



Universell utforming av chatbots



Innhold

| | |
|---|-----------|
| Fakta om rapporten | 3 |
| Sammendrag | 4 |
| Introduksjon | 6 |
| Metoder for testing..... | 6 |
| <i>Teknisk testing.....</i> | <i>6</i> |
| <i>Brukertesting.....</i> | <i>7</i> |
| Resultater..... | 8 |
| <i>Resultater fra teknisk testing.....</i> | <i>8</i> |
| <i>Resultater fra brukertesting.....</i> | <i>9</i> |
| Konklusjon | 12 |
| <i>UU-sjekkliste for chatbots.....</i> | <i>13</i> |
| Om Funka Nu AB..... | 14 |
| Vedlegg | 15 |
| <i>Scenarioer for brukertest.....</i> | <i>15</i> |
| <i>Intervjuguide.....</i> | <i>16</i> |
| <i>Liste med brister fra tekniske tester.....</i> | <i>17</i> |

Fakta om rapporten

Denne rapporten beskriver sluttresultatet av et prosjekt som ble gjennomført med finansiering av Barne-, ungdoms- og familiedirektoratet i løpet av 2019.

Medvirkende i prosjektet:

- Kristian Munter Simonsen (krav, test, analyse og rapportskrivning)
- Andreas Jacobsen (krav, test og analyse)
- Johan Kling (kvalitetssikring)
- Joachim Henstad (kontakt med brukerorganisasjoner og bedrifter)
- Susanna Laurin (prosjektledelse og administrasjon)

Sammendrag

Det blir stadig mer vanlig å bruke tekstbaserte chat-roboter, eller chatbots. Chatbots handler i hovedsak om å gi støtte til bedrifter og individer ved å simulere menneskelig oppførsel gjennom å automatisere dialog med hjelp av programvare eller kunstig intelligens (AI). Bedriftene kan tilby en mer effektiv kundeservice og spare penger, mens enkeltindividene kan for eksempel gjøre bordbestillinger på restaurant eller sørge for at taxien bestilles til rett tid på en enklere måte. På sikt er det mange som tror at chatbots skal kunne hjelpe til med omtrent alle tjenester og bli en naturlig del av andre samtaler i en chat.

Banker og større bedrifter har allerede begynt å ta i bruk chatbots. Og flere bedrifter, organisasjoner og myndigheter er nysgjerrige på hvilke muligheter denne nye teknologien fører med seg og vurderer å ta i bruk teknologien i for å gjøre det enklere for sine kunder, medlemmer eller brukere. Men som alltid, når teknologien gjør et større sprang, vil det være mange uvissheter, brister og potensielle problemer. Blant annet er det i dag ingen som vet om chatbots følger kravene til universell utforming. Målsettingen med prosjektet er å heve kunnskapen om hvordan chatbots fungerer for personer med funksjonsnedsettelse og dokumentere hvilke forbedringstiltak som må iverksettes knyttet til universell utforming.

I prosjektet har vi gransket hvordan noen utvalgte implementasjoner av chatbots fungerer for personer med funksjonsnedsettelse. Utvalget av implementasjoner ble gjort sammen med partnere i prosjektet, slik at det blir relevant for deres ønsker og behov. Mye av grunnlaget for hvilke løsninger som er evaluert baserer seg også på nedslagsfeltet, altså hvilke løsninger som blir brukt av flest personer.

En kombinasjon av brukertester og tekniske inspeksjoner er brukt for å kunne avdekke forbedringspotensialer. Funkas kravunderlag som bygger seg på WCAG 2.1 standarden er anvendt under de tekniske testene hvor totalt 37 suksesskriterier er blitt testet. Utvalget av tester baserer seg delt på det norske lovkravet samt andre aktuelle krav som går utover lovkravet, mens om er naturlig å teste opp mot basert på løsningens natur.

Gjennom rekrutteringsprosessen av testdeltakere var Funka i kontakt med flere brukerorganisasjoner. Interessen var stor og det endelige utvalget av testdeltakere inkluderte både personer med fysiske funksjonsnedsettelser samt personer med kognitive funksjonsnedsettelser. Det var tydelig at personer med ulike utgangspunkt for interaksjon med de aktuelle løsningene påpekte ulike utfordringer.

Alle de testede løsninger har brudd på retningslinjene. Forskjellene bygger seg hovedsakelig på selve implementeringen av ferdigløsninger hos de fleste av aktørene. Flere av de testede aktørene brukte boost.ai som grunnmotor for chatbotene. Selv om det ikke er noen kritiske feil som gjør anvendelsen umulig for enkelte brukere hos de aktørene som bruker boost.ai, er det ofte selve plasseringen av chatboten som kommer til å gjøre at mange brukere ikke vil benytte seg av løsningen. Et fint eksempel på dette er «kommune Kari» som er plassert lengst nede til høyre på skjermen og er tilnærmet ute av blikkfanget på

større oppløselige skjermer. Ikke minst må personer som bruker skjermlesere og tastaturnavigasjon gå igjennom hele innholdet på siden før de kommer til chatboten, da denne er plassert i footerne på siden. Andre løsninger som Tryg sin chatbot må aktivt oppsøkes for å finnes.

For å kvalitetssikre relevansen av resultatene, har vi i prosjektet samarbeidet med Skatteetaten og bransjeorganisasjonen IAAP, som bidrar med kunnskap om kjente brukerproblemer og dialog om eventuelle tenkelige forandringer/forbedringer i hvordan chatbots brukes og utformes. Skatteetaten har også bidratt med sin kunnskap og erfaring som potensielle bestillere og innkjøpere av chatbots, altså kundens perspektiv.

Resultatet av denne undersøkelsen bygger seg på kombinasjonen av resultater fra brukertestingene og granskningene. Resultatene fra disse to undersøkelsene er ikke entydige. På bakgrunn av dette er resultatene listet opp under for å enklere skille mellom de ulike resultatene fra studien.

- Enkelte løsninger som rangerer høyere en gjennomsnittet på brukertestene har resultater godt under gjennomsnittet når det kommer til de tekniske granskningene.
Ut ifra dette kan man se hvor sentralt og viktig det er å inkludere brukerne under utvikling og kvalitetssikringsprosesser. En løsning kan holde en høy teknisk samsvarsrate med forskrift om universell utforming av IKT og fremdeles inneholde hinder for enkelte brukere.
- Mye av ansvaret for at løsningene er universelt uformet tilfaller fremdeles den enkelte aktør, selv om løsningen er levert av en tredjepart. Flere av løsningene som er inkludert i denne undersøkelsen bruker den samme grunnmotoren. Dog skiller disse løsningene seg fra hverandre når man ser på samsvarsraten. Funkas anbefaling er at man påser at implementasjonsteamets kompetanse også inkluderer universell utforming.
- Spesielt personer med synsnedsetninger virker til å ha større utfordringer med chatbotløsningen som er inkludert i testene. Faktorer som formatet med flere hyppige responser fra chatbotten, skjermleserstøtte og zoom fremheves hyppig.
- Generelt mange utfordringer med utydelig identifikasjon av lenker og knapper samt dårlig kontrast. Alle disse utfordringene anslås å stamme fra implementasjonen av løsningen, ikke selve grunnmotoren.
- Språket chatbottene brukte var forståelig. Dog er det utfordringer i henhold til den aktuelle chatbottens tolkning av spørsmålene som blir stilt. Flere tilfeller hvor informasjonsbehovet ikke blir dekket på grunn av misforståelser og dårlig tolkning av forespørsmålene fra brukerne. (dette gitt at brukerne ikke skal klandres for hvordan forespørsmålene blir formulert).

Introduksjon

For at chatbots skal kunne gi støtte til alle, må de også være inkluderende. Digitale løsninger har potensiale til å gi personer med funksjonsnedsettelse økte muligheter til å ta mer makt over sitt eget liv, men det krever at den teknologien som benyttes er tilgjengelig for alle.

I dette prosjektet har vi undersøkt ulike chatbots for å øke kunnskapen om hvordan disse fungerer for personer med funksjonsnedsettelse og hvordan de kan gjøres tilgjengelige for alle.

Vi håper at resultatet fra dette prosjektet vil bidra til å kvalitetssikre chatbots, slik at alle mennesker får muligheten til å dra nytte av teknologien.

Metoder for testing

Funka har igjennom testperioden gjennomført to former for granskning i henhold til prosjektbeskrivelsen. Det ble først utført en ekspertgranskning mot de 31 identifiserte relevante WCAG kravene. Deretter ble det utført brukertester. Det kan leses mer om hvordan hver av metodene ble brukt under avsnittet «Metoder for testing»

Teknisk testing

Funka har lang erfaring med ekspertgranskning av digitale tjenester. Teknisk test ble gjort med erfaring fra evaluering av IT tjenester med WCAG i grunn. Under testingen ble 31 relevante WCAG krav valgt ut, disse WCAG kravene ble så hentet ned fra Funka sitt eget kravgrunnlag og arbeidet imot. Funka sitt kravgrunnlag inneholder mye som ikke står i WCAG, men for denne rapporten ble kun WCAG 2.0 og 2.1 krav tatt med. Totalt er det 37 «Funka krav» som er brukt under testingen. Ingen av Funka sine egenutviklede krav som ikke referer til WCAG krav var med, kun Funka sine oversettelser av WCAG 2.0 og 2.1 krav.

Den tekniske testingen ble utført av to Funka konsulenter, to konsulenter ble brukt fremfor en da WCAG krav kan bli tolket forskjellig av tilgjengelighetseksperter. Ved å ha to Funka konsulenter som tekniske granskere forsikret vi også at disse to var enige om alle resultatene som ble funnet. Se tabellen under for den komplette listen over WCAG krav som ble testet. Under teknisk testing ble WCAG brudd sjekket, kvaliteten på HTML koden og tastaturnavigasjon i chattevinduene både med og uten skjermleser. Ingen andre deler av nettsidene chatbotene befant seg på ble evaluert, kun chatbotene. Det ble også med bruk av manuell kodeinspeksjon evaluert om chatbotene brukte korrekte HTML elementer for å vise innhold. I den tekniske granskningen utgår vi fra de kriterier for universell utforming som ligger til grunn for Likestillings- og Diskrimineringsloven, den internasjonale standarden WCAG 2.0, WCAG 2.1 og EUs direktiv om tilgjengelighet av nettsider og mobilapplikasjoner som etter planen skal innlemmes i norsk lov 1. juli 2019 og som bygger på den europeisk

standarden EN301549 v 2.1. Nedenfor kan man se de tekniske kravene, relaterte kapitler og en kort forklaring som ble testet opp mot under de tekniske testene:

| Kategori | Krav | Navn på krav |
|------------|-------------|-------------------------------|
| Lesbart | 1.1.1 (A) | Ikke-tekstlig innhold |
| Lesbart | 1.3.1 (A) | Info og forhold |
| Lesbart | 1.3.2 (A) | Meningsfull rekkefølge |
| Lesbart | 1.3.3 (A) | Sanseegenskaper |
| Lesbart | 1.4.1 (AA) | Bruk av farger |
| Lesbart | 1.4.3 (AA) | Kontrast |
| Lesbart | 1.4.4 (AA) | Endre størrelse på tekst |
| Lesbart | 1.4.5 (AA) | Bilder av tekst |
| Lesbart | 1.4.6 (AAA) | Kontrast (utvidet) |
| Lesbart | 1.4.8 (AAA) | Visuell presentasjon |
| Lesbart | 1.4.9 (AAA) | Bilder av tekst (uten unntak) |
| Opererbar | 2.1.1 (A) | Tastatur |
| Opererbar | 2.1.2 (A) | Ingen tastaturfelle |
| Opererbar | 2.1.3 (AAA) | Tastatur (uten unntak) |
| Opererbar | 2.1.4 (A) | Snarveier for tegn |
| Opererbar | 2.2.2 (A) | Pause, stopp, skjul |
| Opererbar | 2.3.1 (A) | Tre blitz eller under terskel |
| Opererbar | 2.3.2 (AAA) | Tre blink |
| Opererbar | 2.4.3 (A) | Fokusordre |
| Opererbar | 2.4.6 (AA) | Overskrifter og etiketter |
| Opererbar | 2.4.7 (AA) | Fokus synlig |
| Forståelig | 3.1.1 (A) | Språk på siden |
| Forståelig | 3.2.1 (A) | Ved fokus |
| Forståelig | 3.2.2 (A) | Ved input |
| Forståelig | 3.2.5 (AAA) | Endre på forespørsel |
| Forståelig | 3.3.1 (A) | Feilidentifikasjon |
| Forståelig | 3.3.2 (A) | Etiketter eller instruksjoner |
| Forståelig | 3.3.3 (AA) | Feilforslag |
| Robust | 4.1.1 (A) | Parsing |
| Robust | 4.1.2 (A) | Navn, rolle, verdi |
| Robust | 4.1.3 (AA) | Statusmeldinger |

Funkas granskings- og testmetode av universell utforming er utviklet i nært samarbeid med brukerorganisasjonene og brukes for å kontrollere universell utforming på nett innen EU.

Granskingen gjennomføres av våre eksperter på universell utforming og ulike hjelpemidler. Vi bruker en kombinasjon av ulike tekstverktøy og manuelle tester. I tillegg til granskningene kommer vi til å gjennomføre tester der ulike brukere – med ulike hjelpemidler – tester hvordan chatbots fungerer i forskjellige grensesnitt, ulike typer av hjelpemidler som formidles, i alle relevante versjoner og i kombinasjon med alle typer av nettlere.

Brukertest

Brukertestere ble rekruttert gjennom de ulike brukerorganisasjonene i Norge. Totalt ble 8 brukertestere involvert i testingen, hvor alle fikk godtgjørelse i form

av gavekort. Personer med ulike typer funksjonsnedsettelse er inkludert, da hovedsakelig en person som er blind, en person som har synsnedsettelse, en person med dysleksi. Flere feil ble avdekket som følge av brukertesting som ikke ble avdekket under den tekniske testingen. Resultatene fra brukertesting kan lese om under neste kapittel. Brukertesting ble gjennomført på Funkas kontorer torsdag 10. oktober 2019 og baserte seg på ulike scenarier som simulerte situasjoner hvor en kunne tenke seg å bruke en chatbot for å finne løsninger. Et utvalg av de testede løsningene ble etablert hvor Norwegian, Tryg og DNB ble valgt ut til brukertestene. Scenariene ble naturligvis tilpasset hver enkelt aktør slik at chatbotene kunne komme med relevante svar.

Resultater

Resultater fra både de tekniske testene og brukertestene er gjengitt under. Merk at en bør se på resultatene kombinert, både tekniske tester og brukertester, for å få et representativt uttrykk av tilstanden rundt chatbotløsninger.

Resultater fra teknisk testing

Det var ingen av løsningene som passerte testene uten at feil ble funnet. Dog er det forskjell på mengden feil og alvorlighetsgraden av feil. Mange feil var også repeterende blant de ulike løsningene, noe som ikke kommer som en overraskelse da flere av løsningene bruker samme grunnmotor. En grunnleggende feil som flere løsninger feilet på var å markere tekst med <p> for paragraf. Det er her verdt å nevne at selv om mange aktører bruker samme grunnmotor/leverandør av chatbot systemer er det også en implementeringsfase som åpner for introduksjon av nye universell utformings brudd. Som eksempel kan kontrastfeil introduseres når bedriftens fargeprofil skal implementeres i chatbotløsningen. Enkelte av kravene som ble introdusert i testingen viste seg å ikke være aktuelle grunnet løsningens natur. De spesifikke kriteriene dette gjelder for er markert i oppsummeringen av resultatene fra de tekniske testene. Løsningene feilet gjennomsnittlig på om lag ¼ av de tekniske kriteriene. Ingen løsninger skiller seg drastisk fra de andre, bortsett fra DNB sin chatbot som scoret dårligere på de tekniske testene, men bedre på brukertestene. Formatet med et flertall meldinger som blir sendt etter hverandre kan virke til å være utfordrende for personer med skjermforstørring samt personer med lese-skrivevansker. Dette er dog noe av chatbotens natur og kan være utfordrende å få til å fungere for så mange som mulig.

Under presenteres en liste av de feilene som ble repetert hos tre eller flere aktører. Disse feilene er altså repeterende feil som en bør være ekstra nøye på om man selv skal implementere en lignende løsning.

- Semantisk korrekt HTML. HTML skal ikke inneholde alvorlige feil.
- Kontraster mellom tekst og bakgrunnsfarger
- Tekstlig beskrivelse av grafiske elementer
- Beskrivelse av skjema elementer
- Kommunikasjon om endringer på siden uten at fokus endres
- Forståelse av instruksjoner eller referanser avhengiges ikke av ens evne til å se eller hør

Under presenteres en tabell med oppsummering av resultatene for de ulike løsningene. Merk at det er kombinasjonen av de tekniske testene og brukertestene som burde legges til grunn for eventuelle prioriteringer av rettelser.

Teknisk testresultater:

| | Rikstv | Norwegian | Tryg | Bergen Kommune | Aftenposten | DNB |
|---------------|--------|-----------|------|----------------|-------------|-----|
| # Godkjent | 25 | 22 | 24 | 23 | 18 | 15 |
| # Underkjent | 8 | 8 | 6 | 7 | 9 | 19 |
| #Ikke aktuelt | 4 | 7 | 7 | 7 | 10 | 3 |
| Totalt | 37 | 37 | 37 | 37 | 37 | 37 |
| Samsvarsrate | 76% | 73% | 80% | 77% | 67% | 44% |

Gjennomsnittets teknisk samsvarsrate er på 70%.

Resultater fra brukertesting

Brukertester spiller en sentral rolle i denne testingen da det er brukerne til syvende og sist som skal interagere med løsningene. Enkelte feil som ikke ble avdekket i de tekniske testene blir fremhevet her. Under vil en kunne finne noen generelle observasjoner under brukertestene. Deretter er sitater fra brukertesterne inkludert hvor relevante utsagn er fremhevet.

- Det kan virke som personer som bruker skjermlesere og tastaturnavigasjon vil oppleve utfordringer ved bruk av disse løsningene.
- Personer med lese-skrivevansker vil kunne finne typografien og flere meldinger som ankommer samtidig utfordrende.
- Det er tydelig at brukernes opplevelser av løsningene er på et lavere nivå enn hva de tekniske testene fremhever.

Under er de oppsummerende tankene fra brukertesterne oppsummert. Merk at egne kommentarer og klargjøringer er fremhevet i parenteser, krøllparenteser og klammeparrantereser:

Bruker nr.1:

Den sliter litt med å vite hva den skal gjøre men jeg føler at den begynner å bli bedre. de som ikke er så god på data vil nok ha problemer med å finne frem. Forståelig språk. Delvis dekt informasjonsbehov. De har lett for å misforstå litt. En fordel hadde vært å få det i et eget vindu slik at du kan navigere til der du var. Det å finne tilbake etter å ha fulgt en lenke var litt vanskelig, for da starter du er en ny chat. DNB har for så vidt et eget vindu.

Bruker nr.2:

Lite varierende, DNB var den som best, trygg helt krise. Norwegian mellomting. Ved to av tre anledninger [fikk jeg informasjonsbehovet dekket]. Til slutt skrev jeg bare ønsker om generell informasjon, men det klarte den [Tryg] heller ikke. Hadde vært mye lettere å finne informasjon gjennom nettsiden. Generelt ikke lette å bruke. Flere forbedringspotensialer som hadde gjort det lettere for meg som blind bruker å bruke. F.eks. når jeg har sendt en melding vet jeg ikke om den har blitt sendt. Best på DNB men det samme problemet med at jeg ikke vet om jeg sendte. Hadde valgt mennesker fremfor chatbot [chatbots bør ikke bytte ut kundebehandlere]. [lenker]: Tryg med popup helt krise, først ingenting, trengte hjelp. Noen ganger var det vanskelig å prøve å finne knappene. Hos Norwegian er ikke send-knappen i tabulatorrekkefølge og må derfor navigere med piler (for skjermbrukere). Språket chatboten brukte var forståelig, problemet var mer at den ikke skjønnte hva jeg sa. Popup på Tryg var det verste jeg har sett, hadde vært bedre med et ekte vindu. Dere som har sittet her har begge sett hvor irritert jeg ble da det ikke fungerte for meg. DNB med eget vindu var bra med forskjellige vinduer. Vanskeligere med flere grensesnitt i samme fane.

Bruker nr.3:

Brukervennlighet MEN bruk av farger er dårlig, dårlige ikoner, snarere en tekst med send. Jeg fikk den informasjonen jeg var på utkikk etter, men jeg måtte stille fire spørsmål i fire retninger før jeg fikk svar. DNB best og informer best. Jeg har brukt det før, så jeg vet at du må stille spørsmål på en spesiell måte. Hadde opplevd vanskeligheter som første gangs bruker. Mer info om hvordan du skriver spørsmålene dine. Jeg synes de likner hverandre for mye, alle tre var like dumme og at de minner hverandre om layout på en lignende måte og like vanskelig å lese alle tre, jeg skulle rangere dem 3. Tryg, 2. Norwegian, heller ikke bra men 1.DNB. Tryg var vanskelig å navigere til. Forsøkte å forstørre, men da ble området mindre, og svarene forsvant raskere

Bruker nr.4:

Forklar på samme sted at det er en bot jeg snakker med, ikke i et eget vindu før chaten starter. Det er toleranse for feilstaving som er viktig tenker jeg. Den som jeg stoppa opp på var betingelser for turen, den bomma de litt på. Jeg tenker at det er jo en «late som dialog» da. spørsmålet er om jeg hadde brukt verktøyet. Det er jeg ikke like sikker på. chatbotene vil jeg vanligvis brukt hvis jeg ikke finner det jeg leter etter å prøvd to ulike steder. Veldig fint med chatbot for

sensitive spørsmål. Løsningen i seg selv kan være veldig positiv. Ingen av dem fortalte meg noe om mitt personvern i bruk av chatboten.

Bruker nr.5:

Dekt behov fra Norwegian og god info. Fikk dekt informasjonsbehov om reisebyrå. Ikke dekt behov fra trygg. chatbot fant ikke skadetype selv når den fikk beskjed om vannskade. jeg syntes det kom mye informasjon på en gang som gjør at jeg må scrolle og lese meg opp på lite relevante meldinger ofte. DNB dekt informasjonsbehov. til noen av oppgavene jeg fikk som f.eks. fond og billån var det veldig straight forward.

Bruker nr.6:

Jeg skriver avbestille reise. Da får jeg opp språkvalg og velger norsk. Jeg fikk ikke med meg at det kom to bokser med informasjon. Var ikke klart at det kom flere chatboks dialoger. Gruppering av knapper til samtalen og lenker burde skilles. Fikk det ikke til. Ved zoom forsvinner/viser ikke alle meldingene. Ved zoom forsvinner noe av teksten i knappene. Ikke tydelig hva som er knapper som leder til en ny side og hva som fortsetter i samtalen. Chatboten ligger ikke på forsiden, dette var vanskelig å forstå. Samtaleboksene hopper opp og ved zoom er det ikke tydelig hvilken melding som er den første. Svarene var noe kronglete formulert. Språket bør jobbes med. Ikonbruken var ikke helt synlig. Ingen forskjell på lenker som leder til en ny side kontra de som bare setter inn tekst i chaten. {informasjonsbehov:}Nei. DNB fikk jeg det, men hos Norwegian fikk jeg ikke det. Jeg slet litt når det var flere avsnitt over hverandre. Ting forsvant. Nei det er ikke tydelig at det var mer innhold tilstede som jeg ikke så.

Konklusjon

Konklusjonen i denne studien bygger seg på de tekniske og brukertestene som er gjennomført i prosjektperioden. Selv om de tekniske testene har påvist en høyere enn gjennomsnittet samsvar med WCAG 2.0 og 2.1 har brukertestene fremhevet flere mangler som setter store begrensninger i anvendbarheten og brukervennligheten til chatbotløsninger. Det kan virke som at "ut av boksen"-løsningene mangler sentrale tagger og markeringer i koden som en burde kunne forvente som standard fra den aktuelle leverandøren. Det blir også introdusert flere feil under implementasjon av bedriftens visuelle profil (CSS).

Chatbotløsningene testet er ikke i samsvar med gjellende standarder samt at implementasjonen av disse løsningene hos den enkelte bedrift også skaper WCAG samsvarsbrudd.

Det vil være hensiktsmessig å kvalitetssikre at minst en person i implementasjonsteamet har kunnskap og erfaring fra WCAG standarden. På denne måten kan man i det minste påse at ingen nye feil blir introdusert når løsningen adopteres. Det vil også være mulig å fikse potensielle mangler i "ut av boksen"-løsningen som tilbydes om man påser riktig kompetanse i implementasjonsteamet. Et repeterende problem i markedet generelt er at flere aktører deklarerer løsningene sine som å være i samsvar med WCAG, selv om løsningene egentlig ikke er det. Dette er noe kundene av slike løsninger må bli flinkere til å kvalitetssikre og gjerne stille krav til en uavhengig analyse av løsningen.

Chatboter er av en kompleks natur som ikke støttes fullt ut av HTML elementer. Det er med andre ord spesielt viktig i slike løsninger å påse at man bruker korrekt WAI-ARIA tagger i løsningen. En løsning som chatboter vil etter alt kunne klassifiseres som er rik internettapplikasjon.

Typiske feil som gjentok seg i de tekniske testene er:

- Semantisk korrekt HTML. HTML skal ikke inneholde alvorlige feil.
- Kontraster mellom tekst og bakgrunnsfarger
- Tekstlig beskrivelse av grafiske elementer
- Beskrivelse av skjema elementer
- Kommunikasjon om endringer på siden uten at fokus endres
- Forståelse av instruksjoner eller referanser avhengiges ikke av ens evne til å se eller hør

Det fremgår også av rapporten at det er et avvik mellom de tekniske resultatene og brukertestene. Enkelte av løsningene som får lavere teknisk samsvarsrate var løsningene flest brukere foretrakk. Studien er relativt liten, noe som må tas i betraktning. Denne tendensen er dog verdt å nevne, da spesielt for de av oss som jobber med kvalitetssikrings.

UU-sjekkliste for chatbots

For enhver aktør som vurderer å tilby en chatbotløsningen kommer det her en sjekkliste for at kvaliteten med hensyn på universell utforming skal bli høy:

- Etterspør dokumentasjon fra leverandøren på hvorvidt en opprettholder norske krav til universell utforming av IKT.
- Påse at teamet som får ansvar for selve implementeringen av løsningen har den kompetansen som trengs for å ikke introdusere nye UU feil. Det finnes sertifiseringer for universell utforming som kan være et godt utgangspunkt for å forsikre seg denne kompetansen. WAI-ARIA er spesielt aktuelt for løsninger som chatboter.
- Påse at løsningen er lett tilgjengelig og synlig på nettsiden.
- Gjennomfør brukertester med personer med ulike funksjonsnedsettelse, gjerne tidlig og kontinuerlig gjennom utviklingsprosessen. I det minste burde disse type brukerne bli inkludert i akseptansetesting og/eller brukerakseptansetesting.

For enhver aktør som vurderer en løsning som chatboter er det spesielt tre aspekter utenom universell utforming som vil være ekstra viktig å forholde seg til:

- Chatboter må ikke bytte ut tilbudet av kundebehandlere. Det er ikke bare hvordan løsningene presenteres for brukerne som skaper problemer, men også tolkningen og tilbakemeldingene man får fra slike løsninger. Flere av brukertesterne hadde større utfordringer med å finne frem til informasjonsbehovet som ble etterspurt chatbotene. Hvis det da ikke er mulig å komme i kontakt med en kundebehandler blir brukeren både frustrert og irritert.
- Det må bli kommunisert tydelig at man faktisk snakker med en datamaskin og ikke med en kundebehandler. Man vil kunne føle seg lurert og bedratt om man ut i en samtale skulle oppdage at man ikke snakker til et annet menneske. Dette er et punkt som er spesielt viktig for bedrifter som ønsker å være åpne, transparente og som holder etikk høyt.
- GDPR aspekter relatert til lagring av data og hvilken type data som blir lagret i samtalelogger. Man burde bli informert hvordan og hvor lenge data fra en samtale med en chatbot blir lagret.

Om Funka Nu AB

Funka startet som et ideelt prosjekt innen den svenske handikapbevegelsen, og er i dag markedsledende innen området universell utforming. Siden år 2000 har vi vært et privateid foretak, og vår nære relasjon til brukerorganisasjonene innebærer en unik kvalitetskontroll og forankring.

Vi åpnet vårt Oslokontor 2010, og har blant annet utviklet indikatorene som Difi benytter for å måle etterlevelsen av Diskriminerings- og Tilgjengelighetsloven, DTL. Vi har utarbeidet tekster, illustrasjoner og kodeeksempel til Difis Løsningsforslag for web. Vi har stilt som eksperter når Likestillings- og Diskrimineringsombudet, LDO, utredet klagesak opp mot DTL. Videre har vi hjulpet et stort antall kunder i offentlig, privat og ideell sektor å utforme grensesnitt som fungerer for alle.

Utover den norske virksomheten har vi kontor i Stockholm og Madrid, med kunder over hele verden. Funka er en av grunnleggerne av den internasjonale bransjeorganisasjonen for profesjonelle innen universell utforming, IAAP, og vi arbeider tett med EU-kommisjonen og flere europeiske regjeringer angående implementering av regelverk og politisk metodikk tilknyttet universell utforming.

Funkas eksperter er verdensledende innen universell utforming av innhold, design og teknikk i digitale grensesnitt; web, IT, dokument i alle tenkbare format, apper og automater, men også skilting og annen informasjon. Vi jobber også med universell utforming i fysisk miljø, orientering, akustikk og muntlig kommunikasjon.

Våre anbefalinger bygger på internasjonale retningslinjer, praktisk erfaring og vitenskapelig grunnlag. Omtrent 20 % av Funkas virksomhet finansieres av forskningsmidler, og vi samarbeider med flere ledende universitet.

Gjennom tillitsoppdrag og standardisering utformer Funka krav til universell utforming, og vi setter normer for utvikling og analyse. For EU-kommisjonen har vi blant annet utarbeidet kontrollmetodikk og målt status for universell utforming på nett i samtlige EU-land, Norge, USA, Canada og Australia. På oppdrag fra W3C har vi gjennomført den autoriserte oversettingen av WCAG 2.0 til svensk.

Vi arbeider aktivt med standardisering på nasjonalt nivå i Norge, Sverige og Spania, internasjonalt gjennom W3C og ISO, og på EU-nivå, der vi blant annet har bidratt til felles europeiske krav til universell utforming ved offentlige anbud i EN 301 549.

Funka er EPiServer Solution Partner, Microsoft Partner og Adobe Certified Training Provider. Vi står bak Funkaportalen, med informasjon og nyheter for personer med nedsatt funksjonsevne.

Vedlegg

Scenarier for brukertest

SCENARIO 1 - REISEN

- Du skal nå dra på reise. Du ønsker å vite mer om betingelser og rettigheter du har for turen. Finn ut hvordan man avbestiller reisen.
- Du ønsker deretter å få informasjon om reiseforsikringer fra Tryg.
- Få informasjon om hvordan betale MasterCard regningen fra DNB.

SCENARIO 2 – VANNSKADEN

- Du bor i Fredrikstad og har fått vannskader i leiligheten din, meld fra om skade til Tryg forsikring, be om nødvendig hjelp under melding (ikke lever skademelding).

SCENARIO 3 – OMBESTILLINGSSITUASJONEN

- Du vil ombestille billetter på Norwegian, finn ut hva vilkårene for din Lowfare+ billett er.
- Du kom på at du har bestilt reisen din igjennom ett reisebyrå, hva er vilkårene da?

SCENARIO 4 – LUKSUSFELLEN

- Du vil ha den nye Teslaen, få DNB sin chatbot til å ta deg til siden for å søke nytt billån.
- Rett før du kjøper den nye Teslaen kommer du på at du allerede har en Tesla, du ønsker heller å legge pengene i fond. Få hjelp av DNB sin chatbot til å komme til informasjonssider om fond hos DNB, du vil se på forskjellige typer fond.

Intervjuguide

Intervjuguiden ble brukt etter scenarioene ble gjennomført. Hensikten med spørsmålene var å få et helhetlig bilde av opplevelsen av chatbotene. Ytterligere oppfølgingsspørsmål ble stilt om svarene var uklare.

1. Hva syntes du om brukervennligheten til chatbotene?
2. Var det klart for deg at du snakket med en chatbot?
3. Fikk du dekt ditt informasjonsbehov?
4. Var chatbotene lette å bruke?
5. Foretrekker du chatbots fremfor mennesker?
6. Syntes du chatbotene likner på hverandre?
7. Brukte du kun mus eller også tastatur for navigasjon?
8. Var lenke forståelige?
9. Var det greit å finne frem til hvor chatbotene var?
10. Følte du at språket chatboten brukte var forståelig?
11. Er det enkelt å forstå hva du har skrevet og hva chatboten har svart?
12. Foretrakk du chatbot rett i vinduet eller i ett nytt vindu?
13. Har du noe mer du vil si eller noe du har notert?

Liste med brister fra tekniske tester

Under presenteres resultatene fra de tekniske testene. Her er både krav fra Ligestillings- og Diskrimineringsloven og EUs direktiv om tilgjengelighet av nettsider og mobilapplikasjoner. Merk at et WCAG suksesskriterier kan brytes på flere forskjellige måter. Dette er årsaken til at flere kriterier er inkludert flere ganger.

| Funkas krav | Tema | Relaterte WCAG suksesskriterier | Rikstv | Norwegian | Tryg | Bergen Kommune | Aftenposten | DNB |
|-------------|--------------------------|---|---------------|---------------|---------------|----------------|---------------|---------------|
| HC30 | Html & css | 4.1.1 (A) | Underkjent | Underkjent | Underkjent | Underkjent | Underkjent | Underkjent |
| RD40 | Flexibility of interface | 1.4.4 (AA), 1.4.8 (AAA) | Godkjent | Godkjent | Godkjent | Godkjent | Godkjent | Godkjent |
| NL20 | Navigation & links | 2.1.1 (A), 2.1.2 (A), 2.1.3 (AAA), 3.2.1 (A), 3.2.2 (A) | Godkjent | Godkjent | Godkjent | Godkjent | Godkjent | Underkjent |
| NL30 | Navigation & links | 2.4.3 (A) | Godkjent | Godkjent | Godkjent | Godkjent | Underkjent | Underkjent |
| NL40 | Navigation & links | 2.4.7 (AA) | Underkjent | Godkjent | Godkjent | Godkjent | Godkjent | Underkjent |
| CP10 | Color & presentation | 1.4.3 (AA), 1.4.6 (AAA) | Underkjent | Godkjent | Godkjent | Underkjent | Underkjent | Underkjent |
| CP60 | Color & presentation | 2.2.2 (A), 2.3.1 (A), 2.3.2 (AAA) | Godkjent | Godkjent | Godkjent | Godkjent | Godkjent | Godkjent |
| CP110 | Color & presentation | 3.1.1 (A) | Godkjent | Godkjent | Godkjent | Godkjent | Underkjent | Underkjent |
| ST60 | Structure & formatting | 1.3.1 (A) | Ikke relevant | Ikke relevant | Ikke relevant | Ikke relevant | Ikke relevant | Ikke relevant |
| ST120 | Structure & formatting | 1.3.1 (A) | Ikke relevant | Ikke relevant | Ikke relevant | Ikke relevant | Ikke relevant | Ikke relevant |
| ST180 | Structure & formatting | 1.3.1 (A) | Godkjent | Underkjent | Godkjent | Godkjent | Godkjent | Underkjent |
| ST200 | Structure & formatting | 1.3.1 (A) | Ikke relevant | Ikke relevant | Ikke relevant | Ikke relevant | Ikke relevant | Ikke relevant |
| IM10 | Images | 1.4.5 (AA), 1.4.9 (AAA) | Godkjent | Godkjent | Godkjent | Godkjent | Godkjent | Godkjent |
| IM20 | Images | | Godkjent | Godkjent | Underkjent | Underkjent | Godkjent | Godkjent |
| IM30 | Images | 1.1.1 (A) | Underkjent | Godkjent | Underkjent | Underkjent | Godkjent | Underkjent |
| FS10 | Forms | 1.3.1 (A), 4.1.2 (A) | Underkjent | Godkjent | Godkjent | Godkjent | Godkjent | Underkjent |
| FS50 | Forms | 3.3.2 (A) | Underkjent | Godkjent | Underkjent | Underkjent | Godkjent | Underkjent |

| | | | | | | | | |
|-------|-------------------|--|---------------|---------------|---------------|---------------|---------------|------------|
| FS60 | Forms | 1.3.1 (A), 2.4.6 (AA), 3.3.2 (A), 4.1.2 (A) | Godkjent | Underkjent | Godkjent | Godkjent | Godkjent | Underkjent |
| FS120 | Forms | 4.1.2 (A), 1.1.1 (A), 4.1.2 (A) | Godkjent | Underkjent | Godkjent | Godkjent | Godkjent | Godkjent |
| FS130 | Forms | 1.3.1 (A) | Underkjent | Ikke relevant | Ikke relevant | Ikke relevant | Godkjent | Underkjent |
| FS140 | Forms | 3.3.1 (A) | Godkjent | Godkjent | Godkjent | Godkjent | Godkjent | Godkjent |
| FS150 | Forms | 1.3.1 (A) | Godkjent | Ikke relevant | Ikke relevant | Ikke relevant | Ikke relevant | Godkjent |
| FS170 | Forms | 3.3.1 (A) | Godkjent | Godkjent | Godkjent | Godkjent | Ikke relevant | Godkjent |
| FS180 | Forms | 3.3.3 (AA) | Godkjent | Godkjent | Godkjent | Godkjent | Ikke relevant | Godkjent |
| SW20 | Script & wai-aria | 4.1.2 (A) | Godkjent | Underkjent | Godkjent | Godkjent | Underkjent | Underkjent |
| SW50 | Script & wai-aria | 1.3.2 (A), 2.4.3 (A) | Godkjent | Godkjent | Godkjent | Godkjent | Godkjent | Godkjent |
| SW60 | Script & wai-aria | 2.4.3 (A) | Godkjent | Godkjent | Godkjent | Godkjent | Underkjent | Underkjent |
| SW70 | Script & wai-aria | 4.1.3 (AA) | Godkjent | Underkjent | Godkjent | Godkjent | Underkjent | Underkjent |
| SW80 | Script & wai-aria | 1.4.13 (AA) | Godkjent | Godkjent | Godkjent | Godkjent | Godkjent | Godkjent |
| SW90 | Script & wai-aria | 2.4.3 (A) | Godkjent | Godkjent | Godkjent | Godkjent | Ikke relevant | Godkjent |
| SW100 | Script & wai-aria | 4.1.2 (A) | Godkjent | Ikke relevant | Underkjent | Underkjent | Ikke relevant | Godkjent |
| AE20 | Automatic events | 3.2.1 (A), 3.2.2 (A), 3.2.5 (AAA) | Godkjent | Underkjent | Godkjent | Godkjent | Godkjent | Underkjent |
| AE60 | Automatic events | 2.1.1 (A), 2.1.2 (A), 2.1.4 (A), 3.2.5 (AAA) | Ikke relevant | Ikke relevant | Ikke relevant | Ikke relevant | Ikke relevant | Underkjent |
| AE80 | Automatic events | 4.1.2 (A) | Godkjent | Godkjent | Godkjent | Godkjent | Underkjent | Underkjent |
| MS40 | Help & transition | 1.3.1 (A) | Godkjent | Godkjent | Ikke relevant | Ikke relevant | Ikke relevant | Underkjent |
| US10 | Understandability | 1.4.1 (A) | Godkjent | Godkjent | Godkjent | Godkjent | Godkjent | Godkjent |
| US20 | Understandability | 1.3.3 (A) | Underkjent | Underkjent | Underkjent | Underkjent | Underkjent | Godkjent |