

Inledning

Denna rapport sammanfattar viktiga slutsatser från Steg 1 av projektet *Talresursinsamlaren – För ett tillgängligare Wikipedia genom Wikispeech*. Rapporten kompletterar den ansökan som förberetts för Steg 2 av projektet med ytterligare bakgrund och detaljrikedom om hur *Talresursinsamlaren* knyter an till Wikimediaplattformarna och vilka komponenter som vi kommer att utveckla inom ramen för projektet.

Utvecklingen av Wikispeech är en prioritet för projektdeltagarna som går in med omfattande egenfinansiering. Om projektet blir finansierat finns goda möjligheter för långsiktig drift och arbete med att säkerställa att talresurser på svenska inte längre kommer att vara en bristvara och att Wikipedia, världens femte mest använda webbplats, blir tillgänglig för de som har svårt att läsa.

Slutsatser rörande inspelning och lagring av svenska talresurser som en del av Wikispeech

Efter diskussioner med ett stort antal aktörer under Steg 1 är vår bedömning att behovet av stora mängder taldata som är fritt licensierade är mycket stort och att orsakerna är flera och mångfacetterade. Även om mindre mängder taldata kan ge stort mervärde för specifika, avgränsade användningsområden, finns idag ett mycket stort behov av större mängder taldata med ett bredare perspektiv. En allmänt hållen insamling med utförlig metadata, sammanlänkning och annotering, skulle lägga den grund som behövs för senare specialisering. Att nyttja crowdsourcing bedöms vara en utmärkt möjlighet för detta. Värdet med projektet är större än vad vi initialt trodde.

Att den mjukvara och de processer runt insamling som vi tar fram under projektet kommer att få en stor påverkan för ett litet språk som svenska är ställt utom tvivel. Det kommer att:

- Skapa en fritt licensierad rik svenskspråkig resurs av inspelningar och annoteringar som kan användas i mängder av olika sammanhang, och som jämnar ut spelplanen för mindre aktörer med begränsade resurser (exempelvis ideell sektor eller SME:s).
- Utveckla specialverktyg för annotering av språkdata som kommer att fylla ett gap i verktygslådan. Då fokuset är på svenska kommer dessa verktyg att ha extra stor betydelse för de som vill arbeta vidare med inspelningar av svenskt tal och annoteringar.

- Göra svenskspråkiga Wikipedia mer tillgängligt då de som inte kan läsa kan lyssna på texten istället med talsyntes av hög (och kontinuerligt förbättrad) kvalitet. Wikipedia är den femte mest använda webbplatsen i Sverige.
- Öka Wikipedias värde som lärresurs då många lär sig genom att lyssna.
- Möjliggöra att ta reda på korrekt uttal på svåra ord som exempelvis namn, etc.
- Möjliggöra ökad forskning och utveckling inom området, eftersom det finns ett mycket stort behov av tillgängliga, välstrukturerade och annoterade taldata, samt verktyg och metoder för att samla och tillgängliggöra data. Detta ökar kvaliteten på talsyntes och möjliggör förbättrad taligenkänning.

Våra efterforskningar visar att de finansiärer som kan vara aktuella utöver PTS, främst har sitt fokus på andra språk än svenska. Med det stora intresset vi sett från olika aktörer, samt det starka projektteam vi har satt samman, ser vi en unik chans att vara med och lägga grunden för ett innovativt sätt att samla in inspelningar av tal och annoteringar för svenska.

Verktygen kommer även att kunna användas för andra språk, och ett möjligt användningsområde är exempelvis bevarandet av hotade och snabbt krympande språk. Även med ett uttalat fokus på svenska talresurser kommer projektet att få en global påverkan.

Det faktum att vi kommer att använda oss av länkade öppna data (LOD) för delar av inspelningarna möjliggör också sökningar över olika fält och ämnesområden, vilket kan öppna för idag okända möjligheter med en dylik samling taldata.

Koppling av *talresursinsamlaren* till talsyntes

Syftet med Wikispeech, där *talresursinsamlaren* är en integrerad men separat del, är att göra Wikipedia tillgängligt för de som av olika orsaker inte kan ta del av skriven text. För att verktyget inte skall introducera missförstånd och vara uppskattat att använda är kvaliteten avgörande. Då Wikipedia inte är ett kommersiellt bolag utan lever på donationer är dock möjligheterna att anställa personal för att utveckla de olika språken begränsade. Vad som identifierats som en möjlig väg framåt är att nyttja crowdsourcing för att få in de förbättringar som behövs.

Den data som vi samlar in kommer inledningsvis att fokusera på den specialterminologi som finns i de många Wikipediaartiklarna så att förståelsen för dem kan maximeras genom att uttal förbättras. Insamlingen av inspelat tal och annoteringar av dessa kommer därmed att vara till stort värde för att öka kvaliteten på Wikispeech som talsyntes (text-till-tal).

Förutom denna direkta effekt, kommer all data att vara fritt tillgängliga för externa aktörer, som kan använda dem till att förbättra sina egna produkter och tjänster.

Koppling av *talresursinsamlaren* till taligenkänning

Genom den stora mängd taldata som kommer att samlas in kommer även möjligheten att utveckla en kvalificerad tjänst för taligenkänning att kunna utvecklas med tiden. Det krävs dock mellan 5 000-10 000 timmar med inspelat tal för en fungerande taligenkänning. Inspelningarna behöver därtill vara av en god blandning av olika talare, olika typer av mikrofoner m.m. Ju bättre blandning och ju mer data, desto bättre taligenkänning.

10 000 timmar är dock väldigt mycket och vi ser en del lågt hängande frukter på vägen dit. Exempelvis skulle ett stort antal inspelningar av Wikipedias artikeltitlar innebära att sökfunktionaliteten på Wikipedia men även på andra sökmotorer skulle förbättras genom automatisk taligenkänning av söktermer. Effekten blir desto större tack vare att licensen Creative Commons Zero (CC0), där resurserna är helt fria och kan användas utan begränsningar, har valts.

Integrering med MediaWiki

Mjukvaran MediaWiki är vad som finns under huven på Wikipedia och de andra Wikimediaplattformarna. Genom att utveckla MediaWiki-mjukvaran med ytterligare funktionalitet får arbetet potentiellt ett oerhört stort genomslag då verktygen kan inkluderas på välanvända webbplatser såsom Wikipedia.

Det är även viktigt då crowdsourcing bygger på medvetenhet om och intresse av att delta. Att knyta an till ett varumärke (brand) som är så pass positivt laddat som Wikipedia bedömer vi är en väg in för att engagera ett stort antal personer.

Dessa positiva aspekter har fått oss att välja att knyta *talresursinsamlaren* nära Wikipedia och även vilka val vi gjort gällande vilket material vi ber folk bidra med talinspelningar av. I linje med detta tänker vi fortsätta arbetet med att knyta talsyntesen Wikispeech närmare Wikipedia och aktivera den i encyklopedin. Detta val skapar självklart vissa begränsningar i genomförandet. Tillägg till MediaWiki-mjukvaran kräver att materialet går igenom en rigorös kod- och säkerhetsgranskning vilket tar tid och resurser i anspråk. Vi har lärt oss mycket runt detta i vårt PTS-finansierade projekt Wikispeech – en användargenererad talsyntes på Wikipedia (Dnr: 15-8729) och har satt av mycket mer resurser för detta arbete samt engagerat Wikimedia Deutschland i arbetet (de deltar generöst nog på egenfinansierad tid).

Under de senaste halvåret har vi haft regelbunden kontakt med Wikimedia Foundation för att ta arbetet vidare vilket har varit väldigt värdefullt. Vi är överens med dem om att fortsätta samarbetet även under 2019.

För att de verktyg vi utvecklar ska kunna implementeras behövs därtill en *buy-in* av den aktiva gemenskapen på Wikipedia. Detta innebär att vi måste fokusera på att engagera gemenskapen och aktivt kommunicera med dem på olika sätt.

Införlivning i Wikimediaplattformarna (av en modul för insamling av talresurser)

Talresursinsamlaren kommer att kopplas samman med spelaren i Wikispeech så att de som nyttjar Wikispeech enkelt kan bidra till att förbättra den. Den kommer även att kopplas samman med den underliggande databasen för MediaWiki-tillägget Wikispeechs talsyntes så att uttalslexikonet, som är en central resurs, där kan utvecklas.

Talresursinsamlaren kommer även att kopplas samman med de olika plattformarna som drivs av Wikimedia för att det crowdsourcade materialet automatisk skall tillgängliggöras till en stor publik. Inspelningarna kommer att tillgängliggöras på Wikimedia Commons, den mediadatabas som Wikipedia använder. Delar av ljudinspelningarna kommer att läggas där så att de kan användas i Wikipedia för att exemplifiera uttal av exempelvis artikeltitlar eller av inspelningar av hela artiklar som skapats av volontärer. Vi kommer de närmaste månaderna att arbeta nära med ett team på Wikimedia Foundation för att beskriva hur ljud bäst kan lagras där.

Ljudinspelningarna kommer även att länkas till relevanta artiklar på Wikipedia, den fria encyklopedin, samt ord på Wiktionary, den fria ordboken, för att illustrera hur ordet uttalas. Detta innebär en enorm spridning då de båda plattformarna är mycket välanvända med över en halv miljard besökare per månad.

Därtill kommer inspelningarna att kopplas till Wikidata, en strukturerad databas som är allmänt tillgänglig att söka i och bidra till. Sökbarheten blir därmed stor och det går att göra avancerade s.k. queries för att hitta information över flera databaser och ämnesområden. Vi kommer att jobba nära med Wikimedia Deutschland (Tyskland) för att genom olika specialbyggda lösningar kunna inkludera inspelningar av svenskt tal (länkar till mediafilerna på Wikimedia Commons, se nedan) samt vissa annoteringar i Wikidata i form av länkade öppna data. Det är Wikimedia Deutschland som är huvudansvariga för vidareutvecklingen av Wikidata.

All dessa data på Wikidata och Wikimedia Commons kommer att lagras på och tillgängliggöras via Wikimedia Foundations servrar utan kostnad för projektet.

Integration av existerande mjukvarulösningar

Vi kommer inom ramen för Steg 2 att arbeta med Wikimedia Deutschland för att effektivt kunna inkludera data i Wikidata. Wikimedia Deutschland utvecklar för närvarande mjukvara runt de lexikala resurser som skall inkluderas på Wikidata och vi kommer att koppla våra insamlingsverktyg till Wikidata.

Vi kommer även att arbeta med Wikimedia Foundation med inkluderingen av inspelat tal på mediadatabasen Wikimedia Commons. De arbetar aktivt med att aktivera strukturerad data i mediadatabasen (från att tidigare ha varit baserat på strängar, d.v.s. dum text). Vi kommer att jobba med gemenskapen på Wikimedia Commons och Wikimedia Foundations personal för att inkludera det inspelade talet i mediadatabasen på ett bra sätt.

Språkbanken Tal håller på att bygga upp ett system för lagring av svenska talresurser och teamet på KTH som ansvarar för detta är med i det här projektet. Vi kommer att arbeta för en integrering av materialet som skapas genom *Talresursinsamlaren* och dra nytta av lösningar som utvecklas i respektive projekt.

Stiftelsen Mozillas Foundations projekt *Common Voice* har utvecklat en hel del lösningar som ligger under fria licenser som därmed kan implementeras i vår utveckling av *talresursinsamlaren*. Vi har påbörjat ett samtal och kunskapsutbyte med deras team för att inte återuppfinna hjulet och istället säkerställa att vi arbetar inom kompletterande områden för att gagna varandra.

Integritetsaspekter för insamling och publicering av taldata

Inspelningar av tal kan användas för att identifiera en individ och taldata innefattas därmed i den nya lagstiftningen GDPR.

De taldata som laddas upp som enskilda filer av ett registrerat användarkonto anses vara enklare att hantera ur GDPR-synpunkt då de vid behov kan raderas *en masse*. Taldata är även möjligt att använda som grund för talmodeller i vilka det inte längre är möjligt att identifiera enskilda individer. Sådana modeller påverkas inte av GDPR och kan tillgängliggöras utan problem.

Även om GDPR kan komma att innebära begränsningar för hur inspelningar av tal kan spridas innebär det inte att projektet inte kommer att kunna skapa ett mycket stort värde med många nya värdefulla svenskspråkiga resurser. Samarbetet med Språkbanken Tal

innebär att projektet kan ta del av professionella juridiska tolkningar som kan guida arbetet.

Sammanställning av komponenter som ska ingå i projektet

Inspelare

Visa prompt

Visa en prompt (text) som användaren ska läsa in. Denna hämtas från ett manus som skapats med *Manusskaparen*. Prompten kan vara slumpmässigt vald eller en av en sekvens av prompter om användaren vill spela in en längre text (se "Spela in" nedan om inspelning av en hel Wikipediaartikel samt om "Manusskaparen").

Indata: id för ett manus och eventuellt index för prompt

Utdata: en text att läsa in

Ansvarig utvecklare: WMSE

Spela in

Spela in tal genom webbläsaren. Den ljuddata som skapas kommer sedan att sparas med *Lagring*, förutsatt att den godkänns av *Automatisk validerare*.

En central komponent i projektet består av ett webbaserat, lättanvänt och inbjudande inspelningsverktyg för talare som läser in manus av olika slag, från ord och enstaka meningar till hela artiklar.

På Wikipedia har man tidigare kunnat spela in Wikipediaartiklar på egen hand, och spara inspelningen som en stor ljudfil, länkad till en viss version av artikeln. Eftersom det inte finns någon koppling mellan artikeltexten och talet, kan man inte navigera i ljudfilen från texten, eller tvärtom. Inspelningen riskerar att snabbt föråldras; för att få med de redigeringar av texten som gjorts sedan artikeln lästs in, måste hela artikeln läsas in på nytt. Detta innebär stora insatser från inläsarna, och är inte effektivt.

Inspelningsverktyget i det här projektet kommer att kunna koppla meningarna i en Wikipediaartikel till motsvarande ljudsekvens i det inlästa talet. På så vis behöver man bara läsa in de meningar som ändrats sedan sist för att hålla den inlästa versionen i fas med artikeln.

Detta underlättar också för navigering, om man exempelvis vill börja lyssna från mitten av en artikel. Det betyder också att man kan använda den ljudspelare för text-till-tal som producerades i det första Wikispeechprojektet för att lyssna på en "manuellt" inläst

Wikipediartikel (skillnaden är att ljudet kommer från sparade filer, istället för att genereras av talsyntesen).

Ytterligare ett skäl till att dela in ljuddata i meningar är att det förenklar olika typer av processning för exempelvis tal-till-text- eller text-till-talsystem. Det gör det också lätt att experimentera till exempel med att ha olika röster som läser olika delar av en text.

För att inläsningen skall flyta på bra, behöver inläsaren kunna välja att läsa in en valfri mängd text, och inte bara mening för mening. Verktøget delar då automatiskt upp inspelningen i ljudfiler som matchar meningarna i texten. Oavsett vilken textenhet användaren valt att arbeta med, kommer verktøget att visa användaren vilken del av texten, eller var i texten, som användaren ska läsa.

Inspelningsverktøget fungerar i moderna webbläsare via The Web Audio API (https://developer.mozilla.org/en-US/docs/Web/API/Web_Audio_API) utan att någon extern mjukvara installeras. Det kommer att utvecklas modulärt, för att kunna användas utanför Wikipedia.

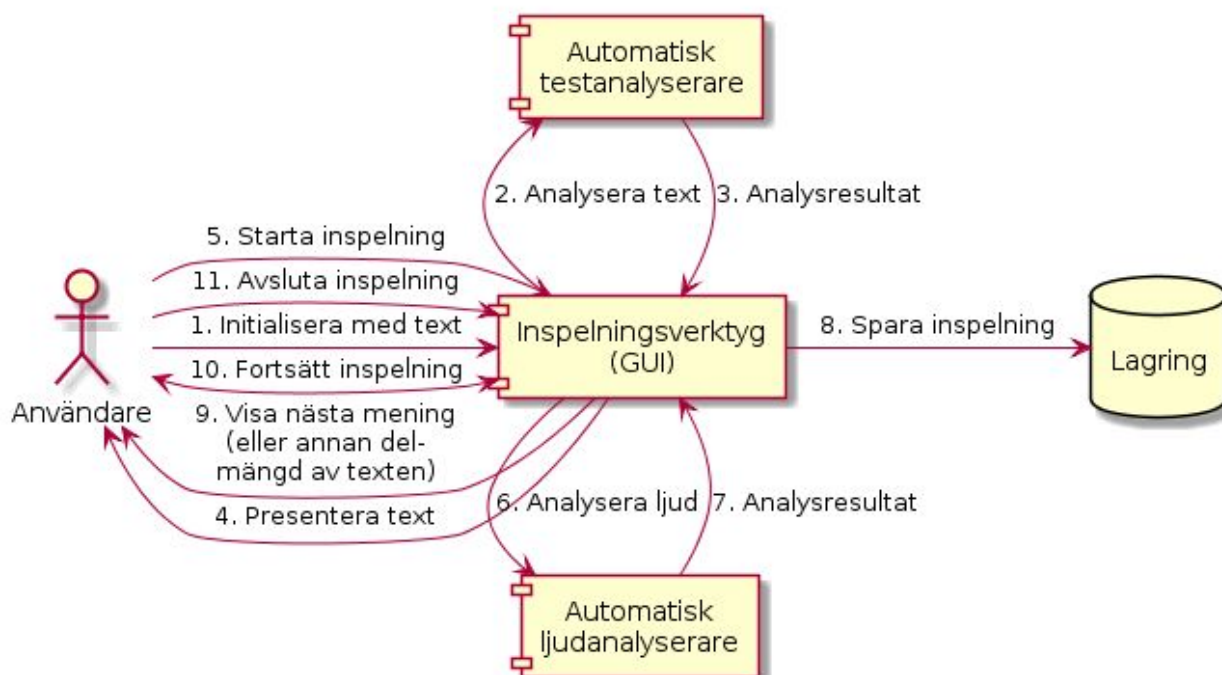
Indata: text från Wikipediaartikel

Utdata: ljuddata bestående av den inlästa inspelningen

Ansvarig utvecklare: WMSE

Taldatainsamlaren

Use case: Spela in Wikipedia-artikel



Uppspelning

Spela upp ljud som har spelats in. Detta gör det möjligt för användaren att kontrollera att det låter bra innan hen väljer att ladda upp det.

Ansvarig utvecklare: WMSE

Automatisk validerare

Validering av signalkvalitet

Automatisk. En enkel signalprocessningsanalys för att diagnostisera talsignalen. Inspelningarna bör hålla en någorlunda god (teknisk) kvalitet, och med hjälp av det här verktyget kan man identifiera eventuella brister i talsignalen.

Indata: ljudsignal

Utdata: numeriska värden för särdrag som kan användas för bedömning

Ansvarig utvecklare: STTS

Normalisering av ljudnivå

Automatisk. Av praktiska skäl bör inspelningarna normaliseras med avseende på ljudnivå. Detta görs i första hand för att undvika obehag och skadligt högt ljud för den som lyssnar (om man höjer volymen vid låg ljudnivå, och sedan får upp en inspelning med hög ljudnivå, kan det bli väldigt högt). Det är också viktigt för att underlätta lyssning, och annan användning samt analys av ljudmaterialet.

Indata: ljudsignal (originalfil)

Utdata: ny, modifierad ljudfil

Ansvarig utvecklare: STTS

Automatisk validering och “sanity check”

Inläsningen valideras automatiskt innan den lagras. I första hand bör det göras någon typ av grundläggande “sanity check” utifrån textunderlaget, för att försäkra sig om att inläsningen sannolikt innehåller den text som avses. Om valideringen tyder på att inspelningen innehåller fel, bör den inte sparas utan vidare granskning (i vissa fall kanske till och med slängas).

Indata: ljudsignal med tillhörande text samt eventuell fonetisk transkription med segmentgränser

Utdata: numeriska eller diskreta värden som indikerar graden av överensstämmelse mellan text och tal

Ansvarig utvecklare: STTS

Manusskapare

Manusgenererare

Ur en stor mängd text väljs en mindre mängd automatiskt ut för att skapa ett manus lämpligt att läsas in. Det kan exempelvis handla om att med så få meningar som möjligt få med så mycket fonetisk variation som möjligt, men det kan också handla om att sälla bort text som av olika anledningar inte är lämpliga för en viss tillämpning.

Indata: Textkorpus, exempelvis hela svenska Wikipedia.

Utdata: Önskat antal meningar som tillsammans utgör ett lämpligt manus för ett visst ändamål.

Ansvarig utvecklare: STTS

Manusinput

För specifika behov kan olika riktade sökningar ske för att samla in material, exempelvis ord inom ett visst område som saknar inspelat tal. Dessa listor kan därefter läggas till i verktyget där orden söks efter på Wikipedia för att plocka ut meningar där de används vilket sedan kan läsas in; alternativt presenteras volontären endast för ordet så att det kan spelas in.

Indata: Textkorpus bestående av i förväg genererade listor av ord inom ett specifikt område. Dessa ord kan komma från exempelvis Wikidata eller från en partner som har ett specifikt behov.

Utdata: Önskat antal meningar som innehåller ordet, vilka tillsammans utgör ett lämpligt manus för ett visst ändamål.

Ansvarig utvecklare: STTS

Lagring

Lagra inspelning

Ladda upp en inspelning, tillhörande prompt och metadata via ett API. Lagra de olika delarna i en databas som kopplar ihop dem och tilldela inspelningen ett id. Ljudet kommer att sparas i ett fritt format. Genomtänkt lagring av ljud och text är viktigt för alla som vill kunna ta del av datan.

Indata: ljuddata, prompt och metadata

Resultat: datan sparas i en databas

Ansvarig utvecklare: WMSE

Hämta inspelningsdata

Hämta inspelningsdata via ett API. I API-anropet anges id för den inspelning som ska hämtas och vilken typ av data som ska hämtas. Data som kan hämtas är ljudet som spelades in, tillhörande prompt och metadata.

Indata: id för inspelning och vilken data som ska hämtas

Utdata: ljuddata, prompt och/eller metadata

Ansvarig utvecklare: WMSE

Ladda upp data till permanenta lagringsplattformar

Ladda kontinuerligt upp inspelningsdata när en inspelning får tillräckligt högt betyg.

Ljuddata laddas upp till Wikimedia Commons. Lägg till kategorier och annan metadata till filen som är kopplat till inspelningen. Data laddas upp via MediaWikis API:er.

Indata: id för en inspelning

Resultat: inspelningsdata läggs till på Wikimedia-plattformar

Ansvarig utvecklare: WMSE

Kassera inspelningsdata

Ta bort all inspelningsdata kopplad till en inspelning.

Indata: id för inspelning

Resultat: ljuddata, prompt och/eller metadata raderas

Ansvarig utvecklare: WMSE

Importera taldata

Ladda upp en större mängd inspelningar från en extern källa. Ljud och prompter konverteras vid behov till stödda format.

Indata: inspelat tal och tillhörande prompter

Resultat: inspelningarna lagras i databasen i rätt format

Ansvarig utvecklare: WMSE

Granskare

Välj inspelning

Presentera en inspelning för användaren. Det går att välja vilket eller vilka språk man vill granska inspelningar för. Det går att hoppa över en inspelning och då få en ny presenterad. Granskaren är viktig för att garantera kvaliteten av insamlad data.

Ansvarig utvecklare: WMSE

Visa prompt

Visa prompten (texten) som användes för att göra inspelningen. Denna används av användaren för att kontrollera att det som spelats in stämmer överens.

Indata: id för inspelning

Utdata: den prompt som visades vid inspelningen

Ansvarig utvecklare: WMSE

Lyssna på inspelning

Spela upp inspelningen. Detta krävs för att användaren ska kunna avgöra huruvida inspelningen låter som den ska.

Indata: id för inspelning

Utdata: ljuddata

Ansvarig utvecklare: WMSE

Ge betyg

Ge ett betyg till en inspelning. Betyget kan vara bra eller dåligt. Om betyget över- eller underskrider ett konfigurerbart tröskelvärde aktiveras *Ladda upp inspelningsdata till permanenta lagringsplattformar* eller *Kassera inspelningsdata*.

Indata: id för inspelning och betyg

Resultat: betyget för inspelningen uppdateras

Ansvarig utvecklare: WMSE

Annoterare

Segmentering

Automatisk uppmärkning av segment i ljudfilen, exempelvis pauser, meningsgränser och i vissa fall även fonemgränser. Om texten spelats in i större sjok delar man automatisk upp inspelningen i mindre enheter motsvarande meningar eller fraser, och kopplar varje enhet till en del av texten. Detta behövs bland annat för att kunna spela upp valda delar av den inspelade texten. För Wikipedia-artiklar gäller också att när texten redigeras behöver man bara spela in de delar som ändrats på nytt, för att täcka hela texten. Uppmärkning av fonemgränser utgår från den fonetiska transkriptionen (ovan), och kan bland annat användas för validering och "sanity check" nedan.

Segmentering av ljudfiler kan betraktas som en form av indexering, och är viktigt för alla som vill kunna hitta i en ljudfil. Med hjälp av segmentering kan man exempelvis koppla meningar, ord och till och med enskilda språkljud (fonem) till exakt position i ljudfilen, vilket möjliggör sökning från en textrepresentation till ljud. Exempel: Om man har märkt ut ordgränser i ljudfilen, kan man klicka på ett ord i en text, och hamna på motsvarande ställe i ljudfilen. Om man har annoterat ljudfilen med fonem och fonemgränser, kan man använda inspelningen för att skapa text-till-tal, etc.

Indata: ljudsignal med tillhörande text, samt eventuellt fonetisk transkription

Utdata: text/transkription uppdelad i segment, tillsammans med tidskodning kopplad till ljudsignalen

Ansvarig utvecklare: STTS

Fonetisk transkription

Det inspelade talet tilldelas automatiskt en transkription baserad på texten. Detta är ofta nödvändigt om inspelningen ska användas för talteknologiska tillämpningar, exempelvis text-till-tal, och viss forskning. Med hjälp av den fonetiska transkriptionen kan man markera, och därmed hitta, enskilda språkljud i ljudfilen. En automatisk transkription blir inte alltid rätt, och det kan därför vara nödvändigt att korrigera den manuellt. De manuellt införda transkriptionerna kan återföras till det generella uttalslexikonet, och därmed komma exempelvis text-till-talsystemet tillgodo.

Indata: inspelat tal med tillhörande text

Utdata: fonetisk transkription (i textformat)

Ansvarig utvecklare: STTS

Konverterare för annoteringsformat/-standard

Om olika annoteringsformat förekommer inom ramen för projektet, behövs ett verktyg för automatisk konvertering mellan formaten. Dessa tas fram vid behov.

Indata: uppmärkning i ett visst format

Utdata: motsvarande uppmärkning i valt output-format

Ansvarig utvecklare: STTS

Publicerare

Lägg till uttal i Wikipediaartikel

Lägg till uttal för hur titeln på en Wikipediaartikel uttalas. Detta är användbart för artiklar där titeln är ett egennamn, t.ex. om den handlar om en person eller namngiven plats. En bot (d.v.s. ett automatiserat skript) kommer att byggas som, med hjälp av API:er inbyggda i MediaWiki, skapar en länk till ljudfilen med inspelningen på Wikimedia Commons.

Indata: id för en inspelning på Wikimedia Commons

Resultat: uttalslänk till inspelning läggs till i artikel

Ansvarig utvecklare: WMSE

Lägg till uttal för ord på Wiktionary

Lägg till en länk till en inspelning av ett ord på Wiktionary-uppslaget för det ordet. Detta gör det lätt att få en bild av hur ett ord uttalas utan att behöva kunna läsa fonetiska transkriptioner. På Wiktionary finns stöd för att på ett standardiserat sätt ange ljud för uttal genom en mall. Denna går att skapa eller utöka med hjälp av en bot som använder MediaWikis API:er. Finns fonetisk transkription för inspelningen kan också denna läggas till, om den saknas.

Indata: id för en inspelning på Wikimedia Commons

Resultat: länk till inspelning och eventuell fonetisk transkription läggs till på Wiktionary

Ansvarig utvecklare: WMSE

Lägg till data på Wikidata

Lägg till länk till inspelat uttal och annoteringar (t.ex. språk och fonetisk transkription) på Wikidata. Inspelningar kan läggas till för lexem (uppslagsord) och items (t.ex. för egennamn). Datan läggs till med hjälp av en bot som använder MediaWikis API:er.

Indata: id för en inspelning på Wikimedia Commons

Resultat: informationen läggs till på Wikidata

Ansvarig utvecklare: WMSE

Exporterare

Ladda ner arkiv

Ladda ner en mängd taldata paketerat som ett arkiv. Det går att välja vilka inspelningar (t.ex. efter språk) och vilken typ av data som ska ingå.

Indata: specifikationer för vilken data som ska laddas ner

Utdata: ett arkiv som innehåller taldata

Ansvarig utvecklare: WMSE

Komponenter utanför projektet

Vi har även identifierat ett antal komponenter som vi inte kommer att fokusera på under projektet:

- *Ortografisk transkription av inspelat tal.* Vi arbetar med scenariot att texten är en utgångspunkt för inspelningen. Vi kommer dock att beakta hur ljudfiler utan text skulle kunna hanteras och om möjligt bygga ett system som är flexibelt nog för att hantera även detta.
- *Kompletterande uppmärkning och metadata.* Det kan finnas behov/intresse av andra typer av uppmärkning, som vi valt att inte prioritera i den aktuella ansökan. Om tid finns och det finns intresse från volontärgemenskapen kan vi komma att komplettera med detta. Exempel:
 - Parametrisering (även anonymisering)
 - Röstkvalitet
 - Talarkaraktäristik
 - Talstil
 - Kategorisering
 - Prosodisk analys