

EXTRA utlysning

Ansökan om bidrag för insatser till stöd för de nationella minoritetsspråken 2022

Ansökan sänds med e-post till revitalisering@isof.se eller med post till

Institutet för språk och folkminnen, Box 135, 751 04 Uppsala.

Sista ansökningsdagen är den 3 oktober 2022.

Sökande/bidragsmottagare

Organisation (huvudansvarig för projektet) Wikimedia Sverige		
Typ av organisation <input checked="" type="checkbox"/> Ideell för. <input type="checkbox"/> Stiftelse <input type="checkbox"/> Annan	Organisationsnummer 802437-8310	Bildades år 2007
Adress Wikimedia Sverige, c/o Norrsken House, Birger Jarlsgatan 57C	Postnummer 113 56	Postort Stockholm
Kontaktperson John Andersson	Roll i föreningen Verksamhetschef	
Telefon 08-409 05 200 John: 073-3965189	E-post för kontakt med föreningen info@wikimedia.se John: John.andersson@wikimedia.se	Ev. webbplats https://wikimedia.se/

Eventuell medsökande

Organisation	Organisationsnummer	
Adress	Postnummer	Postort

Projektamn och tidsperiod

Projektets namn Wikispeech: Ökad tillgänglighet på det egna (minoritets)språket	Startdatum 1 december 2022	Slutdatum 31 mars 2024
Projektledare (om utsedd) André Costa	Telefon 073-396 45 74	E-postadress andre.costa@wikimedia.se
Är projektet flerårigt? <input checked="" type="checkbox"/> Ja <input type="checkbox"/> Nej		

EXTRA utlysning

Ansökan om bidrag för insatser till stöd för de nationella minoritetsspråken 2022

Sökt belopp

Bidrag som söks för projektet (kronor)
400 000 kr

Medel som söks för projektet från andra finansiärer

Annan finansiär	Sökt belopp	Beslutsdatum	Beviljat belopp
Wikimedia Foundation	96 480 kr	Mars 2023	
Annan finansiär	Sökt belopp	Beslutsdatum	Beviljat belopp

Information om projektet

Språk som ansökan avser

Finska | Jiddisch | Meänkieli | Romska | Samiska | Flerspråkigt

Om språket har flera dialekter/varieteter, vilken av dem avser ansökan?

Nordsamiska

Kort beskrivning av projektet (max 4–5 rader)

Wikimedia Sverige har sedan 2015 ansvarat för utvecklingen Wikispeech, en fritt licensierad talsyntes för Wikipedia. Wikispeech är sedan 2021 aktiv som alfa-release (d.v.s. i experimentellt syfte) på svenskspråkiga Wikipedia. Vi vill inom detta projekt lägga grunden för att aktivera Wikispeech på finska samt lägga en grund för aktivering på nordsamiska. För finska finns mer material vi kan återanvända medan det för nordsamiskan kräver mer experimentellt och grundläggande arbete med inkludering av ordlistor, inspelat tal och annat material.

Syfte – varför behövs projektet?

Projektets syfte är att genom tillgänglighetslösningen Wikispeech ge talare av minoritetsspråken, som har en funktionsvariation eller lär sig bättre av att lyssna, möjligheten att tillgodogöra sig kunskapen på Wikipedia på det egna språket. Genom att nyttja crowdsourcing kan vi tillsammans hjälpas åt att förbättra Wikispeech och öka värdet av tjänsten. Då allt material är fritt licensierat och kan återanvändas, även kommersiellt, så skapar projektet värden även utanför Wikispeech.

Mål – vad hoppas ni uppnå med projektet?

1. Inkludering av finska på Wikispeech för att aktivering på finskspråkiga Wikipedia ska vara möjligt.
2. Godkännande från volontärerna på finskspråkiga Wikipedia att aktivera Wikispeech där.
3. Experimentera med insamling av olika språkresurser för att kunna aktivera Wikispeech på nordsamiska.

EXTRA utlysning

Ansökan om bidrag för insatser till stöd för de nationella minoritetsspråken 2022

4. Etablering av minst två samarbeten för fortsatt vidareutveckling av tjänsten samt delning av de talresurser som samlas in.

5. Publicera en rapport om erfarenheterna under punkt 1-4 för att kunna genomföra liknande aktiviteter med andra minoritetsspråk i framtiden.

Målgrupp – för vem görs projektet?

Projektet gör ett viktigt bidrag för att göra Wikipedia mer tillgängligt för de som talar men inte kan läsa på sitt språk. I fokus är finsk- samt nordsamisktalande i Sverige.

Tekniken som utvecklas kan dock även fritt nyttjas av andra webbplatser och gagnar därför finsk- och nordsamisktalande internetanvändare i bred bemärkelse.

Om träffar: Hur många kommer att delta?

10-15 deltagare.

Om träffar: Hur ofta och hur många gånger kommer träffarna att äga rum? Var?

Ett-fem digitala (troligtvis) träffar med företrädare från minoritetsspråksorganisationer, akademien, företag samt myndigheter för att diskutera projektet.

Genomförande – hur kommer ni att genomföra projektet? Vilka konkreta aktiviteter kommer att ingå?

I projektets inledning kommer två grupper från Wikimedia Sverige att arbeta parallellt: Utvecklargruppen och Talresursgruppen.

Utvecklargruppen kommer inledningsvis på att fokusera på att förbereda så att finska språkresurser (såsom lexikon, röster, grammatiska regler m.m.) kan läggas till i Wikispeech. Nya typer av språkmotorer, såsom <https://github.com/synesthesiam/opents>, kommer att behöva aktiveras. När detta är gjort kommer de att arbeta för att lansera Wikispeech på finskspråkiga Wikipedia där de kommer att behöva hantera rapporter om olika typer av fel (något som alltid sker när en ny mjukvara lanseras). Ett nära samarbete sker här med Södermalms talteknologiservice (STTS), en organisation som vi arbetat med i flera år för att öka tillgängligheten på Wikimedias plattformar.

Utvecklargruppen kommer därefter att fokusera på att utveckla nya funktioner (verktyg) som gör det möjligt att lägga till existerande språkresurser för nordsamiska samt att aktivt börja samla in nya språkresurser genom crowdsourcing. Att samla in nya språkresurser innebär ex. att vi ber att de som läser och talar språket att läsa in texter och olika specialord. Vi kan använda deras uttal för att bygga bättre lexikon samt att skapa nya röster (kvinnliga eller manliga röster, röster med olika dialekter m.m.). För att effektivisera arbetet så mycket som möjligt kommer ett manusverktyg att utvecklas. Manusverktyget gör att mindre taldata behöver samlas in då manuset säkerställer att alla olika fonem (språkljud) i språket samlas in vid en enskild inspelning istället för att spridas ut över ett stort antal inspelningar. Färre timmar inspelad taldata behövs därför för att kunna aktivera Wikispeech på ett nytt språk. Även runt utvecklingen av manusverktyget kommer STTS att göra en punktinsats.

En annan viktig komponent i Wikispeech är lexikonet, där uttalet av olika ord beskrivs genom IPA. Även om det finns generella uttalsregler inbyggda i språkmodellen så ökar lexikonet chansen att orden

EXTRA utlysning

Ansökan om bidrag för insatser till stöd för de nationella minoritetsspråken 2022

uttalas på korrekt sätt. Utvecklargruppen kommer därför att utveckla ett nytt finskt och samiskt lexikon där vi drar nytta av s.k. lexeme på Wikidata.org, en databas som drivs av Wikimediarelsen.

Verktyget och tillhörande dokumentation kommer att översättas till finska och nordsamiska för att förenkla användningen av dem för minoritetsspråkstalarare.

Den andra gruppen med personal är *Talresursgruppen*. De arbetar för att identifiera och samla in olika typer av talresurser som är relevanta att stoppa in i den mjukvara som Utvecklargruppen bygger. Det kan ex. handla om ordlistor, framförallt med uttal men även ex. med orsnamn. Ordlistorna kommer att inkluderas i lexikon-komponenten, vilken behövs för att talsyntesen ska fungera på finska eller nordsamiska. Ordlistorna kan även användas som grund för crowdsourcing där de som talar språket ombeds att göra ljudinspelningar när de läser upp ordet eller namnet. Genom dessa inspelningar av olika ord kan vi säkerställa att uttalet i talsyntesen är korrekt och vid behov göra rättningar och tillägg i lexikonet (jämför ex. med hur svårt det kan vara att förstå vilken gata som GPS:en hänvisar en till när det uttalas helt felaktigt då lexikonet i GPS:en är inkomplett).

Talresursgruppen kommer även att arbeta för att inkludera olika typer av fria nordsamiska texter på plattformen Wikisource.org, som drivs av Wikimediarelsen. Precis som ordlistorna har dessa resurser ett egenvärde men de kan även användas som grund för inläsning där det inspelade talet kan användas för att vidareutveckla Wikispeech. Det inspelade talet som bygger på de fria nordsamiska texterna kan då användas för att generera ett manus för manusverktyget.

Vi har inlett samtal med både Google och Europeiska kommissionen om samarbeten runt olika språkresurser. Föreningen har dock pausat diskussionerna då vi inte kunnat gå vidare med de idéer vi diskuterat på grund av bristande resurser. Detta projekt kommer att möjliggöra för oss att återuppta dessa diskussioner och initiera arbetet. Det har potentialen att leda till att det crowdsourcade material som projektet genererar skulle få en mycket större spridning och gagna även andra plattformar. Vi kommer även att fortsätta det nära samarbete vi har med Kungliga tekniska högskolan (KTH) för att diskutera möjlig återanvändning av de talresurser vi samlar in för deras språkforskning.

På vilket sätt bidrar projektet till att stärka deltagarnas språkförmåga?

Att kunna konsumera relevant material på det egna språket är viktigt för att nya generationer ska anamma det. Wikipedia är flerspråkigt och tillgängligt på flera av de svenska minoritetsspråken, vilket är grunden för det här projektet. Wikipedias användare kan därigenom ta till sig kunskap på sitt eget språk. Plattformen är dock textbaserad och de användare som av olika orsaker inte kan läsa (p.g.a. synnedsättning, dyslexi m.m.) stängs då ute från den möjligheten att kontinuerligt använda det egna språket när de konsumerar information. Genom Wikispeech öppnas den möjligheten upp för dem.

Projektet bidrar även till att stärka språkförmågan hos minoritetsspråkstalarare som inte har svårigheter att läsa, då språkresurserna kan användas för deras språkträning. Lexikon, fria texter m.m. skapar andra värdefulla språkresurser för att stärka språkförmågan. Det tillgängliga materialet gör det därigenom möjligt att knyta an till sitt kulturella och språkliga arv.

EXTRA utlysning

Ansökan om bidrag för insatser till stöd för de nationella minoritetsspråken 2022

Beskriv på vilket sätt ni tänker följa upp och utvärdera projektet. Hur vet ni om projektmålen har uppnåtts?

Om Wikispeech aktiveras på finskspråkiga Wikipedia har det första och andra målet uppnåtts fullt ut. Om Wikispeech inte har aktiverats då stöd från gemenskapen saknas kommer vi att utvärdera hur vi ligger till samt vad som kommer att krävas för full aktivering.

Språkresurserna som omnämns i mål 3 kommer att utvärderas både baserat på deras antal samt en bedömning av deras kvalitet (är de ex. ovanliga eller extra värdefulla av någon anledning). Det kan exempelvis handla om hur många fria texter vi kunnat inkludera eller antalet ord vi kunnat samla in.

Vi kommer att utvärdera kvaliteten på de lexikon vi skapat för finska och nordsamiska och hur väl integrerade de är med Wikispeech samt om lexikonerna kan återanvändas av andra (ex. KTH, Google och Europeiska kommissionen).

Manusverktyget, som är en nödvändig del för insamling av språkresurser, kommer att bedömas baserat på om det aktiverats eller ej. Vi kommer även att beskriva hur långt vi kommit med manusverktyget och om det återstår specifika funktioner vi planerar att vidareutveckla efter projektet.

Vi kommer att utvärdera måluppföljningen runt mål 4 baserat på om det finns en överenskommen plan runt hur arbetet kan komma att se ut framöver och/eller om det finns en plan för vidareanvändning av den språkdata som vi samlar in både genom samarbeten och crowdsourcing och om det finns material från ex. Europeiska kommissionen, Google eller KTH som vi kan återanvända på kort eller lång sikt.

Den rapport vi tar fram för att uppnå mål 5 kommer vi att utvärdera baserat på den spridning den får (ex. genom presentationer på konferenser). Spridningen kommer att rapporteras efter projektslut, som en integrerad del av föreningens årliga verksamhetsrapport.

Kostnadssammanställning för projektet

Roll i projektet	Beräknade lönekostnader				
	Månadslön, brutto	% av heltid	Summa per månad	Antal månader	Totalt i kr
Projektledare	52 500	15 %	7 875	16	126 000
Utvecklare	43 200	15 %	6 480	13	84 240
Talresursinsamlare	43 200	15 %	6 480	13	84 240
Summa lönekostnader exkl. sociala kostnader					294 480
Sociala kostnader (arbetsgivaravgift m.m.)					102 000
Summa lönekostnader inkl. sociala kostnader					396 480

EXTRA utlysning

Ansökan om bidrag för insatser till stöd för de nationella minoritetsspråken 2022

Övriga kostnader för projektet (specificera i kr)	
Material och utrustning	10 000
Lokaler	
Kontorsplats 16 månader á 5 000 kr	80 000
Övriga kostnader	
Översättare meänkieli: ca. 2 500 ord á 2 kr = 5 000 kr	
Översättare nordsamiska: ca. 2 500 ord á 2 kr = 5 000 kr	10 000
Summa övriga kostnader	100 000

Tidigare beviljade projekt från Institutet för språk och folkminnen

Projektnamn	År för ansökan	Beviljat belopp
N/A	N/A	N/A

Övriga upplysningar

<p>Övriga upplysningar Om finansiering inte mottas från Wikimedia Foundation kommer Wikimedia Sverige att stå för de 96 480 kronorna som egenfinansiering.</p>

Bilagor

<p>Dokumentation som visar organisationens ordförande/firmatecknare <u>ska bifogas till ansökan</u> Ladda upp eller skicka med ansökan ett justerat protokoll från det senaste årsmötet alternativt ett justerat protokoll som visar vem eller vilka som har valts till organisationens firmatecknare.</p> <p><input checked="" type="checkbox"/> Justerat årsmötesprotokoll som visar organisationens ordförande</p> <p><input type="checkbox"/> Justerat protokoll som visar organisationens firmatecknare</p>
--

Vår förening respekterar demokratins idéer, som till exempel jämställdhet och förbud mot diskriminering.

Underskrift

Med denna underskrift intygar jag att uppgifterna som lämnas i ansökan är korrekta och motsvarar riktiga förhållanden i organisationen.


Personuppgiftsbehandling

Läs om hur institutet behandlar personuppgifter på <http://www.isof.se/personuppgifter>.

Ort Stockholm	Datum 30 september 2022
------------------	----------------------------

EXTRA utlysning

Ansökan om bidrag för insatser till stöd för de nationella minoritetsspråken 2022

Underskrift och försäkran av firmatecknare/ordförande 	Namnförtydligande John Andersson
--	-------------------------------------

Alla projekt som beviljas medel kan komma att bjudas in till ett seminarium för erfarenhetsutbyte och/eller bli besökta av institutet. Eventuella kostnader för detta ska inte medräknas i projektets budget.

Verifikat

Transaktion 09222115557478397293

Dokument

ISOF – Ansökningsblankett extra utlysning 3 oktober
2022 – Wikispeech_ Ökad tillgänglighet på det egna
(minoritets)språket.docx

Huvuddokument

7 sidor

Startades 2022-09-30 15:16:51 CEST (+0200) av John
Andersson (JA)

Färdigställt 2022-09-30 15:17:28 CEST (+0200)

Signerande parter

John Andersson (JA)

Wikimedia Sverige

Personnummer 198603278675

Org. nr 802437-8310

john.andersson@wikimedia.se

+46733965189



Signerade 2022-09-30 15:17:28 CEST (+0200)

Detta verifikat är utfärdat av Scrive. Information i kursiv stil är säkert verifierad av Scrive. Se de dolda bilagorna för mer information/bevis om detta dokument. Använd en PDF-läsare som t ex Adobe Reader som kan visa dolda bilagor för att se bilagorna. Observera att om dokumentet skrivs ut kan inte integriteten i papperskopian bevisas enligt nedan och att en vanlig papperutskrift saknar innehållet i de dolda bilagorna. Den digitala signaturen (elektroniska förseglingen) säkerställer att integriteten av detta dokument, inklusive de dolda bilagorna, kan bevisas matematiskt och oberoende av Scrive. För er bekvämlighet tillhandahåller Scrive även en tjänst för att kontrollera dokumentets integritet automatiskt på: <https://scrive.com/verify>

