

A photograph of three people sitting on a blue sofa in a modern, brightly lit room with large windows. They are all holding and reading a magazine titled 'LUMAGASIN'. The person on the left is a man with a shaved head wearing a black jacket and light-colored trousers. The person in the middle is a woman wearing a grey cardigan over a dark top and dark trousers. The person on the right is a man with dark hair wearing a black t-shirt and grey trousers. The magazine covers feature various images, including a person in a white lab coat, a person in a red shirt, and a person in a blue shirt. The text 'Transkribering on prem' is overlaid in large white letters across the center of the image.

Transkribering on prem

Forskningsstöd IT Gbg 2023-11-07



LINKÖPINGS
UNIVERSITET

Bakgrund

- Inom forskning finns ett stort behov av transkribering av ljudfiler.
- De innehåller i de flesta fall personliga uppgifter och ofta också känsliga sådana.
- De flesta tillgängliga lösningar är molntjänster men på senare tid har bland annat Microsoft tillgängliggjort containers som gör det möjligt att köra en lokal transkriberingsmotor på ett relativt enkelt sätt.
- Medverkande:
 - Linköping: Lena Strömbäck, Johan Peterson, Lennart Johansson, Marcus Liw, Andreas Larsson, Bo Taube Nicolin.
 - Umeå: Mikael Gustafsson, Per Hörnblad
 - Göteborg: Ola Ljungkrona
 - Luleå: Marcus Nättedal

Nuläge

	Linköping	Umeå
Manuell transkribering	Använder till exempel DOTE eller ELAN	Använder Olympus dss + pedal eller OCTRA
Automatisk transkribering	Rekommenderar första hand M365s verktyg för icke känsligt material	
Inhyrd hjälp	Anlitar till exempel studenter och personal från regionen. Inget centralt avtal med transkriberingsbyrå.	

Behov – bättre stöd av transkribering för känsliga intervjuer

Mål med projektet

1

Ta fram en lösning för automatisk transkribering som driftsätts på minst två lärosäten.

2

På sikt skall lösningen kunna paketeras och driftsättas för att användas av flera lärosäten.

3

Ta fram dokumentation och instruktioner för att sätta upp lösningen.

4

Ge förslag på vidare aktiviteter för att lösningen skall kunna produktifieras och förvaltas som gemensam tjänst för lärosäten.

Kravinsamling

	Linköping	Umeå
Format	.wav .flac .mp3 .ogg .m4a .wma, dss, mm	wma, mp3, wav
Omfattning	1 timme eller mer per fil. Ofta 10-50 sådana i en studier.	Dialogformat, två till 8 talare, 10 sessioner á 1h st.
Frekvens	Forskarna arbetar med transkribering i en fas om ett par månader per projekt. (Dvs datainsamling – transkribering – analys i ett projekt.)	Sällananvändning, typiskt stintar 1ggr/år
Språk	Svenska (dialekter, bruten svenska, barnspråk), engelska, andra språk som till exempel somaliska. Ofta dialoger och gruppsamtal.	Svenska (med dialekter) och engelska

Önskemål funktionalitet

Linköping	Umeå
Önskemål om att skicka med domänspecifika ord.	Skicka med domän-ord till transkriberingsmotor
Önskemål om att kunna identifiera talare i en grupp	Röstigenkänning med talar-tagging.
Vissa forskare vill hålla data på sin egen dator/hårddisk.	Webtjänst ok för de flesta, lånad dator med transkribering för paranoida.
Många vill transkribera själva och ser detta som en möjlig hjälp. Möjlighet till interaktivitet eller efterbearbetning viktig.	Interaktivt, möjlighet att spela upp ljud och ändra i transkribering.
I många projekt är alla detaljer viktiga, harklingar, pauser, tvekan. I andra sammanhang vill man få fram det som sades på ett enkelt sätt.	Byta läge mellan att: Ta med "Rumskänsla", pauser och humör till Filtrera för bra flytande språk utan harklingar, stakningar och ofullständiga meningar.

Övergripande funktionalitetsbeslut

Så enkel lösning som möjligt med möjlighet till senare vidareutveckling.

Webtjänst mot en lokal server på lärosätet där transkriberingen sker.

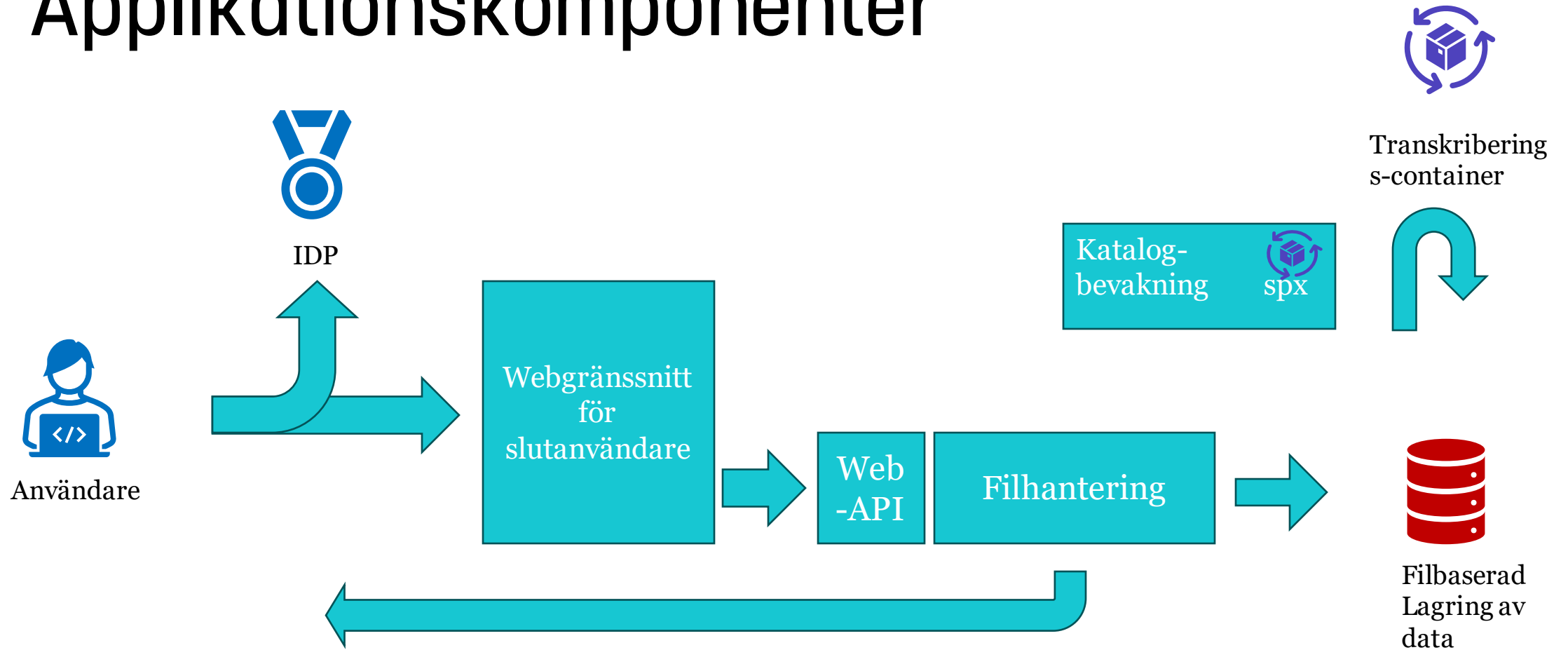
Användaren loggar in och laddar upp filer och ordlista för det som ska transkriberas.

Inga filer lagras längre än nödvändigt. Ljudfiler och resultat tas automatiskt bort.

Möjlighet att debitera användaren förbereds.

Ingen efterbearbetning eller liknande i detta projekt. (Nedladdning till andra verktyg en möjlig lösning)

Applikationskomponenter




Leverabler

- Kod och documentation
 - <https://github.com/ITCF-projects/Transcription>
 - Öppen att ladda ner och vidareutveckla
- Lösning i test vid Linköping och Umeå.
 - Många förfrågningar om transkribering när man set funktionaliteten i AI-system
- Plan för driftsättning i säker miljö som tillåter känsliga personuppgifter.
- Förslag på vidare aktiviteter för att lösningen ska kunna produktifieras och förvaltas på flera lärosäten eller som en gemensam tjänst.





Testversion tillgänglig i Linköping och Umeå

Transcribe

AI-powered transcription service that prioritizes the safety and security of your data

 Create new transcription
Upload a file

My transcriptions

Filename	Language	Status	Upload date ↓	
Röst 007_sd (2).m4a	English	 Completed	2023-12-15 16:39	Download  >
Röst 006_sd.m4a	English	 Completed	2023-12-15 16:04	Download  >

A COLLABORATION BETWEEN

 LINKÖPING UNIVERSITY  UNIVERSITY OF GOTHENBURG

 UMEÅ UNIVERSITY  LULEÅ
TEKNISKA
UNIVERSITET

Lite axplock från användartest

- 8 användare 4 från LiU 4 från UmU

4. Hur karakteriserar du språket på dina inspelningar (0 poäng)

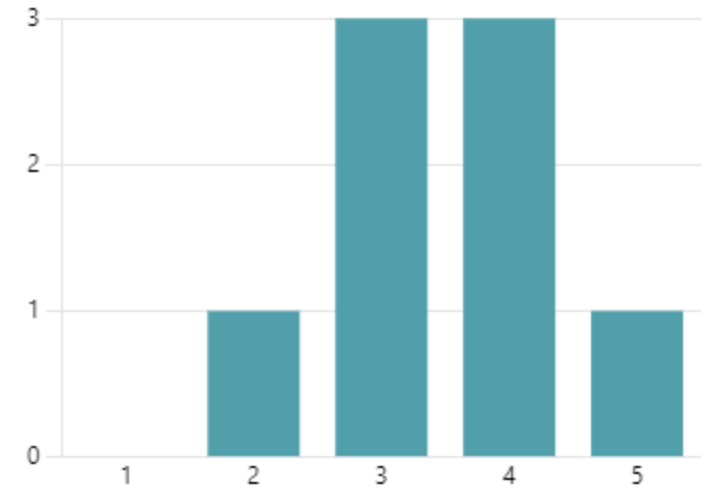
[Mer information](#)

● Det är i första hand vad som ka...	2
● Dialekter förekommer	4
● Barnspråk förekommer	1
● Annat	1



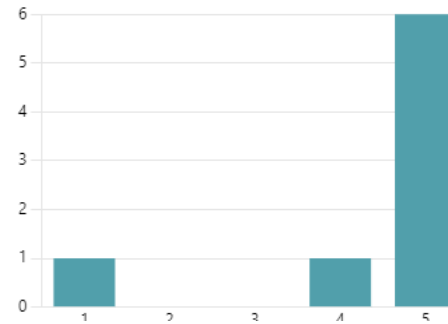
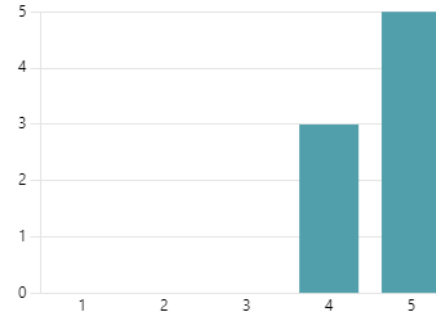
Transkriberingskvalitet

- Positivt:
 - Fångade alla ord, korrekt svenska,
 - Går att få ett sammanhang
 - Positivt överraskad, men kräver genomgång efteråt
 - Fångar upp det som sades
- Negativt:
 - Saknar information om talare, ny rad vid varje röst
 - Texten delas konstigt i flera delar
 - Svårt när fler pratar samtidigt
 - Gissade ord som blir felaktiga



Systemet

- Enkelt att komma igång med
- Format på Ljudfiler
- Format på utdatafiler



Skulle du använda systemet om det infördes?

- De flesta inser att efterarbete krävs.



- Bra format för efterbearbetning är önskat.
- Önskar talarigenkänning.
- Nöjda med att systemet klarade dialekter.
- Smidigt att ladda upp.
- Positivt med tidsstämplar.
- Enkelt att använda.

Driftsättning

- Planer vid de olika universiteten
 - Linköping: Vill komma i drift under året
 - Umeå: Vidare utredning och utveckling för att komma i drift
 - Luleå: Vill komma igång med test för att utreda behov
 - Göteborg: Har Whisperbaserad lösning

Produktifiering och förvaltning

- Effektiv förvaltning: olika driftsmiljöer ställer olika krav
- Administration av tjänsten behöver utvecklas för effektiv förvaltning
- Gemensam drift ställer ytterligare krav på inloggning (SWAMID) och säkerhet

Transkribering steg 2

- Vidareutveckla lösningen med avseende på felhantering och förvaltning. I testversionen av tjänsten loggas transkriberingstiderna men det finns ingen möjlighet att se dessa för en användare. Vi vill förbättra strukturen på dessa loggar och införa en administratörsroll så att det går enkelt att ta ut ett underlag för debitering från systemet. Vi behöver också få en möjlighet att logga fel och införa driftövervakning för att öka driftsäkerheten.
- Vidareutveckla funktionaliteten för att möta en del av de synpunkter vi fick in vid användartesterna. Det innebär bland framförallt viss efterbearbetning som gör det enklare att ladda upp resultatet till andra verktyg, till exempel DOTE, där forskaren kan jobba vidare med det.
- Förbättring av köhanteringen som ger bättre möjlighet att förutsäga när min transkribering är färdig. Man kan vid behov även införa möjlighet till flera språkmotorer för att ytterligare öka effektiviteten.
- Undersöka förutsättningar för ytterligare förbättringar av transkriberingskvalitén. Bland annat har önskemål om talaridentifiering kommit in, men det är oklart om och hur detta kan realiseras i Microsofts containerlösning.
- Undersöka förutsättningar för gemensam drift för lärosäten som inte har möjlighet att driva en egen lösning. En möjlighet skulle vara en transkriberingstjänst som erbjuds via SUNET.

Diskussion/Frågor