

Tittel: 2022 - Språk - Søknadsskjema

➤ Søkeren

Kontaktperson

Navn OLE ESPEN KLEM
E-postadresse espen.klem@knowit.no
Telefon 40221940

Organisasjon

Navn	KNOWIT AMENDE AS
Adresse	Universitetsgata 7 0164 OSLO Norge
Organisasjonsnummer	991513833
E-postadresse	post@knowit.no
Telefon	02486

➤ Om prosjektet

Navn på prosjektet

Stoppordlister for nordsamisk, lulesamisk og sørsamisk for bedre digital samisk språkforståelse

Beskrivelse av prosjektet

Prosjektet skal identifisere stoppord for nordsamisk, lulesamisk og sørsamisk for bedre digital, samisk språkforståelse.

Stoppord er ord som har lite meningsbærende innhold. De er derfor lite egnet til bruk i digital tekstanalyse. Eksempler fra bokmål er “og”, “eller”, “men”, “for”, “å”, i tillegg til mange andre.

Fjerning/identifisering av stoppord for et språk er en enkel måte å gjøre digitale løsninger som søkemotorer, chatbot'er, maskinlæring og duplikat-/plagiat-identifisering bedre.

Mål med prosjektet

Prosjektets målsetting er å kunne generere gode stoppordlister for nordsamisk, lulesamisk og sørsamisk under en åpen kildekode-lisens for alle som vil til å benytte seg av. Dette så hvem som helst skal kunne lage/utvikle gode teknologiløsninger hvor du trenger en form for språkforståelse for de samiske språkene nordsamisk, lulesamisk og sørsamisk.

Begrunnelse for søknaden

Programmeringen for prosjektet er allerede gjort. Arbeidet som gjenstår og vi søker penger om er:



- Luke ut artikler med norsk tekst
- Finne personer som kan lule- og sørsamisk. For nordsamisk har vi allerede en internt som kan bistå
- Koordinere arbeidet med disse personene
- Gjennomgå automatisk genererte lister og rødliste ord som ikke skal være med
- Definere lengde på ferdige lister
- Generere ferdige lister
- Siste kvalitetskontroll av listene
- Publisere stoppordslistene
- Bloggpost om prosjektet

Hvilken språkgruppe?

Nordsamisk

Målgruppe

Flerkulturelt

Hvordan bidrar prosjektet til å fremme samisk språk?

Prosjektet vil gjøre det lettere å lage digitale løsninger med samisk språkforståelse. De digitale løsningene kan være søkemotorer, chatbot'er i tillegg til andre maskinlærings-løsninger. Prosjektet publiseres under den åpne kildekode-lisensen MIT som gjør at alle kan benytte seg av prosjektet i egne prosjekter. Dette gjelder uansett om dette er åpen eller lukket kildekode.

Planlagt dato for ferdigstillelse

31.12.2022

Fremdriftsplan

	Aktivitet	Fra dato	Til dato
1	Gjennomgang av innhold hentet fra NRK og luke ut kort-artikler med norsk tekst.	01.09.2022	07.09.2022
2	Finne mennesker som kan lulesamisk og sørsamisk - ansvarlige for språkforståelsen.	01.09.2022	30.09.2022
3	Forberede lyntale til NDC Oslo	01.09.2022	23.09.2022
4	Koordinere arbeidet med personene ansvarlig for språkforståelse. Forklare hva stoppord er og få dem til å gå gjennom listene	01.10.2022	07.10.2022
5	Gjennom automatisk genererte lister og melde tilbake ord som burde være rødlistede og hvor lange de ferdige listene kan være	10.10.2022	22.10.2022
6	Legge ord til rødlistene	24.10.2022	28.10.2022
7	Definere hvor vi skal sette grensene for hver enkelt stoppordsliste. Kort diskusjon med hver enkelt av de ansvarlige for språkforståelse	31.10.2022	04.11.2022

Aktivitet	Fra dato	Til dato
8 Generere ferdige lister, klare til publisering	07.11.2022	11.11.2022
9 En siste kvalitetskontroll av listene	14.11.2022	18.11.2022
10 Publisere stoppordslistene på stopword-sami-kodebiblioteket.	21.11.2022	25.11.2022
11 Generere tester og publisere stoppordslistene på stopword-kodebiblioteket	28.11.2022	02.12.2022
12 Bloggpost om prosjektet og hva det ferdige resultatet betyr	05.12.2022	16.12.2022

Tidligere tilskudd fra offentlige finansieringskilder de tre siste årene?

Nei

➤ Budsjett

Søknadsbeløp

73 500

Finansiering

Finansieringskilde	Sektor	Type	Status	Beløp
Sametinget	Offentlig	Tilskudd	Skal søke	73 500 1)
Total	100% Off.		0% Bekr.	73 500

1) Saksnr 12332

Kostnader

	Beløp	Kommentarer
Utgifter	73 500	
Gjennomgang av innhold hentet fra NRK og luke ut kort-artikler med norsk tekst.	6 000	4 timers arbeid
Finne mennesker som kan lulesamisk og sørsamisk. Ansvarlig for språkforståelsen.	12 000	8 timers arbeid
Forberede lyntale til NDC Oslo	4 500	3 timers arbeid
Koordinere arbeidet med språkressursene.	9 000	6 timers arbeid
Gjennom automatisk genererte lister og melde tilbake ord som burde være	13 500	9 timers arbeid

	Beløp	Kommentarer
rødlistede og hvor lange de ferdige listene kan være		
Legge ord til rødlistene	3 000	2 times arbeid
Definere hvor vi skal sette grensene for hver enkelt stoppordsliste. Kort diskusjon med hver enkelt av de ansvarlige for språkforståelse.	4 500	3 timers arbeid
Generere ferdige lister, klare til publisering.	1 500	1 times arbeid
En siste kvalitetskontroll av listene.	4 500	3 timers arbeid
Publisere stoppordslistene på stopword-sami-kodebiblioteket.	1 500	1 times arbeid
Generere tester og publisere stoppordslistene på stopword-kodebiblioteket	4 500	3 timers arbeid
Bloggpost om prosjektet og hva det ferdige resultatet betyr	9 000	6 timers arbeid
Total	73 500	

> Vedlegg

Fullstendig prosjektbeskrivelse

Andre

 [application-sametinget.pdf](#)

Oversetting eller tilpasning av applikasjoner til samisk?

Nei

Samiskkurs?

Nei

Andre vedlegg

Andre

Ingen vedlegg lastet opp

> Egenerklæring

Personvern

Jeg bekrefter at jeg er kjent med Sametingets personvernerklæring

Offentlighetsloven

Jeg er innforstått med at Sametinget følger offentlighetsloven, og at alle saksdokumenter registreres i offentlig postjournal på innsyn.