CAN’T YOU READ I’M FRUSTRATED?”

konversationer innehöll fler svordomar och sexuellt anspelande ord spekuleras inte i, men en anledning skulle kunna vara att användarna ‘testar’ chatboten för att se vilken reaktion de får (se avsnitt 2.2.2 nedan).

Eftersom en chatbot i egenskap av kommunikationsmedium inbjuder till en särskild typ av interaktion, där denna både är avhängande av input/output samt tillåter användaren att göra så kallade *message construction repairs* (Meredith & Stokoe, 2014), skiljer sig förutsättningarna för textanalys från de som gäller vid individ-till-individ-kommunikation. Vi tycks dessutom anpassa språket efter den eller det vi chattar med (Hill et al., 2015). Detta antyder att det behöver framgå för användaren att det är relevant för denne att kommunicera sitt emotionella tillstånd, för att gränssnittet ska kunna ta hänsyn till eller anpassa sig efter detta. Möjliga sätt att uppmuntra till det torde vara att konversationen upplevs som tillräckligt naturlig för att användaren ska uppleva det som relevant, alternativt att gränssnittet erbjuder det på annat sätt.

2.2.2 Emotionell investering vid chatbotinteraktion

Sundar, Bellur, Oh, Jia och Kim visar i sin studie (2016) “Theoretical Importance of Contingency in Human-Computer Interaction: Effects of Message Interactivity on User Engagement” på att det kan vara önskvärt med den typ av interaktiva funktion som behandlas i denna studie i ett webbaserat gränssnitt. Detta eftersom en uppfattning hos användaren av att det finns en responsiv mottagare på andra sidan skärmen bidrar till att förstärka användarens känsla av förbindelse till tjänsten. En webbaserad tjänst som tillåter interaktion som upplevs som ‘*contingent*’, det vill säga, där meddelandet som tas emot upplevs vara avhängande av det meddelande som sänts ut, bidrar både till förbindelse samt en uppfattning av ‘reciprocitet’ mellan användare och gränssnitt (*contingency* betyder här ungefär ‘villkorlig’ eller ‘kausaltetsbaserad’) (Sundar et al., 2016). Reciprocitet betyder i detta sammanhang ungefär upplevelsen av att ge och ta mellan användare och gränssnitt, och blir viktigt i och med frågan om huruvida vi upplever ‘robotisk feedback’ som likvärdig mänsklig sådan (Hingston, 2012).

Sundar med kollegors studie (2016) bestod av ett mellangruppsexperiment med fem olika grupper, där två av dessa hade möjlighet att chatta med varsin typ av kundtjänst, en mänsklig och en chatbot. Tre grupper saknade möjlighet att interagera med någon form av kundtjänst. Resultaten visade att nivån av upplevd *contingency* skilde sig åt mellan de tre grupper som saknade kundtjänstfunktionen och de två grupper som interagerade med hemsidan då detta element fanns tillgängligt. De två grupper som hade tillgång till chattbaserad kommunikation upplevde nivån av *contingency* som högre.

Detta är av två anledningar relevant i denna studie; för det första gällande sambandet mellan *contingency* och uppvisande av negativa emotioner. Stöd för att en mer villkorlig konversation skulle bidra till en större emotionell investering kan finnas sett till den korrelation som finns mellan hög kognitiv belastning och tendens att ge uttryck för negativa emotioner (Storbeck, Davidson, Dahl, Blass, & Yung, 2015). Den typ av synkroniserade interaktion som krävs för att uppehålla *contingency* bidrar nämligen till en systematisk snarare än heuristisk processering av meningsinnehållet, vilket tvingar oss att vara

“CAN’T YOU READ I’M FRUSTRATED?”

involverade i den dialog som genereras (Sundar et al., 2016). För det andra manipulerades deltagarna i studien endast till att tro att de pratade med en chatbot - i själva verket pratade de med en människa (Sundar et al., 2016).

2.3 Vikten av tillit

I genomsnitt avbryts personal på arbetsplatser i USA en gång var tredje minut i sitt arbete. Detta talar Klingberg om i sin bok “Overflowing brain” (2009), och han menar att en anledning till detta är att de som arbetar vid en dator ofta har så många som 8 flikar på sin webbläsare öppna samtidigt. I boken används begreppet ‘informationsöverflöd’ (egen översättning) för att peka på ett modernt problem som uppstått när tekniken tillåter oss att kommunicera via flera medium samtidigt, samt att internet med sina distraktionsmoment i form av reklam och andra interaktiva element blivit huvudkällan till information.

En webbaserad tjänst med syfte att avlasta användarens arbetsminne behöver inte bara vara utformad på ett sätt som motverkar informationsöverflöd. Den bör också med tillgängliga medel sträva mot att vinna användarens förtroende, för att användaren ska kunna känna sig bekväm med att använda den som hjälpverktyg (Pu & Chen, 2007). Ett gott användarförtroende inom människa-robotinteraktion tycks korrelera med uppnådd tillit (Hingston, 2012). I nästa avsnitt avser vi utveckla varför tillit är en viktig komponent även i människa-chatbotinteraktion, samt hur detta kan åstadkommas.

2.3.1 Den sociala roboten

“Människor tenderar att respondera på ett annat sätt gentemot autonoma interaktiva system än vad de gör mot ‘vanliga datorsystem’ (...), inte bara svaren skiljer sig, utan även förväntningarna varierar och tenderar att gå i en mer social riktning ju mer antropomorferad systemdesignen är.” (s. 1) (egen översättning). Det skriver Hingston i sin bok “Believable bots” från 2012. Ett begrepp som behöver redas ut i det citatet är ‘social’. I Svenska Akademiens ordlista (2009) definieras social som ‘samhällelig’ eller ‘sällskaplig’ - beskrivningar som inte direkt för tankarna till automatiserade system. Detta till trots tycks människor se på sådana som sociala aktörer, åtminstone så länge de ger sken av att inneha intentioner (Hingston, 2012). Intentioner förklaras av Jensen (2016) som en vilja eller drivkraft, och är “sammansatta av två nödvändiga delar: mål och medel” (s. 74). Således har de en klar koppling till de mål som Oatley och Johnson-Laird (1987) talar om som avgörande för vilka emotioner vi upplever.

Vi tycks alltså applicera förmågan att inneha och tolka emotioner på en artificiell agent, så länge vi uppfattar att denna också besitter ett mål (Hingston, 2012). Detta har vissa likheter med ‘*theory of mind*’, som kan förklaras som möjligheten att förstå vad som rör sig i andra individers tankar, samt insikten att detta kan skilja sig från vad som finns i våra (Linblom, Susi & Tysk, 2012). Clavel och Callejas (2016) menar att ett sätt att öka systemets trovärdighet vid människa-datorinteraktion är att låta kommunikationen präglas av vad som brukar kallas för ‘spegling’, det vill säga, att återanvända ord eller fraser som användaren lämnar ifrån sig i den text som chatboten genererar. Vanligtvis är spegling en automatisk respons hos en individ gentemot en annan, bestående av en subtil imitation av gester, tonläge,

“CAN’T YOU READ I’M FRUSTRATED?”

skratt eller ansiktsuttryck, och förmågan är starkt förknippad med *theory of mind* (Linblom et al., 2012).

En annan faktor som kan vara avgörande för om vi uppfattar ett livlöst objekt som innehavande av sociala egenskaper är huruvida det rör sig på ett sätt som kan uppfattas som motiverat (Heider & Simmel, 1944). Vid en situation där interaktionen är textberoende skulle detta eventuellt kunna översättas till att gälla hur snabbt chatboten svarar, samt att svaren uppfyller användarens förväntningar, baserat på dennes egna input till chatboten. I ett fall då vi är medvetna om att den andra parten inte är en människa tycks vi i början vilja ‘testa’ den för att se om den verkligen beter sig så som den utger sig för att kunna bete sig (Sundar et al., 2016). Det vill säga, beter sig en automatiserad chattfunktion i enlighet med våra förväntningar på en textmässig konversation med en annan individ torde vi bete oss mot den som vi hade betett oss vid kommunikation med en människa.

2.4 Metodologisk bakgrund

I denna studie är tanken att användaren av chatboten har för avseende att lösa ett antal givna problem, och chatbotens främsta uppgift är att hjälpa till med att lösa dessa. Tecken på emotioner såsom frustration eller ilska torde vara en signal på att användaren helt eller delvis misslyckats med sin plan, och att interaktionen således inte är framgångsrik. Negativa emotioner innebär en risk att användaren överger sin plan och lämnar gränssnittet, och i det långa loppet kan det föra med sig att hen väljer att inte fortsätta använda tjänsten. Vad som skulle väcka negativa emotioner i detta sammanhang blir enligt Oatley och Johnson-Lairds (1987) teori att planen som användaren nyttjar för att nå målet avbryts på grund av något som ligger utanför dennes makt. Genom att utföra en textanalys med avsikten att detektera användarens emotionella läge före och efter avbrottet går det möjligen att se hur den textmässiga inputen från användaren förändras, och således huruvida användarens emotionella tillstånd förmedlats i texten eller inte.

För att kunna ge en sammanhängande motivering till val och utformning av metod utvecklas begreppen ‘sentimentanalys’ och ‘ansiktsuttrycksanalys’ i varsitt avsnitt. Eftersom begreppet sentimentanalys dessutom är relativt brett förtydligas här vilken typ som används i denna uppsats.

2.4.1 Sentimentanalys

En sentimentanalys ger ledtrådar kring hur vi kan formulera vissa kognitiva mekanismer ur ett perspektiv som går bortom renodlade emotionsteorier. Via sentimentanalys går det att finna ledtrådar till upphovspersonens emotioner och attityder via textbaserad data, såsom exempelvis twitterinlägg eller filmrecensioner (Clavel & Callejas, 2016). Dessa analysmetoder har blivit ett allt mer eftersträvanvärt element inom *text mining* och *opinion mining* (bearbetande av stora mängder text), eftersom det finns ett värde i att veta huruvida en användares attityd gentemot exempelvis ett gränssnitt är positiv eller negativ, särskilt i kontexter där gränssnittet kan eller behöver förändra sitt beteende efter användarens tillstånd (Clavel & Callejas, 2016).

“CAN’T YOU READ I’M FRUSTRATED?”

I en sentimentanalys genereras ett övergripande sentimentvärde för texten, vilket blir textens ‘laddning’ eller ‘polaritet’. Analysen i det här fallet sker med hjälp av databasen SentiWordNet 3.0. SentiWordNet är en lexikalisk databas för sentimentanalys baserad på den engelskspråkiga databasen WordNet 3.0 (Princeton University, 2015). I WordNet har alla ord (155,287 stycken) försetts med etiketter, beskrivningar, samt relationer till varandra på ett semantiskt och lexikaliskt plan. Orden har grupperats in i 117,659 så kallade ‘synsets’, vilka består av vad Princeton University (2015) kallar kognitiva synonymer. Med kognitiva synonymer menas kategorier där det framgår vad för laddning ordet i fråga har beroende på semantisk kontext. I och med de lexikaliska och semantiska relationerna som framgår via *synset*-grupperingen av ord kan SentiWordNet i viss mån ta hänsyn till kontextualitet. Orden från WordNet har i SentiWordNet försetts med ett positivt, negativt, eller objektivt värde, alternativt har ordet både ett positivt och ett negativt värde associerat till sig och då bestämmer kontexten vilken laddning som blir mer framstående (Princeton University, 2015; Baccianella, Esuli & Sebastiani, 2010).

Värdet i en sentimentanalys av den här typen kan sträcka sig från -1 som mest negativ, till 1 som mest positiv. För att visa på sentimentanalysens förmåga att utrona nyansskillnader i språket har vi formulerat tre exempelstycken och analyserat dessa med hjälp av SentiWordNet. I exemplena nedan går det att se att en text som vid första anblick kan tyckas vara mycket emotionellt laddad får ett sentimentvärde som kan ses som relativt lågt.

- **Positiv**

This flower smells amazing, the sun is shining! I am so grateful, my head is filled with positive thoughts. This is a beautiful place, I love being here. How fantastic it is to be able to experience this. My mind is at ease, my body feels light. I wish I could stay here forever, I’ve found a true courteous accepting friend with a benign smile. She is the wisest and most advisable angel. It’s beautiful, this gift of life, a precious present from above. Life is a party with laughing, joy, friend. What a lush life. I feel creative, artistic, ravishing, perfect.

Sentimentvärde: 0,205

- **Negativ**

Everything is terrible. The walls are ugly, they’re making me feel bad and afraid, I don’t want to be here. It’s stressful, bad, it’s terrible. I feel sick, agitated and angry. I wanna tear these walls down aggressively with a sledgehammer. This house is making me hate everything and I’m bitter. My uncle is a schizophrenic shark. He is looking for a scapegoat and called me an evil, manipulative sadist. He stole my money and robbed my cousins, calling my mom a trollop and loser.

Sentimentvärde: -0,224

- **Objektiv**

I read in a book that cognitive science is an interdisciplinary field. Unlike for example the science of medicine, cognitive science deals with areas very different from each other, but areas that are still connected. These include philosophy, neurobiology, and artificial

“CAN’T YOU READ I’M FRUSTRATED?”

intelligence. The professor I had in my classes was a normal, middle-aged man. He was always on time and punctual. The classes were objective.

Sentimentvärde: -0,003

2.4.2 Ansiktsuttrycksanalys

Tidigare studier visar att så mycket som 55% av den emotionella meningen i ett budskap framgår icke-verbalt, det vill säga via gester, ansiktsuttryck eller kroppshållning (Fitrianie et al., 2003). Detta tillsammans med vad som av Meredith och Stokoe (2014) omnämns som ‘reducerad spontanitet’ vid textmässig kommunikation motiverade oss att vidga analysmetoden till att även involvera ansiktsuttrycksanalys. För att ha möjlighet att fånga upp eventuella emotionella uttryck som inte skulle framgå i text filmades därför deltagarnas ansikten under tiden experimentet pågick. Detta material analyserades sedan med hjälp av mjukvaran FaceReader 6.1. Bland de emotionskategorier som FaceReader sorterar in ansiktsuttryck under återfinns: ‘*Unknown*’, ‘*Happy*’, ‘*Neutral*’, ‘*Angry*’, ‘*Sad*’, ‘*Disgusted*’ och ‘*Surprised*’ (Okänd, Glad, Neutral, Ledsen, Äcklad och Överraskad, egen översättning) (Noldus, 2017). Av dessa kategoriserade vi ‘*Angry*’, ‘*Sad*’ och ‘*Disgusted*’ som negativa.

“CAN’T YOU READ I’M FRUSTRATED?”

3 Metod

Nedan beskrivs material, metodens utformning, samt hur studien genomfördes i praktiken, före, under, och efter experimentet.

3.1 Material

- Kamera: Blackmagic pocket camera.
- Slack: Onlinebaserat verktyg (www.slack.com), användes här för interaktionen mellan deltagarna och en simulerad chatbot (simulerad av försöksledare 1).
- MacBook Pro 1: Användes av deltagarna under experimentet, kompletterad med extern datorskärm, tangentbord och mus.
- MacBook Pro 2: Användes av försöksledare 1 för att simulera chatboten.
- Quicktime player 10.4: Användes för att spela in skärmaktivitet på MacBook Pro 1.
- Analysverktyg för text: RapidMiner Studio 7.4 och SentiWordNet 3.0.
- Analysverktyg för video: FaceReader 6.1.
- Analysverktyg för statistik: IBM SPSS Statistics 24.

3.2 Experimentets utformning

Interaktionselementet i studien bestod av ett experiment där deltagaren och en simulerad chatbot höll en ca 30 minuter lång konversation. För detta användes verktyget Slack via webbläsaren Google Chrome (se fig. 1), där användarna och den simulerade chatboten (försöksledare 1) kopplades upp mot varandra med hjälp av individuella mailkonton. I chatttrutan var chatboten döpt till “*bookingassistant*” och den hade en tillhörande profilbild (se fig. 2). Samma försöksledare agerade chatbot genom varje session, och var ej synlig för någon av deltagarna under tiden denna pågick.

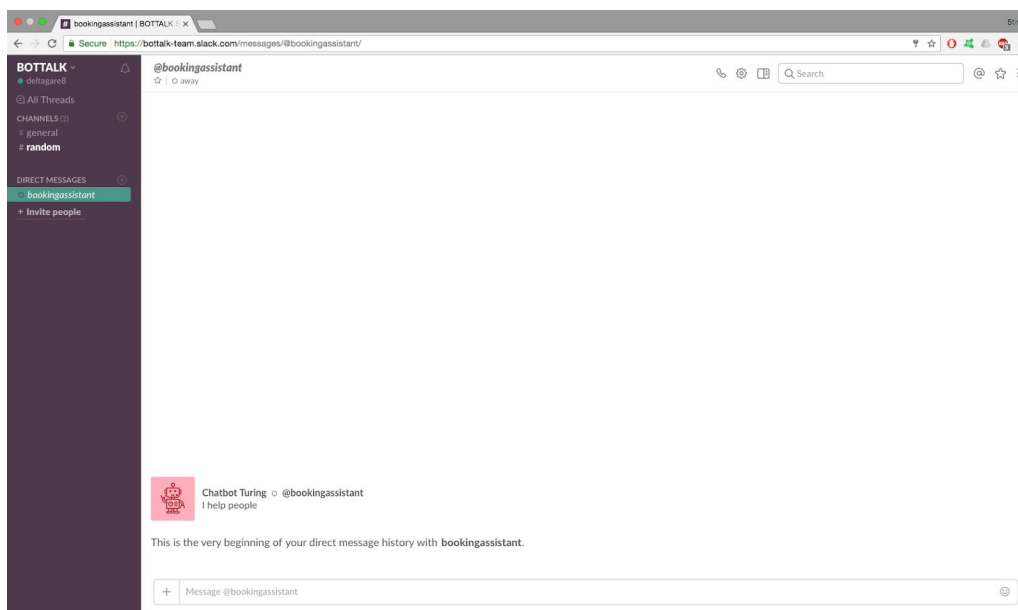


Fig. 1: Överblick av Slack i webbläsaren Google Chrome.

“CAN’T YOU READ I’M FRUSTRATED?”

I Slack är det möjligt att skapa en separat chattkanal för interaktion mellan specifika användare. En sådan skapades i förväg och samtliga deltagare registrerades inom kanalen för att få åtkomst till chatboten. Varje deltagare hade således ett unikt konto för sin interaktion, och samtliga kunde enbart se sin egen konversation.

Slack användes med motiveringen att försöksdeltagarna inte skulle vara för bekanta med gränssnittet sedan tidigare. Verktöget hade i oktober 2016 cirka fem miljoner dagliga användare (DMR, 2017), jämfört med exempelvis Facebook Messenger som i mars 2017 hade runt 1,2 miljarder dagliga användare (Facebook, 2017). I och med att Slack har avsevärt färre användare bedömdes försöksdeltagarna ha mindre erfarenhet av gränssnittet.

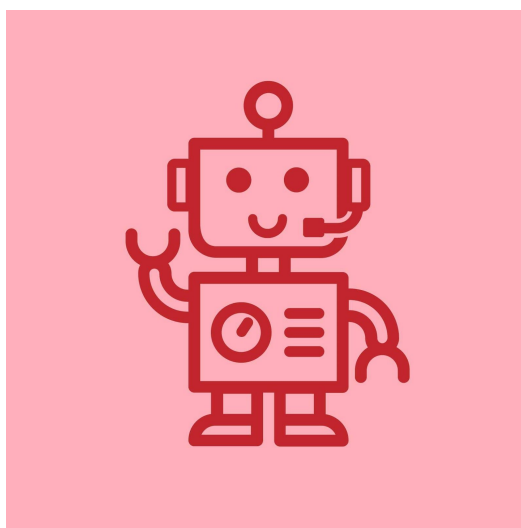


Fig. 2: Chatbotens profilbild.

Slacks gränssnitt är, precis som instruktionerna och uppgifterna för experimentet, utformat på engelska. Således kunde en kontinuitet hållas i studien, vilken behövdes hållas på engelska eftersom sentimentanalysen sker med hjälp av en lexikalisk databas med engelska ord. Gränssnittet är till viss del begränsat enligt vad som är önskvärt vid denna typ av undersökning, eftersom vissa element kan sorteras bort (till exempel information om andra användare aktiva i samma kanal, se fig. 3, punkt 1). Eftersom deltagarna lämnades ensamma under interaktionen kunde vi dock inte veta huruvida denne valde att lämna chatttrutan. För att ha möjlighet att kontrollera för detta spelades skärmaktiviteten in.

3.2.1 Procedur

Experimenten ägde rum i Chalmers Media Lab på Lindholmen. Alla deltagare skrev under en medgivandeblankett (se bilaga A) där de gick med på att sessionen samt skärmaktiviteten spelades in. Samtliga deltagare blev informerade om hur experimentet skulle gå till av försöksledare 2 (se bilaga B), efter vilket de delgavs ett instruktionsblad (se bilaga C) som förklarade hur de skulle gå tillväga för att utföra ett antal uppgifter genom att interagera med chatboten. Därefter lämnades de ensamma i rummet.

Som ett försök att förstärka illusionen av att de pratade med ett automatiserat system försågs deltagarna med ett antal bokningsrelaterade scenarier (A, B och C) som sades vara

“CAN’T YOU READ I’M FRUSTRATED?”

förprogrammerade, och de instruerades att ladda in varje scenario via en viss procedur (se bilaga C). Scenario A och B gick vid interaktionen felfritt, det vill säga, chatboten bidrog med assistans och svarade korrekt på det användaren skrev. I scenario C bidrog inte chatboten med någon hjälp alls, utan missförstod och svarade fel på varje fråga, fram till den punkt då försöksledaren som agerade chatbot ansåg att det var dags att gå vidare till nästa uppgift. Efter avslutad session fick varje deltagare fylla i ett frågeformulär (se bilaga D).

Genom att ha två scenarier som gick bra och ett som gick mindre bra, avsåg vi skapa en mer gediget etablerad relation mellan användare och chatbot än om det enbart hade varit ett scenario av vardera slag (för vidare läsning se Bessière et al., 2004). Hädanefter refererar vi till A och B som A+B, eftersom dessa båda scenarier kan ses som ett helhetligt scenario separat från C, i den mening att interaktionen skiljer sig åt. Scenario A+B kommer att analyseras ihop i sentimentanalysen på grund av detta.

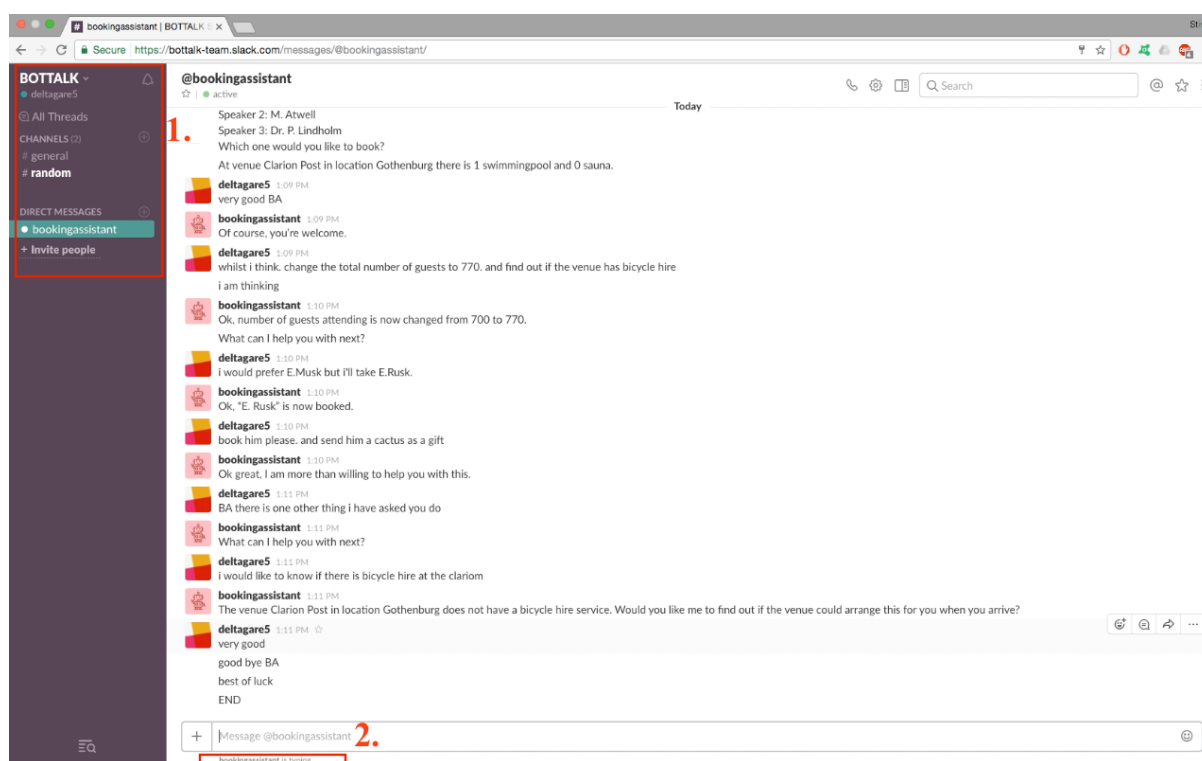


Fig. 3. Skärmbild på konversation mellan deltagare och chatbot.

1. I denna ruta går det att justera vilka element som är synliga för användaren (gamla chattloggar, kanaler, slacks egen chatbot).
2. När avsändaren formulerar ett svar får användaren återkoppling.

Sundar et al. (2016) använde sig av en simulerad chatbot i sin studie, och valde då att använda samma hälsningsfraser vid varje interaktion för att etablera konversationen samt hålla den konsekvent. Med inspiration från detta valde vi att utforma ett chatbot-manus (se bilaga E), som vi under experimentets gång kunde använda oss av. Chatbotens intention att assistera användaren fastställdes med hjälp av hälsningsfrasen “Hello! My name is bookingassistant. I can do a number of things! Anything from guest contact to booking caterers - I’m your bot!” (se bilaga E). Med hjälp av vissa förbestämda riktlinjer (om

“CAN’T YOU READ I’M FRUSTRATED?”

användaren skriver x, svarar chatboten y) tillät manuset dessutom att interaktionen kunde styras till att först vara framgångsrik, för att sedan gå mot att bli mindre framgångsrik.

Vid en konversation mellan två individer går det i Slack-chattfönstret att se när den andra individen skriver (se fig. 3, punkt 2). Ytterligare en anledning till att använda ett manus med förbestämda fraser blev således för att undvika att en observant deltagare reagerade på att chatboten var aktiv, men att ett svar dröjde eller uteblev. Denna avvägning var delvis motiverad av Heider och Simmels (1944) slutsats att vi i högre grad uppfattar livlösa objekt som innehavande sociala egenskaper i de fall då vi uppfattar dessas rörelsemönster som motiverade. Ett omotiverat beteende i form av att inte svara användaren torde inte ha uppmuntrat till den typ av kommunikation vi var intresserade av att undersöka.

3.3 Analys av text

Före analys gällades den text som genererats vid interaktionerna så att endast det deltagarna skrivit återstod. Eventuella stavfel korrigerades för att sentimentanalysen skulle kunna innefatta samtliga ord, och varje deltagares text analyserades för sig. Scenario A+B analyserades ihop, medan texten från C analyserades separat. Datan från scenario A+B analyserades tillsammans eftersom dessa scenarier hade en likvärdig interaktion, i den mening att chatboten var hjälpsam och på så vis underlättade för användaren.

Analysen genomfördes med hjälp av mjukvaran RapidMiner Studio 7.4 och tillägget SentiWordNet 3.0. SentiWordNet implementerades i RapidMiner enligt fig. 4, där även en överblick ges över hur sentimentanalysen var utformad. Modulerna organiserades i RapidMiner på detta vis för att vi enkelt skulle kunna jämföra utfallet mellan scenario A+B med scenario C för varje deltagare. Sentimentvärdet från A+B jämfördes sedan med sentimentvärdet från C i ett beroende *t*-test.

“CAN’T YOU READ I’M FRUSTRATED?”

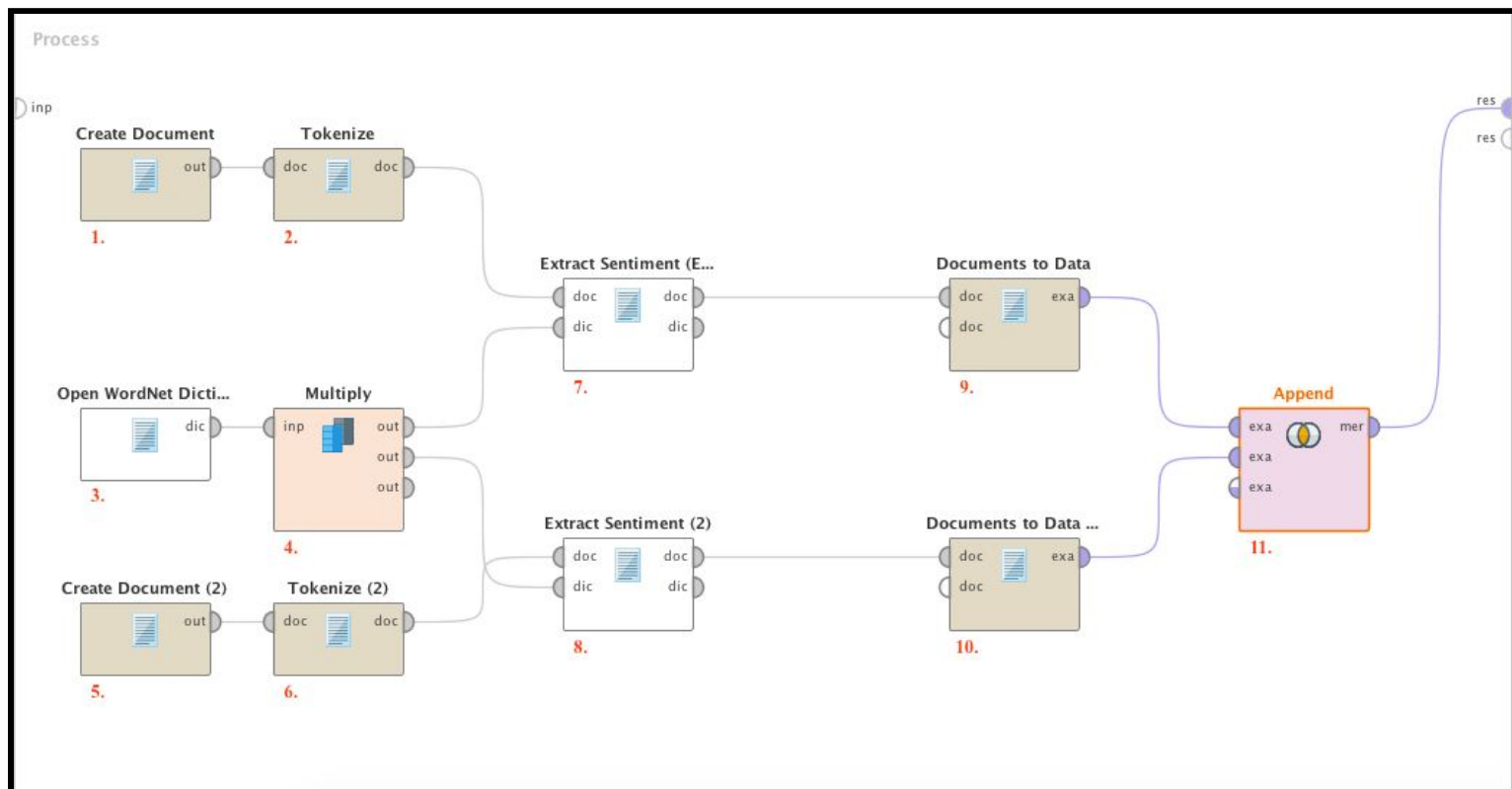


Fig. 4: Skärmbild på sentimentanalysprocessen i RapidMiner. Nedan följer en beskrivning av delkomponenterna, numrerade enligt bilden.

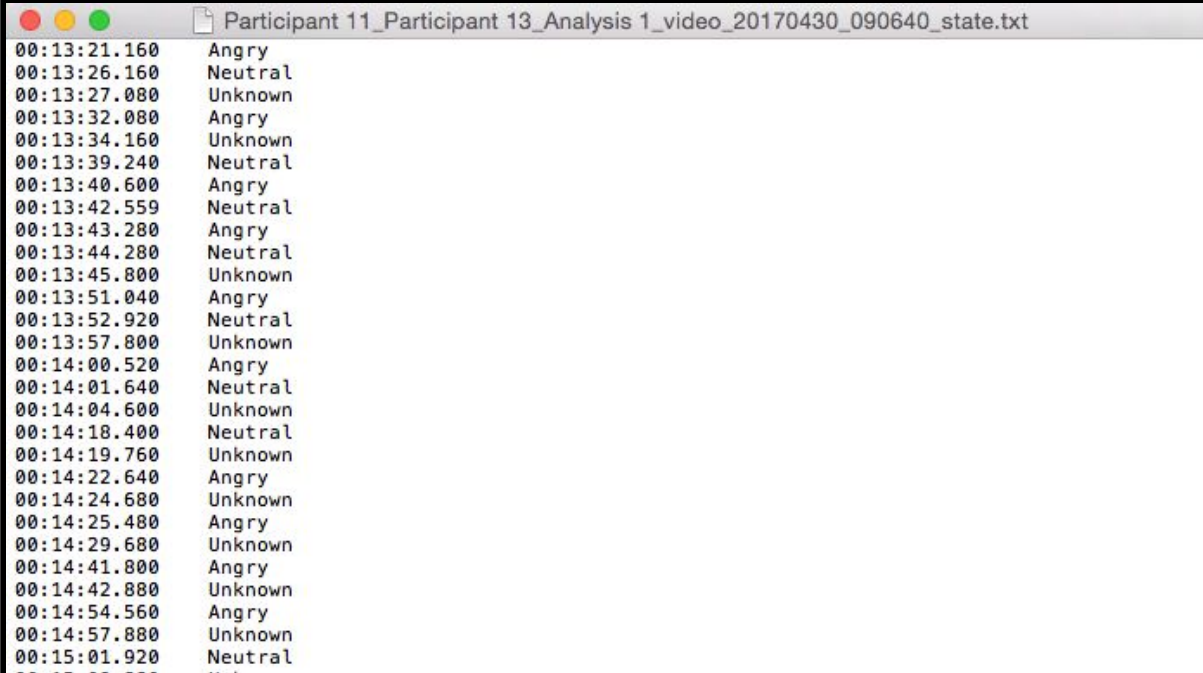
1. Input skapas i form av ett dokument med användarens interaktionsdata från scenario A+B.
2. I denna modul tokeniseras texten till mindre beståndsdelar. Detta för att kunna utföra sentimentanalys.
3. Den här modulen innehåller databasen WordNet.
4. Multiply-modulens funktion är att kombinera input-modulen (här WordNet) till alla moduler som är kopplade till output-portarna (här modulerna för sentimentextrahering).
5. Input skapas i form av ett dokument med användarens interaktionsdata från scenario C.
6. I denna modul tokeniseras texten från modulen i föregående steg.
7. Ett sentimentvärde extraheras ur interaktionsdatan från scenario A+B med hjälp av WordNet- och SentiWordNet-databaserna.
8. Ett sentimentvärde extraheras ur interaktionsdatan från scenario C med hjälp av WordNet- och SentiWordNet-databaserna.
9. Ett dataset genereras utifrån dokumentet som utvärderats i steg 7.
10. Ett dataset genereras utifrån dokumentet som utvärderats i steg 8.
11. Append-modulen bygger slutligen samman dataseten till en komplett tabell.

3.4 Analys av video

Varje session filmades. Videomaterialet analyserades i programmet FaceReader 6.1. FaceReader angav resultaten i en lista där den dominanta emotionen samt vilken tid den uppstod visades (se fig. 5). Den totala mängden registrerade ansiktsuttryck antecknades och jämfördes med i hur många fall FaceReader misslyckades med att uttröna en emotion ur datan.

“CAN’T YOU READ I’M FRUSTRATED?”

Antalet registrerade negativa emotioner sammanställdes genom att den tid varje deltagare lade på scenario C beräknades, varpå antalet negativa emotioner som deltagaren uppvisade under detta tidsspänn beräknades. Även antalet negativa emotioner som FaceReader registrerade för det tidsspänn deltagaren lade på scenario A+B beräknades. Dessa värden jämfördes sedan i ett beroende *t*-test.



Timestamp	Emotion
00:13:21.160	Angry
00:13:26.160	Neutral
00:13:27.080	Unknown
00:13:32.080	Angry
00:13:34.160	Unknown
00:13:39.240	Neutral
00:13:40.600	Angry
00:13:42.559	Neutral
00:13:43.280	Angry
00:13:44.280	Neutral
00:13:45.800	Unknown
00:13:51.040	Angry
00:13:52.920	Neutral
00:13:57.800	Unknown
00:14:00.520	Angry
00:14:01.640	Neutral
00:14:04.600	Unknown
00:14:18.400	Neutral
00:14:19.760	Unknown
00:14:22.640	Angry
00:14:24.680	Unknown
00:14:25.480	Angry
00:14:29.680	Unknown
00:14:41.800	Angry
00:14:42.880	Unknown
00:14:54.560	Angry
00:14:57.880	Unknown
00:15:01.920	Neutral

Fig. 5: Lista över dominant emotion som FaceReader 6.1 utrönat vid redovisad tidpunkt i videofilen.

3.5 Analys av frågeformulär

Svaren från frågeformuläret som samtliga deltagare fick besvara (se bilaga D) sammanställdes och medelvärden beräknades där det var lämpligt.

3.6 Pilotstudier

Två pilotstudier utfördes i syfte att arbeta fram en lämplig experimentmetod. Här diskuteras båda med avseende att motivera val av slutgiltig metod, det vill säga, varför vi valde att använda oss av scenarion framför exempelvis en fri interaktion.

I den första pilotstudien försågs två deltagare med uppgifter liknande de i bilaga C. Ett liknande manus som det i bilaga B användes av försöksledaren som simulerade chatboten. Deltagarna tycktes gå med på illusionen att de pratade med en chatbot, dock gick det ej att utläsa några tecken på negativa emotioner i den text som genererats med hjälp av den analysmetod som användes. Analysmetoden gav ett resultat som enbart beskrev textens beräknade polaritet, det vill säga huruvida texten var positiv, negativ eller objektiv. Hur stark polariteten var framgick dock inte. En deltagare uppgav att hen valde att inte uttrycka sin frustration på grund av en medvetenhet om att försöksledarna skulle läsa konversationen i efterhand, vilket ledde till ytterligare osäkerhet kring metoden. Illusionen av att de interagerade med en chatbot lyckades dock så pass väl att en av deltagarna valde att avsluta

“CAN’T YOU READ I’M FRUSTRATED?”

scenariot med chatboten (skriva “*END*”, se bilaga C) även efter att försöksledaren återvänt och avbrutit sessionen.

En andra pilotstudie utfördes för att se om en friare interaktion skulle ge ett annat resultat. Experimentdeltagaren uppmanades att chatta fritt med chatboten, samt att utföra ett antal bokningsrelaterade uppgifter. Eftersom ett manus blev svårt att arbeta fram till en så pass fri interaktion blev det svårt för försöksledaren som simulerade chatboten att ge adekvata och konsekventa svar på experimentdeltagarens frågor. Slarvfel i form av stavfel begicks, och deltagaren undrade efteråt om det varit försöksledarna hen chattat med. Det konstaterades att en omarbetning av pilotstudie 1, i kombination med en känsligare analysmetod, skulle vara mest fördelaktigt för studien.

4. Resultat

Till experimentet rekryterades 16 individer (4 kvinnor och 12 män) i åldrarna 24-34. Den genomsnittliga tiden för att slutföra scenario A+B var 15,5 minuter per deltagare. För scenario C var samma tid 14,3 minuter. Den totala snitttiden för interaktion med chatboten var således 29,8 minuter. En analys av skärminspelningarna visade att samtliga deltagare höll sig till sin givna uppgift under hela tiden experimentet pågick.

4.1 Resultat från sentimentanalys

Den genomsnittliga mängden ord som varje deltagare kommunicerade till chatboten i scenario A+B var 179, med ett minimumvärde på 93 och ett maximumvärde på 328. I scenario C var den genomsnittliga mängden ord 170, med ett minimumvärde på 82 och ett maximumvärde på 238. Resultatet från sentimentanalysen redovisas i fig. 6.

Det beroende *t*-testet tyder på att det finns en statistiskt signifikant skillnad i sentimentvärde mellan de olika scenarierna, där sentimentvärdet är högre vid scenario A+B ($M = 0,10$, $SD = 0,47$) och lägre vid scenario C ($M = -0,37$, $SD = 0,081$), $t(15) = 6,66$, $p < 0,001$, $d = 1,67$.

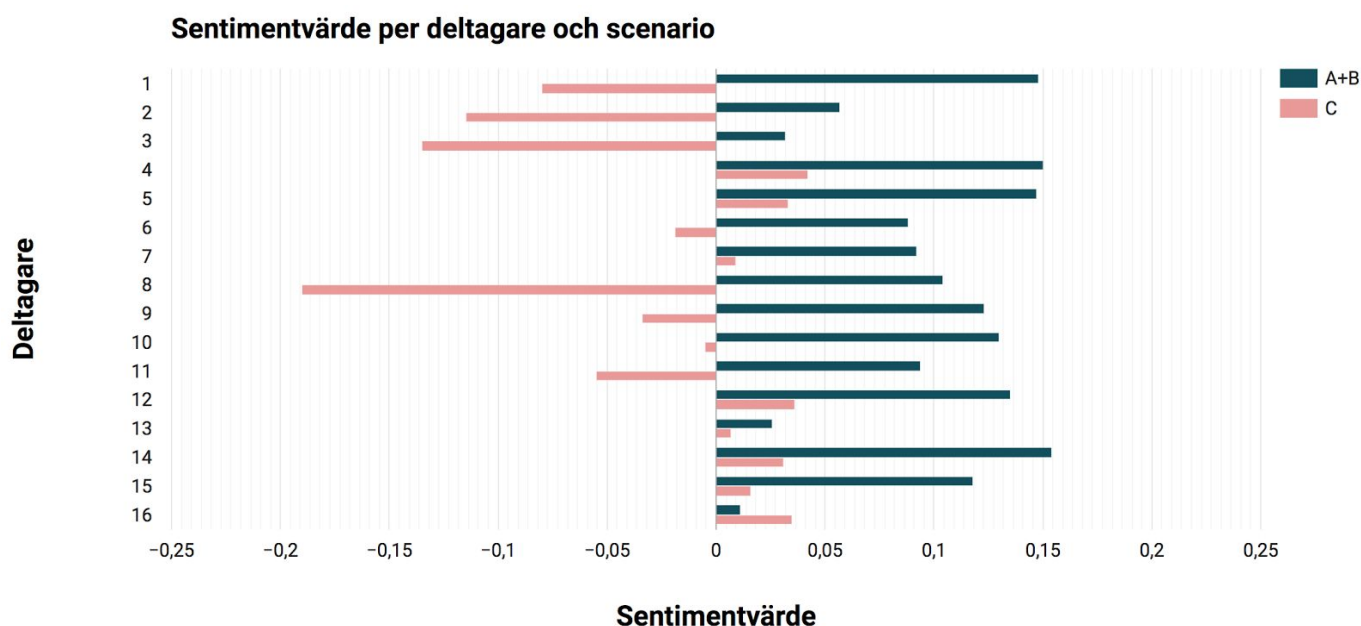


Fig. 6. Graf över resultat från sentimentanalysen. I grönt visas sentimentvärdet från respektive deltagares input från scenario A+B. I rosa syns sentimentvärdet från respektive deltagares input från scenario C.

4.2 Resultat från videoanalys

Två av de 16 deltagarnas videoupptagningar misslyckades. Totala mängden videoupptagning uppgick till 7,2 timmar. Den totala mängd ansiktsuttryck som FaceReader registrerade var 4721, varav 1649 (34,9%) registrerades som ‘unknown’.

“CAN’T YOU READ I’M FRUSTRATED?”

Det beroende *t*-testet tyder på att det inte finns någon statistiskt signifikant skillnad mellan antalet uppvisade negativa ansiktsuttryck i scenario A+B ($M = 15,64$, $SD = 29,60$) jämfört med scenario C ($M = 16,29$, $SD = 31,03$), $t(13) = -0,24$, $p = 0,82$, $d = 0,064$.

4.3 Resultat från frågeformulär

Tre deltagare uppgav engelska som modersmål. 15 deltagare uppgav att de erfarit en negativ emotion under experimentets gång och 14 av dessa specificerade denna som frustration. Ytterligare emotioner som angavs var ‘ilska’, ‘förakt’ och ‘utmattning’. På frågan kring hur svårt deltagarna ansåg det vara att interagera med chatboten samt få instruktioner/uppgifter på engelska blev det genomsnittliga svaret för gruppen 2,4 (1 = *easy*, 5 = *hard*). På frågan kring hur vana deltagarna var vid onlinebaserade bokningssystem blev det genomsnittliga svaret för gruppen 3,3 (1 = *less*, 5 = *more*). På frågan kring hur vana deltagarna var vid liknande chatbotar blev det genomsnittliga svaret för gruppen 1,9 (1 = *less*, 5 = *more*).

5 Diskussion

Nedan följer avsnitt där studiens utfall, procedur och utformning diskuteras. Vi introducerar även termen ‘botparadoxen’, ett metodologiskt dilemma som vi behövde förhålla oss till eftersom chatboten simulerades av en individ. Vi ger också förslag på vad som skulle kunna förbättras i studien samt föreslår ämnen till framtida studier.

5.1 Analys av text

Den text som skulle genomgå analys togs fram genom att radera allting chatboten skrev i varje deltagares interaktion, så att endast deras egen input analyserades. Eftersom WordNet och SentiWordNet är engelskspråkiga databaser var interaktionen med chatboten tvungen att ske på engelska (Baccianella et al., 2010; Princeton University, 2015). Vissa ord som föll sig naturliga att använda i interaktionen i och med uppgifternas utformning (exempelvis ‘cancel’, ‘remove’ och ‘allergy’) har en positiv eller negativ laddning i SentiWordNet, och skulle därmed kunna ha påverkat resultatet. Detta bedömdes dock inte ha varit en risk i och med två faktorer. Ord av liknande slag förekommer i alla tre scenarier och därmed bör denna faktor vara utbalanserad. Dessutom är SentiWordNet utformat till att ta hänsyn till en viss grad av kontextualitet i sin analys. Detta eftersom varje ord i databasen WordNet, som SentiWordNet baserats på, tillhör olika *synsets*, vilka är kategorier där det framgår vad för laddning ordet i fråga har beroende på semantisk kontext (Princeton University, 2015).

En viss risk vid sentimentanalys av den här typen står dock kvar; att somliga element möjligen misstolkas eller helt enkelt inte plockas upp på grund av att kontexten feltolkas i SentiWordNet. Vissa ord kan ha flera olika betydelser beroende på kontext som möjligen faller utanför analysen. Andra språkliga fenomen så som ironi, sarkasm, emoticons eller metaforer kan möjligen bidra till ett missvisande resultat i sentimentanalysen.

Det framgår inte huruvida SentiWordNet tar hänsyn till ord skrivna i enbart versaler och i så fall hur sådana ord tolkas, dock var detta så pass sällsynt (18 stycken, av totalt 5586 ord) att det inte anses ha påverkat resultatet. Ett fåtal emoticons förekom, varav en (automatiskt, av ordbehandlingsprogrammet) omtolkades till “*slightly smiling face*” när den omformaterades till text.

5.1.1 Resultat från sentimentanalys

En viss variation observerades i resultaten när antalet ord per deltagare och scenario utvärderades, vilket var föga förvånande med tanke på att Hill et al. (2015) i sin studie visar att vi tycks kommunicera med kortare, om än fler, meningar vid interaktion med en chatbot än vid chattbaserad kommunikation med en annan individ. Variationen uppstod i denna studie utan att det genomsnittliga värdet på tid per scenario varierade särskilt mycket alls (15,5 minuter för A+B respektive 14,3 för C). Möjligen berodde detta på att vissa experimentdeltagare försökte väcka respons eller funktioner hos chatboten genom att presentera nyckelord för denne, snarare än att skriva en komplett mening. Det faktum att vi erhöll ett resultat från sentimentanalysen som var av statistiskt signifikans talar för att ordantalet inte är av särskilt hög relevans vid sentimentanalys. Dock är detta något att ta i

“CAN’T YOU READ I’M FRUSTRATED?”

beaktande vid skapandet av de algoritmer som skulle vara bakomliggande sentimentanalysen respektive chatbotens respons på denna vid implementering.

Det beroende *t*-testet pekar på att implicita tecken gällande användarens emotionella tillstånd framgår i sentimentanalysen. Med undantag för deltagare 16 (se fig. 6) blev sentimentvärdet för scenario A+B högre än för scenario C för samtliga försöksdeltagare. En möjlig förklaring till varför värdet för deltagare 16 avvek skulle kunna vara bruk av ironi, men eftersom detta faller utanför ramarna för vad vår undersökning innefattar har vi valt att inte utreda det närmare. Resultaten från sentimentanalysen visar ett maximumvärde på 0,154 och ett minimumvärde på -0,19 (se fig. 6). Genom en jämförelse med de sentiment- värden som genererades med hjälp av exempeltexterna (se avsnitt 2.4.1) framgår det hur känslig analysen är, och att små skillnader plockas upp.

5.2 Resultat från ansiktsuttrycksanalys

En analys av skärminspelningarna visade att samtliga deltagare höll sig till sin givna uppgift. Här kunde vi också observera ett flertal fall av *message construction repair*. Ett försök att kontrollera för huruvida detta eller liknande fenomen påverkade utfallet i sentimentanalysen gjordes med hjälp av videoinspelning och ansiktsuttrycksanalys. Det faktum att FaceReader återgav 34,9% av de uppfattade ansiktsuttrycken som ‘*unknown*’ talar dock för att analysen inte var särskilt framgångsrik. Om videoinspelningarna kan sägas att både kvaliteten och andra faktorer som vi ej hade möjlighet att kontrollera för varierade från deltagare till deltagare. Både skärpa och hur ofta en deltagare kollar nedåt eller åt sidan (varpå deltagarens ansiktsdrag rimligtvis inte längre uppfattas av FaceReader) skiljer sig mellan de olika videofilerna, trots upprepade försök att kontrollera dessa faktorer (kontrollera kamera-inställningarna, höja/sänka kamerastativ och datorskärm, samt uppmana deltagarna att sitta rakt).

Den deltagare som hade lägst antal procent ‘*unknown*’ bland sina rapporterade ansiktsuttryck hade 17%, och den deltagare som hade högst antal fick 45% av sina ansiktsuttryck rapporterade som *unknown*. Den deltagare som enligt FaceReader uppvisade flest antal negativa emotioner hade ‘*angry*’ som 202 av sina 540 rapporterade ansiktsuttryck, av vilka totalt 25% var *unknown*. Ur detta kan två potentiella slutsatser argumenteras för, där den ena talar för mjukvarans precision och den andra talar emot. Att FaceReader vid så pass många tillfällen lyckades utläsa denna specifika negativa emotion ur deltagarens ansikte kan tyckas exceptionellt. Dock ställer vi oss kritiska till huruvida deltagaren faktiskt upplevde emotionen i fråga så pass frekvent under experimentets gång som analysen pekar på. Här tycks det snarare handla om att individen i fråga har ett neutralt ansiktsdrag som FaceReader tolkar som *angry*. En sådan tolkning av programmet kan bero på att ansiktet inte riktas rakt mot kameran, och att FaceReader då får svårigheter att läsa av ansiktsuttrycket på korrekt sätt. Dessutom kan subtila signaler, som att individens läppar drar en aning bakåt eller att individen kisar med ögonen, bidra till att vilseleda programmets tolkning. Detta talar för att den kalibrering som rimligtvis behövs för att väga upp för hur individuella ansiktsdrag påverkar hur mjukvaran tolkar dem misslyckades, vilket i sin tur leder till ett ifrågasättande av samtliga resultat från ansiktsuttrycksanalysen (Gibson, 2016, 2 Februari).

“CAN’T YOU READ I’M FRUSTRATED?”

5.3 Deltagare

Till en början var avsikten att endast studera interaktion mellan chatboten och individer som hade engelska som modersmål, eller som talade/kommunicerade via text på engelska dagligen. Att inkludera IT-studenter i studien ville undvikas eftersom dessa rimligtvis har en bättre uppfattning kring hur en chatbot fungerar, något vi hypotiserade skulle ha en icke önskvärd effekt på resultatet. På grund av studiens tidsmässiga omfattning blev vi dock tvungna att frånga dessa önskemål, för att kunna komma upp i det antal deltagare som önskades. Hur detta samt faktorer som ålder och könsmissig spridning kan ha påverkat resultatet hamnar utanför ramarna för denna studie, men kan vara värt att undersöka i ett senare skede.

5.4 Botparadoxen

Vid utformning av chatbotens manus och profilbild blev det relevant att reflektera kring hur nödvändigt det egentligen var med en ‘robotisk framtoning’ för att åstadkomma trovärdighet hos simuleringen. Vi kom att kalla denna problematik för ‘botparadoxen’. Botparadoxen handlar alltså om problematiken kring chatbotens trovärdighet som chatbot, eftersom den i vår studie simuleras av en individ. Å ena sidan är det eftersträvansvärt att försöksdeltagarna tror att de chattar med ett automatiserat system, eftersom detta är en nödvändig premis för att mäta det vi ville mäta. Å andra sidan är chatbotar numera så pass kompetenta att även en mer ‘mänsklig’ interaktion går att tillskriva en chatbot, och därmed blir frågan mindre relevant. Således kontrollerar vi inte för huruvida deltagarna trodde att de chattade med ett system i frågeformuläret. Ett positivt svar på en fråga i stil med “Fick du någon gång under experimentet intrycket av att du chattade med en människa?” skulle i ett fall då interaktionen sker med en faktisk chatbot endast betyda att chatboten uppnår det den syftar till att göra - att efterlikna mänsklig kommunikation i så hög utsträckning som möjligt (mer om detta i Fitriani et al. 2003; Hingston, 2012; Jenkins, Churchill, Cox & Smith, 2007).

De chatbotar som finns i olika former och forum är i dagsläget så pass naturtroga och utvecklade att människor generellt inte ifrågasätter att en chatbot är en chatbot. Som ett exempel klarade chatboten Cleverbot Turingtestet år 2011, då 59,3% av de 1334 människor som interagerat med chatboten under fyra minuter trodde att den i själva verket var en människa (jämfört med att 63,3% av deltagarna trodde att människorna de interagerade med var faktiska människor) (Aron, 2011). I vårt experiment användes till stor del för- bestämda repliker av chatboten, men även en del improviserade sådana, vilket ansågs bidra till en rimlig balans av automatiserade svar och flexibilitet hos dess förmåga till interaktion.

5.5 Chatbotens profilbild

Hingston (2012) skriver: “När en ovan användare behöver interagera med en robot för första gången är det första intrycket av största vikt för att på ett framgångsrikt sätt kunna initiera och uppehålla interaktionen. Därför är det viktigt att utseendet på roboten matchar med dess uppgift för att öka dess trovärdighet.” (s. 2) (egen översättning). Med detta i åtanke lät vi en grafisk designer utforma chatbotens profilfoto (se fig. 2). I ett verkligt gränssnitt skulle en vidare avvägning behöva göras kring utseendet. Här var vi ute efter att åstadkomma en

“CAN’T YOU READ I’M FRUSTRATED?”

illusion av att det var en chatbot och inte en människa som deltagaren interagerade med, vilket valet av profilbild antyder.

Dock uppstår en frågeställning kring vad detta uppmanar till för sorts kommunikation. Hill med kollegors (2015) tidigare omnämnda studieresultat visar på att det finns en skillnad i ordantal beroende på huruvida en användare chattar med en annan individ eller med en chatbot, där användaren tenderar att uttrycka sig med fler meningar innehållande färre ord vid chatbotkommunikation. I ett fall där det är önskvärt att använda sig av realtidsbaserad sentimentanalys kan det således vara relevant att användaren tror att den interagerar med en annan människa. Skulle en sådan avvägning göras är det dock viktigt att chatboten lyckas leva upp till det en mänsklig profilbild antyder, det vill säga, kommunicerar på det vis en människa skulle ha gjort i en liknande situation. En interaktion där chatboten inte lever upp till användarens förväntningar sker på bekostnad av dennes tillit till gränssnittet (Hingston, 2012).

5.6 Interaktion med chatboten

Chatboten Cleverbot, som vi till viss del hämtat inspiration ifrån, är utformad så att det står “*thinking...*” under tiden den input användaren lämnar ifrån sig bearbetas, vilket skulle kunna vara önskvärt att efterlikna vid en implementering av den chatbot som behandlas här. Detta för att återkoppla till användaren att hans input är under bearbetning, utan att användaren upplever att det dröjer. Medan Cleverbot ‘tänker’ går det dessutom inte att skriva något till den, vilket gör att den endast behöver bearbeta en input åt gången. Även denna funktion kan vara önskvärd att efterlikna i ett faktiskt gränssnitt eftersom det förhindrar att användaren snabbt skickar två meddelanden efter varandra, som var och ett framkallar en specifik respons från systemet. I de fall detta skedde under interaktionselementet i vår studie svarade den försöksledare som simulerade chatboten antingen på en av fraserna, eller på båda efter varandra. Likt Cleverbot noterade chatboten inte heller några av användarnas stavfel (förutom i scenario C, där sådana direkt uppmärksammandes i syfte att frustrera). Detta var ett medvetet val som grundade sig i att en faktisk chatbot av det slag vi föreslår rimligtvis har en ordlista som korrigerar för stavfel, synonymer och slangord (Fitriani et al., 2003).

Vid ett flertal tillfällen gick deltagarna utanför sina givna uppgifter, och med hjälp av förberedda fraser (exempelvis “*Of course, give me a second and I will look for alternatives.*”) kunde ett passande svar arbetas fram i ett separat dokument, där försöksledaren hade möjlighet att kontrollera för stavfel i förväg. Detta var förväntat med tanke på att instruktionerna uppmanade till kreativa lösningar, samt att deltagarna ombads att själva lista ut hur de skulle slutföra uppgifterna.

Det ingick i instruktionerna att avsluta varje avklarat scenario genom att skriva “*END*” till chatboten. Sju av 16 experimentdeltagare valde att avsluta interaktionen med att skriva “*END*”, även efter det att försöksledare 2 kommit in i rummet för att avbryta sessionen. Detta tolkar vi som ett tecken på att försöket att åstadkomma en illusion av att de interagerade med en chatbot (och inte en människa) varit framgångsrikt. Eftersom vi medvetet ej kontrollerar för huruvida deltagarna accepterade illusionen eller inte, kan vi inte diskutera denna faktor bortom implicita resultat. Till implicita resultat räknar vi också det

“CAN’T YOU READ I’M FRUSTRATED?”

faktum att ingen av deltagarna efteråt frågade om det varit en människa de kommunicerat med, vilket deltagaren i pilotstudie 2 gjorde.

5.6.1 Manipulation av interaktioner

Genom att använda scenarier på det vis vi redogör för i metodavsnittet kunde interaktionen på ett smidigt sätt styras till att först vara framgångsrik, för att sedan gå mot att deltagaren misslyckades med sin givna uppgift. Detta i linje med studier som visar att negativa emotioner, och specifikt frustration inom människa-datorinteraktion, uppstår när ett mål hotas och en plan avbryts efter att först ha varit framgångsrik (Bessière et al., 2004; Freud, 1912; Oatley & Johnson-Laird, 1987).

Scenarierna bestod av ett antal uppgifter som alla sades vara av ungefär samma komplexitet, och som alla hade samma instruktioner (se bilaga C). Efter att scenario A gick helt smärtfritt (chatboten svarade snabbt, hjälpsamt och effektivt) förväntades användaren vara tillfreds med interaktionen, eftersom målet att slutföra uppgifterna var utom märkbara hot. Förväntningarna på systemet var här intakta eftersom alla uppgifter gick att lösa utan problem.

I scenario B gick interaktionen lika bra, bortsett från att chatboten ställer aningen fler motfrågor, exempelvis i ett fall för att kontrollera huruvida användaren menar “*Royal Castle Bar & Restaurant*” när denne frågar om öppettider till det kungliga slottet i Stockholm (se bilaga E). Detta eftersom chatboten fortfarande skulle vara hjälpsam och målet fortfarande skulle gå att uppnå, om än med något mer friktion än i scenario A. Syftet här var att etablera en känsla av tillit till chatbotens problemlösningsförmåga (se till exempel Bessière et al., 2004; Hingston 2012).

I scenario C sker en markant skillnad i interaktionens smidighet. Chatboten missförstår förfrågningar, skickar reklam till användaren, och mailar olika mottagare utan att först få användarens godkännande till detta. Systemets prestation går här under deltagarnas förmodade förväntningar, vilket torde väcka frustration (Charlton, 2009). Sentimentanalysen pekar på en skiftning av textvärde i negativ riktning under tidsspannet för detta scenario, och flertalet deltagare nämnde i frågeformuläret specifikt scenario C som källan till deras negativa emotion.

5.7 Brister i studien

Nedan följer en diskussion kring studien gällande metodologisk utformning, analysrelaterade tillkortakommanden, och vad den mänskliga faktorn kan ha haft för inverkan. Diskussionen förs med avseende att reda ut hur dessa faktorer kan ha påverkat resultatet.

5.7.1 Miljö och procedur

Miljön där experimentet ägde rum var i viss utsträckning onaturlig - försöksdeltagarna befann sig i en främmande miljö och filmades av en kamera som var riktad mot ansiktet under hela tiden experimentet pågick. Dock är också de flesta omständigheter naturliga för uppgiften i fråga. Deltagarna satt placerade vid en dator och fick interagera med ett webbaserat gräns-

“CAN’T YOU READ I’M FRUSTRATED?”

snitt. I instruktionerna fick de veta att gränssnittet var under prototyputveckling, eftersom ett utforskande förhållningssätt till chatboten skulle uppmuntras.

Deltagarna hade innan inspelningen började givit sitt godkännande till filmupptagningen, vilket skulle kunna ha påverkat deras självmedvetenhet och således i hur stor grad de uttryckte sig med emotionella ansiktsuttryck. I de instruktioner de fick ta del av påstods det att all analys skulle ske med hjälp av automatiserad mjukvara, som ett försök att avdramatisera detta. Deltagarna uppmanades också till att *“Treat the chatbot as your assistant!”* (se bilaga B).

5.7.2 Analysmetod

En närmare analys av videomaterialet från ansiktsuttrycksanalysen tyder på att det finns ett samband mellan hur ofta individen kollar nedåt eller åt sidan och hur ofta den analytiska mjukvaran misslyckas med sin tolkning. Analysen hade möjligtvis dragit fördel av att videoinspelningen istället skett med den sortens kameror som finns integrerade i datorskärmar, eftersom dessa rimligtvis skapar en bättre kameravinkel med större potential att fånga samtliga ansiktsdrag.

Eftersom ansiktsuttrycksanalysen, där vi förväntade oss ett resultat som skulle gå att korrelera med resultatet från sentimentanalysen, inte var framgångsrik kan vi också konstatera att vi borde ha ställt fler frågor gällande emotionella tillstånd samt när under experimentet de erfarades. Deltagare 11 uppgav sig inte erfarit någon negativ emotion, men har ändå ett lägre sentimentvärde i C vid en jämförelse med A+B (se fig. 6). Deltagare 16 har i sin tur ett högre sentimentvärde i C, trots att hen uppgav att frustration uppstod. De förmodade anledningarna till detta faller utanför ramarna för vad vi valt att undersöka, och kan därför inte diskuteras i relation till resultatet på ett trovärdigt sätt. En teori vi lägger fram i bakgrunden, gällande hur en individ möjligen uppvisar en större tendens att förmedla sitt emotionella tillstånd vid interaktion med ett system som uppfattas inneha intentioner (se avsnitt 2.2.2), erbjuder dock en potentiell förklaring. En ökad förståelse för hur ett automatiserat system av detta slag fungerar torde nämligen minska antalet förmedlade emotioner, eftersom en sådan användare rimligtvis har en god insikt i hur chatboten faktiskt frambringar sin respons.

5.7.3 Mänskliga faktorn

Den mänskliga faktorn orsakade ett par brister. Detta då manuset för chatboten vid ett tillfälle inte stämde överens med deltagarens instruktioner. Deltagaren reagerade på detta under tiden sessionen pågick, och bad chatboten att korrigeras i sina inställningar för respektive scenario. Eftersom detta skedde i scenario C lades ingen större vikt vid felet av försöksledarna, på grund av att interaktionen här skulle gå mot att chatboten inte uppfyllde deltagarens förväntningar. Två inspelningar föll bort på grund av tekniska problem under experiment-sessionen, vilket resulterade i 14 videoupptagningar från experimentet istället för 16, som var målet.

Ett försök att etablera tillit samt att efterlikna hur en faktisk chatbot skulle gå tillväga för att ge en passande respons på input, var att återspegla det ord som deltagaren använde sig

“CAN’T YOU READ I’M FRUSTRATED?”

av i sin formulering (Fitrianie et al., 2003). Det vill säga, i ett fall då användaren till exempel använder sig av ordet *‘hotel’* istället för *‘venue’* när hen har för avseende att göra ändringar i hotellbokeningen, borde chatboten anpassa sitt språk efter det begrepp användaren använt. Försök att göra detta gjordes i flera fall, dock ej så pass konsekvent att det kan påstås ha haft någon effekt, vilket eventuellt är en brist i metoden. I detta avseende kan valet att ha använt ett färdigt manus för chatboten ses som en nackdel.

På grund av en formulering i frågeformuläret, där försöksdeltagaren ombeds besvara frågan “*How easy was it to communicate in english during this task (before and during the interaction with the bot)?*” (se bilaga D), ifrågasätter vi snittresultatet 2,4. Några av de deltagare som hade engelska som modersmål angav högsta värde på skalan, som i detta fall gick från 1 till 5. Eftersom formuleringen lyder “*How easy was it...*” anser vi att det kan finnas en risk att deltagarna missuppfattat skalangivelsen och fyllt i en högre siffra för att ange enkelhetsgrad snarare än svårighetsgrad. Syftet med frågan var att, i ett fall då sentimentanalysen inte gav ett förväntat resultat, undersöka om det fanns ett samband mellan språklig förmåga och emotionellt tillstånd förmedlat i text. På grund av det resultat vi erhöll, samt hur interaktionerna såg ut när de väl genomfördes, ser vi dock ingen anledning att diskutera denna aspekt bortom det faktum att frågeformuläret kunde ha utformats bättre.

5.8 Realiserbarhet utanför studien

Den frågeställning som undersöks syftar till att att besvara huruvida ett automatiserat system har möjlighet att utröna användarens emotionella tillstånd endast utifrån det som kommuniceras via chatt. Studiens metod var begränsad i relation till syftet i och med att det varken fanns möjlighet att programmera en faktisk chatbot, eller att utföra sentimentanalysen i realtid. Den första av dessa faktorer anser vi oss kunna avfärda som ett icke-problem (se avsnitt 5.5). Den andra faktorn är något svårare att kringgå, men eftersom implementering ej behandlas i denna studie har vi heller inte möjlighet att gå djupare in på ämnet. Om detta kan dock sägas att studiens resultat visar på en stor potential hos SentiWordnet och dess framtida användningsområden i relation till automatiserade system.

5.9 Framtida studier

Som vi påpekar ovan finns det ett antal studier som tyder på att ett gränssnitts utseende påverkar hur vi väljer att kommunicera med det (se till exempel Hingston, 2012). Vid en situation där målet är att chatboten ska kunna upprätthålla en så människolik konversation som möjligt bör gränssnittet utformas därefter. Således kan det vara intressant att undersöka hur olika antropomorferade gränssnittelement påverkar sentimentvärdet i den text som användaren genererar vid interaktion. En vidare undersökning av hur olika personlighetsdrag ur både användar- och chatbotperspektiv påverkar interaktionen kan också vara intressant.

En annan faktor som vore intressant att undersöka i ett skede då chatboten står färdig att använda, är hur målgruppen använder en sådan tjänst. Den chatbot som står i fokus här undersöks parallellt med Mohamed Abdelnabys och Claudia Bertones arbete med att ta fram en prototyp för en komplett konferensbokningstjänst. Dessa bör därför integreras och ut-

“CAN’T YOU READ I’M FRUSTRATED?”

värderas av individer som bokar konferenser. Vidare bör det även undersökas hur frustrerad en användare kan bli innan den väljer att överge gränssnittet, alternativt innan tilliten försvinner, och vad systemets åtgärd blir när denna gräns uppnås. En chatbot skulle exempelvis kunna slussa vidare användaren till bemannad kundtjänst, men hur och när under interaktionen detta bör ske lämnas åt framtida studier.

En optimal textanalys av den interaktion som sker mellan användare och chatbot borde ske i realtid och ha ett kontinuerligt uppdaterat värde för konversationens sentiment. På så vis kan förändringar i språket och därmed även emotioner och attityder detekteras, och systemet kan med hjälp av den datan skapa en mer framgångsrik interaktion genom att aktivt förebygga och följa upp dessa förändringar. Dessutom borde analysen utvecklas till att involvera ironi, sarkasm, emoticons, metaforer och (högre grad av) kontext.

Eftersom WordNet 3.0 och SentiWordNet 3.0 är engelskspråkiga databaser är vår sentimentanalys begränsad till engelskspråkig textdata. En liknande databas på svenska kan vara aktuell att utveckla, beroende på var sentimentanalysen ska implementeras. Databasen kan med fördel även vidgas till att inkludera ett specialiserat fackspråk, där varje ord får ett värde baserat på sin betydelse i den kontext den ska användas.

Detta väcker frågor kring hur implementering av ett sådant system bör se ut. Ett möjligt svar på detta ges i Fitriani med kollegors studie från 2003, där ett förslag till ett liknande systemramverk lyfts fram. Andra aspekter som kan vara av relevans att ta upp vid implementering är hur en långvarig relation mellan användare och chatbot byggs upp. Om detta är målet kan det vara värt att undersöka hur den typen av relationer påverkas av interaktionsstrategier som till exempel *‘tit for tat’* (mer om detta i till exempel Clavel & Callejas, 2016; Hingston, 2012).

I ett skede där ansiktsuttrycksanalys ska implementeras som ett komplement till sentimentanalys behöver undersökningar göras gällande hur systemet ska behandla de emotionella uttryck som eventuellt plockas upp. Detta blir viktigt eftersom en användare, som Klingberg (2009) påpekar, tenderar att kontinuerligt skifta mellan de flikar hen väljer att ha öppna i en webbläsare. Således väcks frågor kring hur systemet ska kunna veta vad som gav upphov till vilken emotion. Dessutom behöver en avvägning ske gällande analytisk mjukvara och användarens datorkapacitet, eftersom mjukvaror av denna typ kan vara krävande för en processor.

“CAN’T YOU READ I’M FRUSTRATED?”

6 Slutsats

Studier har visat att en interaktion där användarens emotionella tillstånd tas i beaktande bidrar till en högre nivå av tillit mellan användare och gränssnitt. Ett sätt att hitta tecken på användarens emotioner är att via sentimentanalys söka efter ord med implicit emotionell laddning. Denna studie syftade till att undersöka huruvida ett automatiserat chattsystem har möjlighet att avgöra om interaktionen är framgångsrik eller ej, endast utifrån det som kommuniceras av en användare. Större mängder textdata analyserades i sin helhet, och resultatet tyder på att negativa emotioner är möjliga att detektera i den textbaserade data som genereras när en användare interagerar med en chatbot.

Försöket att vidga analysmetoden med ansiktsuttrycksanalys kan i detta fall inte ses som särskilt lyckat, vilket tyder på att FaceReader 6.1 inte lämpar sig som komplement till den realtidsbaserade sentimentanalys som skulle vara aktuell att implementera. Detta på grund av dess oförmåga att utifrån begränsad data kalibrera individuella ansiktsdrag korrekt, samt hur känslig mjukvaran tycks vara för inkompleta ansiktsdrag.

En chatbot liknande den vi undersökt kommer möjligen att implementeras i en webbaserad tjänst för konferensbokning. Vad vi visar på här är att det verkar genomförbart att skapa ett gränssnitt som känner av och i viss mån anpassar sig efter användarens emotionella tillstånd. I framtiden kommer gränssnitt som lyckas med detta att bidra till bättre förutsättningar för utvecklingen av de områden som relaterar till människa-datorinteraktion.

“CAN’T YOU READ I’M FRUSTRATED?”

Referenser

- Aron, J. (2011, 6 september). Software tricks people into thinking it is human. *New Scientist*. Hämtad 2017-04-27 från <https://www.newscientist.com/article/dn20865-software-tricks-people-into-thinking-it-is-human/>
- Averill, J. R. (2012). *Anger and aggression: An essay on emotion*. New York, NY: Springer.
- Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. *LREC, 10*, 2200-2204.
- Bessière, K., Newhagen, J. E., Robinson, J. P., & Shneiderman, B. (2006). A model for computer frustration: The role of instrumental and dispositional factors on incident, session, and post-session frustration and mood. *Computers in Human Behavior, 22*(6), 941-961. doi:10.1016/j.chb.2004.03.015.
- Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems, 46*, 109-132. doi:10.1016/j.knosys.2013.03.012
- Brennan, S. (1995). Interaction and feedback in a spoken language system: A theoretical framework. *Knowledge-Based Systems, 8*(2-3), 143-151. doi:10.1016/0950-7051(95)98376-H.
- Charlton, J. P. (2009). The determinants and expression of computer-related anger. *Computers in Human Behavior, 25*(6), 1213-1221. <https://doi.org/10.1016/j.chb.2009.07.001>
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives on socially shared cognition, 13*(1991), 127-149.
- Clavel, C., & Callejas, Z. (2016). Sentiment analysis: From opinion mining to human-agent interaction. *IEEE Transactions on Affective Computing, 7*(1), 74-93. doi:10.1109/TAFFC.2015.2444846.
- Derks, D., Fischer, A. H., & Bos, A. E. R. (2008). The role of emotion in computer-mediated communication: A review. *Computers in Human Behavior, 24*(3), 766-785. doi:10.1016/j.chb.2007.04.004.
- DMR. (2017). *40 Amazing Slack Statistics and Facts (March 2017)*. Hämtad 2017-05-10 från <http://expandedramblings.com/index.php/slack-statistics/>

“CAN’T YOU READ I’M FRUSTRATED?”

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6(3), 169-200. doi:10.1080/02699939208411068.

Facebook. (2017). *Stats*. Hämtad 2017-05-10, från <https://newsroom.fb.com/company-info/#statistics>

Fitrianie, S., Wiggers, P., & Rothkrantz, L. J. M. (2003). A multi-modal eliza using natural language processing and emotion recognition. *Paper presented at the International Conference on Text, Speech and Dialogue, 2807*, 394-399. doi: 10.1007/978-3-540-39398-6_56

Freud, S. (1921). *The standard edition of the complete psychological works of Sigmund Freud*. London: Hogarth Press.

Ghazi, D., Inkpen, D., & Szpakowicz, S. (2014). Prior and contextual emotion of words in sentential context. *Computer Speech & Language*, 28(1), 76. doi:10.1016/j.csl.2013.04.009.

Gibson, C. (2016, 2 Februari). Scientists have discovered what causes Resting Bitch Face. *Washington Post*. Hämtad 2017-05-05, från https://www.washingtonpost.com/news/arts-and-entertainment/wp/2016/02/02/scientists-have-discovered-the-source-of-your-resting-bitch-face/?utm_term=.eb62d136e492

Hingston, P. (2012). *Believable bots*. DE: Springer Verlag.

Hill, J., Randolph Ford, W., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, 49, 245-250. doi:10.1016/j.chb.2015.02.026.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, 57(2), 243-259.

Jenkins, M., Churchill, R., Cox, S., & Smith, D. (2007). Analysis of user interaction with service oriented chatbot systems. *Paper presented at the International Conference on Human-Computer Interaction, 4552(3)* 76-83. doi: 10.1007/978-3-540-73110-8_9

Jensen, M. (2016). *Lärande och lärandeteorier: Om den intentionella människan*. Lund: Studentlitteratur.

Klingberg, T. (2009). *The overflowing brain: Information overload and the limits of working memory*. Oxford: Oxford University Press.

“CAN’T YOU READ I’M FRUSTRATED?”

Linblom, J., Susi, T., & Tysk, A. (2012). Social Kognition. J. Allwood, & M. Jensen (Red.). *Kognitionsvetenskap* (s. 383-392). Lund: Studentlitteratur.

Noldus. (2017). FaceReader. Hämtad 2017-04-11, från <http://www.noldus.com/human-behavior-research/products/facereader>

Norman, D. A. (2013). *The design of everyday things: revised and expanded*. New York: Basic Books.

Meredith, J., & Stokoe, E. (2014). Repair: Comparing facebook ‘chat’ with spoken interaction. *Discourse & Communication*, 8(2), 181-207. doi:10.1177/1750481313510815.

Oatley, K., & Johnson-Laird, P. N. (1987). Towards a cognitive theory of emotions. *Cognition and Emotion*, 1(1), 29-50. doi:10.1080/02699938708408362.

Princeton University. (2015). *What is WordNet?*. Hämtad 2017-05-03, från <https://wordnet.princeton.edu/>

Pu, P., & Chen, L. (2007). Trust-inspiring explanation interfaces for recommender systems. *Knowledge-Based Systems*, 20(6), 542-556. doi:10.1016/j.knosys.2007.04.004.

Shawar, B. A., & Atwell, E. (2005). Using corpora in machine-learning chatbot systems. *International Journal of Corpus Linguistics*, 10(4), 489–516.

Smith, P., Lee, M., Barnden, J., & Hancox, P. (2011). *Sentiment analysis: beyond polarity* (Doctoral dissertation, Thesis Proposal, School of Computer Science, University of Birmingham, UK). Hämtad 2017-05-12, från <https://www.cs.bham.ac.uk/~pxs697/publications/documents/rsmg3.pdf>

Statistiska Centralbyrån. (2016). *Privatpersoners användning av datorer och internet 2016*. Stockholm: SCB, enheten för industriindikatorer, Fou och IT

Storbeck, J., Davidson, N. A., Dahl, C. F., Blass, S., & Yung, E. (2015). Emotion, working memory task demands and individual differences predict behavior, cognitive effort and negative affect. *Cognition and Emotion*, 29(1), 95-117. doi:10.1080/02699931.2014.904222.

Sundar, S. S., Bellur, S., Oh, J., Jia, H., & Kim, H. (2016). Theoretical importance of contingency in human-computer interaction: Effects of message interactivity on user engagement. *Communication Research*, 43(5), 595-625. doi:10.1177/0093650214534962.

Svenska Akademiens Ordlista. (2006). *Social*. Tillgänglig: http://sok.saol.se/Saol13_Sidor2.pl?ord=a&grad=M

“CAN’T YOU READ I’M FRUSTRATED?”

Bilaga A: Medgivandeblankett

Informed Consent Form

The Department of Applied IT at the University of Gothenburg and Chalmers supports the practice of protection of human participants in research. The following will provide you with information about the experiment that will help you in deciding whether or not you wish to participate. If you agree to participate, please be aware that you are free to withdraw at any point throughout the duration of the experiment.

This study will ask you to perform a couple of tasks using a chatbot-interface. Your participation in this study will require approximately 40 minutes. Video, screen activity and audio will be recorded and used only for research and educational purposes. All information will be kept confidential and your name, image, or sound will not be associated with any research findings.

When this study is complete you will be provided with the results of the experiment if you request them, and you will be free to ask any questions. It is important that you don’t share the contents of the experiment with other people, as they could be potential participants.

If you have any further questions concerning this study please feel free to contact us: Stina Ljungkvist or Evelina Larsson, email stina.lju@gmail.com or ewelarsson@gmail.com.

Please indicate with your signature below that you understand your rights and agree to participate in the experiment.

Date

Signature

Print name

“CAN’T YOU READ I’M FRUSTRATED?”

Bilaga B: Manus till försöksledare 2.

Hello and welcome to this study!

For practical reasons, this study will be held in English.

As a part of this study I’m going to record this session using this camera. The video, audio and text recorded will not be shown to anyone. We will use an automated software for the analysis later on.

If we have your permission to record this, please sign this paper.

You may at any time cancel the experiment, in which case we will delete all data recorded.

Your mission today is to try out and explore a prototyped interface. The interface is a chatbot that acts as an assistant for conference bookings.

I will provide you with a number of tasks to help you test the prototype. The tasks are all solvable in a number of different ways, and they are of about the same complexity. Treat the chatbot as your assistant! Feel free to find creative solutions; try to explore how the chatbot can help you.

Read through the instructions one page at a time and start whenever you’re ready!

I will give you about 30 minutes to interact with the chatbot, after which I will return and give you a questionnaire to answer. Have fun!

“CAN’T YOU READ I’M FRUSTRATED?”

Bilaga C: Instruktioner.

INSTRUCTIONS

The chatbot is pre-programmed with three different scenarios. Load each scenario by typing “**LOAD A**”, “**LOAD B**” or “**LOAD C**”.

The scenarios consist of booking details that you will need to change or find information about.

To complete your tasks, write to the chatbot and communicate the changes you want to make in your bookings.

When you have completed a scenario, type **END** and the chatbot will “forget” all previous interaction and you can move on to the next scenario.

All the tasks are of about the same complexity. Solving all the tasks should take about 30 minutes.

“CAN’T YOU READ I’M FRUSTRATED?”

Scenario A. Type “LOAD A” to the chatbot to load this scenario

Location: Gothenburg

Venue: Clarion Post

Number of days: 3

Price: 150 000 SEK

Number of guests: 700

Speakers: M. Danielsson (Stocks)

Special diet requests: 100 vegans, 147 vegetarians, 8 allergic to seafood

Other facts: Elevator, Swimming Pool

Task:

- Cancel the speaker.
- Find and book new speakers within topic “Trends in AI” and “Positive thinking/Mindfulness”.
- Find out if there is a sauna in connection to the swimming pool.
- Change the total number of guests to 770.
- Find out if venue has bicycle hire service.

Type END to finish, then load next scenario

“CAN’T YOU READ I’M FRUSTRATED?”

Scenario B. Type “LOAD B” to the chatbot to load this scenario

Location: Stockholm

Venue: Best Western Karlaplan

Number of days: 1

Price: 50 000 SEK

Number of guests: 100

Speakers: B. Loblaw (Law)

Special diet requests: 2 vegans, 15 vegetarians, 5 gluten intolerant

Other facts: No elevator, 4th floor

Task:

- A person in a wheelchair is attending the conference. Change the venue to one that is accessible by wheelchair.
- The conference will be set over two days instead of one. Change venue booking accordingly.
- Find out the new price.
- Find out the check-out time for the guests.
- A few of the attendees want to visit the Royal Castle while in town. What are the opening hours for the Royal Castle?

Type END to finish, then load next scenario

“CAN’T YOU READ I’M FRUSTRATED?”

Scenario C. Type “LOAD C” to the chatbot to load this scenario

Location: London

Venue: Hilton London Metropole

Number of days: 2

Price: 150 000 SEK

Number of guests: 40

Speakers: M. Johnson (Journalism)

Special diet requests: 40 vegetarians, 1 allergic to peanuts, 2 vegans

Other facts: Gym, Swimming pool

Task:

- There are 4 attendees allergic to peanuts, and 3 allergic to seafood. Change these special diet requests accordingly.
- Find out if there is a shuttle bus to the airport.
- Find out what kind of weather it will most likely be in London during your conference (the conference is in June)
- Find out if there will be a breakfast buffet.
- Find out if there is WiFi at the venue.
- Find out if the hotel rooms have Swedish television channels.

Type END to finish scenario.

“CAN’T YOU READ I’M FRUSTRATED?”

Bilaga D: Frågeformulär.

Questionnaire:

- E-mail:
- Age:
- Is english your first language? **Yes No**
- How easy was it to communicate in english during this task (before and during the interaction with the bot)?

Easy 1 2 3 4 5 Hard

- How familiar are you with using internet booking services for travels, accomodations, etc.? (less 1-5 more)

Less 1 2 3 4 5 More

- How familiar are you with interaction with this type of bot/interface?

Less 1 2 3 4 5 More

- Did you at any time experience any degree of the following emotions during your interaction with the chatbot: anger, anxiety, sadness, contempt, frustration? **Yes No**
- If yes, which one(s)?

“CAN’T YOU READ I’M FRUSTRATED?”

Bilaga E: Chatbotmanus.

SCENARIO A

Hello! My name is bookingassistant. I can do a number of things! Anything from guest contact to booking caterers - I’m your bot!

No scenario loaded. Type LOAD A, B or C to load a scenario.

Ok great, I am more than willing to help you with this.

Of course, give me a second and I will look for alternatives.

Of course, you’re welcome.

What can I help you with next?

Would you like to cancel speaker “M. Danielsson”?

Ok, speaker “M. Danielsson” is now cancelled.

What subject are you interested in?

That’s all the information I have. I’m sorry.

I was able to find 1 available speaker within the subject “Trends in AI”:

“E. Rusk”, would you like to book this speaker?

Ok, “E. Rusk” is now booked.

I was able to find 3 available speakers within the subject “Positive thinking/Mindfulness”

Speaker 1: Y. Nielsen

Speaker 2: M. Atwell

Speaker 3: Dr. P. Lindholm

Which one would you like to book?

Speaker 2: M. Atwell has the most bookings.

The speakers you have booked are “E. Rusk” and “M. Atwell”, would you like to change anything?

Ok, “E. Rusk” and “Dr. P. Lindholm” is now booked.

Ok, “Y. Nielsen” is now booked.

Ok, “M. Atwell” is now booked.

Ok, “Dr. P. Lindholm” is now booked.

At venue Clarion Post in location Gothenburg there is 1 Swimming pool and 0 Sauna.

How many guests will attend the event?

Ok, number of guests attending is now changed from 700 to 770.

The venue Clarion Post in location Gothenburg does not have a bicycle hire service. Would you like me to find out if the venue could arrange this for you when you arrive?

An e-mail has been sent to the venue with your inquiry.

SCENARIO B

Hello! My name is bookingassistant. I can do a number of things! Anything from guest contact to booking caterers - I’m your bot!

No scenario loaded. Type LOAD A, B or C to load a scenario.

Ok great, I am more than willing to help you with this.

Of course, give me a second and I will look for alternatives.

“CAN’T YOU READ I’M FRUSTRATED?”

Of course, you’re welcome.

What can I help you with next?

Sorry, I did not get that. Could you rephrase your inquiry?

Of course, I am more than willing to help you with this. Give me a second and I will look for alternatives.

Not that I can find. I’m sorry.

That’s all the information I have. I’m sorry.

The venue Best Western Karlaplan is not accessible by wheelchair.

Venue “Best Western Karlaplan” is now cancelled.

Ok, I found 3 available venues that are accessible by wheelchair that also match your other preferences:

Venue 1: Hotell Ibis

Venue 2: Radisson Blu Royal Viking

Venue 3: Hotel Kung Carl

There might be a difference in the price. Which one would you like to book?

Ok, “Hotell Ibis” is now booked.

Ok, “Radisson Blu Royal Viking” is now booked.

Ok, “Hotel Kung Carl” is now booked.

Your previous reservation has been cancelled.

Would you like to change venue to Best Western Karlaplan?

Would you like to increase the number of days by one?

The venue “Hotell Ibis” is booked for 1 day.

Ok, the booking is now increased by one day.

Best Western Karlaplan costs 1200 SEK per person per night.

Hotel Kung Carl costs 1500 SEK per person per night.

Hotell Ibis costs 1200 SEK per person per night.

Radisson Blu Royal Viking costs 1400 SEK per person per night.

Which one would you like to book?

The new price for your booking will be 200 000 SEK.

Your booked venue is Hotel Kung Carl.

Your booked venue is Radisson Blu Royal Viking.

Your booked venue is Hotell Ibis.

Would you like to change anything?

Would you like to change budget to 50 000?

The check-out time for Hotel Kung Carl is 12.

The check-out time for Radisson Blu Royal Viking is 12.

The check-out time for Hotell Ibis is 12.

Did you mean The Royal Castle bar and restaurant?

Did you mean The Royal Castle in Stockholm?

The Royal Castle in Stockholm is open:

October–April:

Tuesday–Sunday 10:00–16:00

“CAN’T YOU READ I’M FRUSTRATED?”

May–June:

Daily 10:00–17:00

July–August:

Daily 09:00–17:00

September:

10:00–17:00

SCENARIO C

Hello! My name is bookingassistant. I can do a number of things! Anything from guest contact to booking caterers - I’m your bot!

No scenario loaded. Type LOAD A, B or C to load a scenario.

Ok great, I am more than willing to help you with this.

Of course, I am more than willing to help you with this. Give me a second and I will look for alternatives.

Of course, give me a second and I will look for alternatives.

Of course, you’re welcome.

What can I help you with next?

Sorry, I did not get that. Could you rephrase your inquiry?

Sorry, the word “Swedish” is not in my vocabulary. Could you rephrase your inquiry?

Should I send a reminder to the guests 2 weeks before the conference?

Should I ask the guests about food preferences?

Still searching...

That’s all the information I have. I’m sorry.

Should I send this information to the catering service?

An e-mail has been sent to the venue with your inquiry.

An e-mail has been sent to all the guests asking for their food preferences. Awaiting confirmation.

An e-mail will be sent to the guests asking for their feedback.

Would you like to change the special food requests?

The Special food request is currently 40 Vegetarian, 4 Allergic to peanuts and 2 Vegan.

Would you like to change anything?

Ok, would you like to change Special food requests to 40 Vegetarian, 4 Allergic to peanuts, 2 Vegan and 3 allergic to seafood?

The Special food request is currently 40 Vegetarian, 5 Allergic to peanuts, 2 Vegan and 3 allergic to seafood. Would you like to change anything?

Ok, the Special food requests is now 40 Vegetarian, 4 Allergic to peanuts, 2 Vegan and 3 allergic to seafood.

Ok, the Special food requests is changed to 40 Vegetarian, 9 Allergic to peanuts and 2 Vegan.

Did you mean “aircraft”?

Sorry, the word “shuttle” is not in my vocabulary. Could you rephrase your inquiry?

The London Gatwick Airport is located approximately 27 miles from the city centre.

“CAN’T YOU READ I’M FRUSTRATED?”

The Hilton London Metropole is happy to provide all their guests with transportation to and from the airport.

Yes, would you like me to send a request for price information?

Ok, an e-mail has been sent asking for price information according to your request.

Type “YES” if you would like to receive hourly updates about the weather in London.

You will now receive an hourly update about the weather forecast in London.

The weather in London today is mainly dry during the evening with clear spells and light winds. Becoming breezier during the second half of the night with clouds thickening and some occasional rain likely.

The average temperature in June is around 18-20 degrees Celsius.

The Hilton London Metropole is happy to provide you with a dinner buffet from 5pm-10pm every night. In June, we have a special offer for you! Check in at the Hilton London Metropole using Facebook and you get a 10% discount when purchasing the dinner buffet!

Thank you!

At The Hilton London Metropole breakfast is included for all guests.

Did you mean “Wi-Fi”?

What about Wi-Fi?

Did you mean “breakdance”?

All of the rooms at The Hilton London Metropole have complimentary Wi-Fi for their guests.

Would you like me to send an e-mail with your inquiry to The Hilton London Metropole?

An e-mail with your inquiry has been sent to the venue.