

Utvikling av system for

Mottak og Bearbeiding av Digitalisert Audiovisuelt Materiale

Prosjektrapport

Utført av Arbeiderbevegelsens arkiv og bibliotek, med støtte fra Arkivverkets utviklingsmidler under programområde 1, *En innovasjonsarena for arkiv*, innsatsområde 1.1 *Arkivspira*.

Prosjektdeltakere: Jonas Nilsson og Odd-Rune Hansen

30. mars 2026

Innhold

Innledning	3
Funksjonalitet.....	4
Teknologi.....	5
Brukergrensesnitt.....	7
Transkribering og kvalitetssikring.....	9
Utfordringer og erfaringer	15
Beskrivelse av plattform	16
Læring.....	17
Oppsummering.....	19

Innledning

Dette prosjektet har sitt utspring i et identifisert behov for en løsning som kan håndtere audiovisuelt materiale i digital form, aktualisert i forbindelse med mottak av en stor mengde materiale mediekonvertert av Nasjonalbibliotekets Senter for Kulturarvdigitalisering.

Etter hvert som vi arbeidet med problemstillingen ble det klart at det var flere behov som naturlig kunne dekkes av en slik løsning, og at de problemstillingene vi sto overfor sannsynligvis var felles for flere institusjoner. Vi besluttet å søke Arkivverket om midler for utforsking av muligheter og utvikling av et sett med funksjoner og en plattform for å dekke behovene rundt håndtering av slikt materiale, og som kunne implementeres hos andre på tilsvarende måte som vi hadde planlagt hos oss.

Utgangsrammene var disse:

- En løsning for å håndtere slikt materiale må være fullstendig in-house
- Det digitaliserte materialet må kunne oppbevares samlet ett sted
- Det må være mulighet for oppslag i materialet
- Det må være mulighet for å kunne søke i materialet med fritekst, slik at spesifikke opptak enkelt kan gjenfinnes
- Løsningen må kunne håndtere en datamengde som dekker behovet vårt, anslagsvis minst 10.000 digitaliserte audio- og videoopptak.

Nokså tidlig under forberedelsene til og planleggingen av prosjektet ble det klart at den største hindringen for å kunne søke i og gjenfinne digitalt audiovisuelt materiale var manglende metadata. Dette er en egenskap som er «arvet» fra det analoge materialet; ofte finnes det en kort overordnet tekstbeskrivelse av innholdet på et opptak, men det kan være begrenset til f.eks. «LO-kongressen 2009», og i noen tilfeller var beskrivelsen så lite informativ som «Ukjent opptak».

Fordelen med slikt materiale i digitalisert form er at det finnes verktøy som kan maskinelt analysere og ekstrahere informasjon, og gjøre denne søkbar. Vi ønsket å ta i bruk en spesifikk funksjonalitet, automatisk transkribering, for å trekke ut søkbar tekst fra opptakene, og på den måte øke metadatamengden for hvert opptak betraktelig.

Funksjonalitet

Utgangspunktet for løsningen ble nedfelt i en funksjonsspesifikasjon som skulle beskrive funksjonaliteten vi ønsket å bygge. (Se vedlegg 1.)

Denne ble utarbeidet i samarbeid med arkivarer som arbeider med audiovisuelt materiale, og kan sammenfattes slik:

Brukere (arkivarer) skal kunne få tilgang til løsningen via nettleser, og brukergrensesnittet skal være enkelt og intuitivt. Materialet skal presenteres som en katalog, der brukere skal kunne bla i opptak, søke i forhåndsregistrerte metadata, og søke i transkribert tekst. Opptak som er lagret i løsningen skal kunne spilles av direkte i nettleseren, fortrinnsvis i en visning som lar bruker se transkribert tekst samtidig. Brukere må kunne legge inn merknader og eventuelt registrere ytterligere metadata direkte i løsningen, og må kunne sende opptak til transkribering direkte fra brukergrensesnittet. Det må også være mulig å laste ned opptak til lokal lagring for videre bearbeiding og tilgjengeliggjøring.

Teknologi

En forutsetning ved utarbeidelse av IT-løsninger for behandling av digitalt og digitalisert arkivmateriale er at all lagring og prosessering skal skje internt på egen lokal infrastruktur. Dette gjelder spesielt dersom KI-baserte tjenester tas i bruk. Dette for å kunne ha kontroll med hvordan data flyter i systemet og kunne begrense mulighetene for datalekkasje.

Opprinnelig skisserte vi en firedelt løsning, der forskjellige oppgaver var distribuert mellom forskjellige maskiner, og utveksling av informasjon og data skjedde via en meldingskøtjeneste (MQ). Underveis i prosjektet skjønte vi at dette tilførte en unødvendig og uønsket grad av kompleksitet, så dette ble forkastet. I stedet gikk vi for et enklere oppsett, der alle funksjoner unntatt KI-behandling er samlet på en server («Portalen»), og alle KI-oppgaver avlastes til en egen maskin bestykket med GPU for bedre KI-ytelse («KI-server»).

Vi valgte å ta utgangspunkt i en tradisjonell LAMP-stack, da det har lav kostnad og er en kjent plattform for oss. MySQL ble brukt som databasemotor, og databasene opprettet i denne inneholder både lagring av systemspesifikke data som brukerkontoer og brukerprofildata, katalog over materialet som finnes i løsningen, alle metadata som er knyttet til arkivmaterialet, og søkbare indekser over maskinelt utøkte metadata.

I utgangspunktet ønsket vi også å benytte Linux som plattform for KI-tjenester, men i første utgave av plattformen ble det besluttet å kjøre Windows på KI-serveren da det muliggjorde at flere ansatte internt kunne benytte denne maskinen for testing og utprøving av teknologien. Fordelen var at det ga oss en plattform med grafisk grensesnitt for installasjon og oppsett av programvare, ulempen at det ga en litt mer komplisert interaksjon mellom elementene i løsningen.

For fulltekstindeksering valgte vi å benytte MySQLs egne funksjoner. Vi definerte et felt «Longtext» som kan inneholde opptil 4 GB med data, med et index av typen «Fulltext», og la hele den transkriberte teksten inn i dette feltet. Det gir fleksibilitet i utforming av søkestrenger, men søketiden kan variere ganske mye, og øker ved komplekse søk eller ved søk som gir mange treff.

For transkribering valgte vi å benytte programvaren Whisper fra OpenAI, med språkmodeller opptrent og gjort tilgjengelig av Nasjonalbibliotekets AI-lab. Vi gjorde en del manuell testing av de respektive modellene og et utvalg av parametere for å forsøke å finne et oppsett som ga best mulig resultater med lavest mulig ressursbruk, og samtidig minsket mengden avvik.

Løsningen er satt opp slik at flere brukere kan benytte den samtidig, men innehar ingen spesielle samarbeidsfunksjoner. Transkriberingsoppgaver legges i en kø, og KI-serveren kjører en task som plukker oppgaver fra denne køen fortløpende så lenge køen inneholder elementer.

Ferdigstilte transkriberinger returneres til portalen i form av tekstfiler som blir plukket opp og importert inn i databasen.

Brukergrensesnitt

Vi utviklet et svært enkelt brukergrensesnitt som et utgangspunkt for løsningen, og for å få gitt arkivarene tilgang til tiltrengt funksjonalitet så raskt som mulig.

Dette grensesnittet benytter tradisjonell HTML for utforming, med en tabellvisning av registrert materiale og metadata som hovedvisning.

Utsnitt av dette grensesnittet ser slik ut:

Søkefilter: Submit Query							
ID	mappe	filnavn	Metadata				
10358	ABM000030	AV1000014822_01_0001.mp4	Arkiv-ID ARK-2670/U/ Ub/L0001/46	Date 07/05/1982	Medietype Kassett	NB-ID AV1000014822	Sekvens 46
10359	ABM000030	AV1000014823_01_0001.mp4	Arkiv-ID ARK-2670/U/ Ub/L0001/48	Date 21/04/1982	Medietype Kassett	NB-ID AV1000014823	Sekvens 48
10360	ABM000030	AV1000014824_01_0001.mp4	Arkiv-ID ARK-2670/U/ Ub/L0001/47	Date 17/03/1982	Innhold Kjelleberg, Bergljot	Medietype Kassett	NB-ID AV1000014824
10361	ABM000030	AV1000014825_01_0001.mp4	Arkiv-ID ARK-2670/U/	Date	Medietype	NB-ID	Sekvens

Målet var å lage et mer moderne grensesnitt ved bruk av HTML5 og CSS, og noe arbeid er gjort på det området.

Eksempel på forbedret grensesnitt:

Oppslag

Import

Eksport

Søk

arkivid	innhold	analog_enhet	dato	file_id
ARK-1640/U/L0002/1	Landsmøtet 1953, 1.dag, bånd nr. 1	Spolebånd	1953-06-04	AV10000044974
ARK-1640/U/L0002/2	Landsmøtet 1953, 1. dag, bånd nr. 2	Spolebånd	1953-06-04	AV10000044975
ARK-1640/U/L0002/3	Landsmøtet 1953, 1. dag, bånd nr. 3	Spolebånd	1953-06-04	AV10000044976
ARK-1640/U/L0002/4	Landsmøtet 1953, 1. dag, bånd nr. 4	Spolebånd	1953-06-04	AV10000044977

Brukergrensesnittet er todelt, med en knappebasert meny på venstre side med hovedfunksjonene i løsningen, og et visningsvindu til høyre. Hovedvisningen er en lineær oversikt over alt innholdet, det vil si at det ikke tilbyr noen form for hierarkisk visning, men bare speiler tabellen over mottatte filer. Tilgjengelige metadatafelt er Arkiv-ID (Asta-sti), Dato, Medietype, ID og beskrivelse. Bruker gis mulighet til å filtrere visningen basert på innholdet i metadatafeltene, f.eks. basert på årstall, arkivnummer, medietype, eller beskrivelse. Filtringen fungerer godt, men mangelen på mulighet for å kunne se det i et hierarki med mulighet for å kunne drille seg inn i materialet er en mangel.

Filnavnet til det enkelte opptaket er presentert som en hyperlenke, der bruker kan få avspilt opptaket ved å klikke på lenken, eller lastet ned filen ved å høyreklikke på lenken

og velge «lagre». Lyddopptak avspilles i et lite mediespillerobjekt nederst på siden, mens videoopptak åpnes i et nytt skilleark.

Ved fritekstsøk i transkripsjoner, vil man få opp et bilde over søketreff som f.eks. ser slik ut:

ID	Filnavn	Transkriberingsutdrag
11604	AV1000017878 01 0002.mp3	<u>...ettogjeld i forhold til et brutto nasjonalprodukt</u> <u>7</u> 00:00:48,480 --> 00:00:52,760 <u>gir en svak indikator på den gjeldsbyrden vi har.</u> <u>8</u> 00:00:52,760 --> 00:00:57,440 <u>En bedre indikator er n...</u>
11518	AV1000017820 01 0001.mp3	<u>...an komme til å skje innen borgerligheten.</u> <u>193</u> 00:19:19,000 --> 00:19:25,200 <u>Vi fikk en liten indikator på det under utenriksdebatten i Stortinget i går.</u> <u>194</u> 00:19:25,200 --> 00:19:30,000 ...
11657	AV1000018016 01 0001.mp3	<u>...mmer ut med ulemper på alle levekårsfag.</u> <u>59</u> 00:05:59,260 --> 00:06:06,620 <u>På alle levekårsindikatorer og bydeler hvor man kommer ut med topp score på alt.</u> <u>60</u> 00:06:06,620 --> 00:06:12,4...
		<u>...ingens muligheter</u>

Dette bildet har samme funksjonalitet som hovedvisningen, der bruker kan klikke på lenken til filnavnet for å få avspilt filen, og klikke på lenken i utdraget av transkripsjonen for å få opp hele transkripsjonsteksten.

Transkribering og kvalitetssikring

Som motor for maskinell transkribering (ASR) valgte vi å benytte Whisper fra OpenAI, i form av programmet whisper.cpp kompilert for Windows, og språkmodellene utarbeidet av Nasjonalbibliotekets AI-lab og delt på Huggingface.

(<https://huggingface.co/NbAiLab>)

Testing indikerte at den foretrukne modellen ville være nb-whisper-large på grunn av høyest nøyaktighet, men at modellen nb-whisper-medium hadde nokså tilsvarende presisjon, og i noen tilfeller noe mindre hallusinerer. Ytelsesforskjellen mellom de to modellene var i vårt tilfelle neglisjerbar, da det ikke ble stilt noen tidsmessige prestasjonskrav til transkriberingsfunksjonen.

Når det gjelder transkriberingshastighet var resultatene overbevisende. Med en GPU med i overkant av 4000 CUDA-kjerner og 16 GB VRAM oppnådde vi en hastighet på ca. 1:12, der en time transkribering tok ca. fem minutter, inkludert preprosessering og overhead.

Når det gjelder kvalitet på transkripsjonene oppnådde vi noe varierende resultater: Normalt kan man, dersom man har en manuell transkripsjon å måle mot (såkalt «Ground Truth»), bruke verdier som WER (Word Error Rate) og CER (Character Error Rate) for å få en indikasjon på om den automatiske transkripsjonen er av tilstrekkelig høy kvalitet. Vi hadde imidlertid ikke noen manuelle transkripsjoner for våre opptak som vi kunne måle resultatene mot, og tillegg var det stor spennvidde i alder, kvalitet på fysisk medium, kvalitet på utstyr som var brukt ved opptak, og kvalitet på opptak/mengde bakgrunnsstøy, så for en så heterogen datamengde som dette ville vi uansett ikke kunnet oppnå noen felles indikasjon av kvalitet på transkriberingene.

I prosjektet hadde vi imidlertid noen forhold som gjorde dette av mindre betydning:

- Vi vet at Whisper-løsningen yter svært bra innenfor ASR av vanlig tale, og at Nasjonalbibliotekets språkmodeller har vært trent på større mengder materiale i forhold til norsk språk og norske dialekter enn de internasjonalt produserte norskspråklige modellene. Vi aksepterte at denne kombinasjonen sannsynligvis ville være noe av det beste vi kunne finne i forhold til ytelse og presisjon innenfor rammene av dette prosjektet, så vi unnlot å gjøre noen kvalifisering av alternativer.
- Løsningen skulle i utgangspunktet ikke presentere transkripsjonene som et ferdig «produkt» som skulle tilgjengeliggjøres sammen med de tilhørende opptakene; de skulle kun inngå som en del av en søkbar indeks over innhold i materialet for å forbedre gjenfinnbarheten.

Ut fra dette kunne vi ta det standpunktet at enhver utøkning av metadata i form av transkribert tekst ville være en tilførelse av metadataverdi til materialet, men

selvfølgelig, jo bedre kvalitet og presisjon, desto større grad av merverdi ville materialet få.

Vi ønsket likevel å få en eller annen slags automatisk vurdering av kvaliteten til transkripsjonen til det enkelte opptaket, slik at vi ville kunne manuelt gå inn og vurdere nødvendigheten av manuell korrigering eller eventuelt fullstendig manuell transkripsjon på enkeltfiler der resultatet var svært dårlig.

For å få et grunnlag for dette gjennomgikk vi en del transkripsjoner og så på hvilke problemstillinger som dukket opp. Eksempler på problemer var:

- Gjentakelse av ord og setninger/hallusinerer
- Feiltolking av ord/navn/stedsnavn
- Ufullstendige setninger/avbrudd

Disse kunne ha forskjellige årsaker, noen skyldtes problemstillinger med programvaren og språkmodellen, noen var forårsaket av problemer med det enkelte opptaket, og noen skyldtes til og med dialekten eller måten personer på opptaket snakket og språkmodellenes manglende evne til å tyde den aktuelle dialekten.

Vi fant ut at vi kunne lage en automatisk analyse av noen av disse problemene som manifesterte seg i transkripsjonene, og benytte disse som en kvalitetsindikator (eller problemindikator):

- Setningsgjentakelse: Dersom en setning/linje forekommer mange ganger i den transkriberte teksten indikerer det at transkripsjonen ikke er presis eller at KI-prosessen har «hallusinert», så ved å lage en tabell over unike setninger og antall forekomster av disse kan man grovt beregne forekomsten av dette.
- Ordgjentakelse: I noen tilfeller forekommer hallusinerer på enkeltord, så ved å lage en tabell over unike ord med antall forekomster i transkripsjonen, og se på antall distinkte ord i forhold til totalt antall ord får vi også en indikator på kvaliteten, om enn noe svakere enn i forhold til setningsrepetisjon.
- ...-forekomster: Dersom transkripsjonen «gir opp» underveis i en setning blir dette ofte markert med «...». Ved å telle antall forekomster av disse får vi også en indikasjon på kvaliteten, om enn enda noe svakere enn for ordgjentakelse.

Samlet sett bruker vi disse tre kalkulerte parameterne, via en trinnvis skala, til å gi en indikasjon på den enkelte transkripsjonens kvalitet. Dette er ikke en veldig nøyaktig kvalitetsindikator, men på samme måte som at transkripsjonen ikke trenger å være 100% perfekt for vårt formål trenger heller ikke kvalitetsindikatoren å være 100% perfekt, den trenger bare å gi oss en pekepinn i forhold til hvilke opptak vi eventuelt bør etterbehandle manuelt.

Dette kan f.eks. se slik ut i portalen:

k- et	trans-L	trans-O	trans-D
	100	38.8102	1
	99.5816	21.3109	119
	99.7647	20.4829	87
	99.8214	18.1699	87
	100	22.6499	77
	100	21.0208	89
	99.8466	20.608	86
	99.8043	19.3777	72
	99.8049	17.4161	72

Ved hjelp av fargekoding kan man raskt og enkelt få en formening om den vurderte kvaliteten på den enkelte transkripsjonen.

Eksempel på en transkripsjon av bra kvalitet:

Indikasjoner i portal:

100	37.5459	1
-----	---------	---

1

00:00:00,320 --> 00:00:02,880

Krav til innsats for felles framtid.

2

00:00:02,880 --> 00:00:07,440

Arbeiderpartiet vil ha et samfunn preget av frihet og mangfold.

3

00:00:07,440 --> 00:00:11,800

Sosial og økonomisk trygghet for alle er grunnleggende.

4

00:00:11,800 --> 00:00:15,840

Den viktigste forutsetningen for dette er arbeid til alle.

5

00:00:15,840 --> 00:00:19,400

Felles løsninger bygd på solidaritet mellom menneskene

6

00:00:19,400 --> 00:00:23,160

er viktige redskaper for å sikre likhet og rettferdighet.

7

00:00:23,160 --> 00:00:27,280

Arbeiderpartiet vil fornye og styrke det norske velferdssamfunnet.

8

00:00:27,280 --> 00:00:28,920

Målet er at flest mulig gjennom egen innsats ...

9

00:00:29,380 --> 00:00:34,380

Gjennom egen innsats og felles rettigheter blir de i stand til å ta vare på seg sjøl og andre.

10

00:00:34,380 --> 00:00:38,220

For å få det til må vi ha en sterk og smidig offentlig sektor

11

00:00:38,220 --> 00:00:42,740

og ordninger som stimulerer den enkelte til å ta ansvar for seg sjøl og for andre.

Eksempel på problematisk transkripsjon:

Verdier i portal:

57.8801	23.1769	21
---------	---------	----

Utdrag fra transkripsjon:

5
00:00:16,780 --> 00:00:20,740
En stor forskjell er i ferd med å bli forvirret.

6
00:00:20,740 --> 00:00:25,780
En stor forskjell er i ferd med å bli forvirret.

7
00:00:25,780 --> 00:00:33,140
En stor forskjell er i ferd med å bli forvirret.

8
00:00:33,140 --> 00:00:37,060
En stor forskjell er i ferd med å bli forvirret.

9
00:00:37,060 --> 00:00:41,860
Det er sjokkerende å være der.

10
00:00:41,860 --> 00:00:46,440
Det er bra.

11
00:00:46,440 --> 00:00:49,820
Det er bra.

12
00:00:49,820 --> 00:00:52,080
Ja!

13
00:00:52,080 --> 00:00:54,700
Det er bra.

Et litt overraskende eksempel der disse mekanismene har hatt god nytte er et opptak der de to lydkanalene var i motfase. Ved avspilling hørtes dette helt normalt ut. Preprosessering for transkribering innebærer imidlertid å mikse lydspor ned til en kanal (mono), og i dette tilfellet kansellerte kanalene hverandre og vi sto igjen med et opptak som i praksis var nesten helt uten lyd. Kvalitetsverdien vi fikk ut fra dette opptaket var 7%, en klar indikator på at automatisk transkribering hadde feilet, og opptaket måtte behandles manuelt. Etter å ha justert parameterne for konvertering på en slik måte at det ble utført en automatisk fasevending av en av kanalene ved behov, endte dette opptaket opp med en normal transkripsjon og kvalitetsvurdering.

Dersom parameterne viser at det er lite setningsrepetisjon og bra ordspreddning, men mange tilfeller av «...», så kan det indikere at personen(e) som snakker på opptaket har et veldig nølende språk med mange pauser og avbrutte setninger, som for så vidt også kan være nyttig informasjon.

Et eksempel på dette:

```
30
00:02:39,020 --> 00:02:44,560
Men far dro ut til skogen og kjørte og havnet sånn ... hva heter det?

31
00:02:44,560 --> 00:02:49,680
Som du bada i, Brisk. Så det var ikke noe feil.

32
00:02:49,740 --> 00:02:54,060
Så det var ikke noe fett, akkurat. Men vi hadde det ...

33
00:02:54,060 --> 00:02:59,260
Søsknene dine, er dere omtrent sånn tett etter hverandre?

34
00:02:59,260 --> 00:03:07,360
Ja. Det er bare halvannet år til hver dag som det var. Så det var ...

35
00:03:07,360 --> 00:03:11,980
Så ... Hva slags bilde prøver jeg å danne meg av?
```

Her indikerer det ikke problemer med transkripsjonen, men at personen som er intervjuet tar pauser i setningene.

Ved å vekte disse måleparameterne mot hverandre og sette dem i en trinnvis skala kan man beregne en overordnet «kvalitetsindikator» for det enkelte opptaket.

Utfordringer og erfaringer

Konseptet såkornprosjekt var nytt, så vi ser i etterkant at vi i større grad skulle fokusert på selve utviklingen og testingen av mulighetsrommet i ny teknologi heller enn å prøve å få definert en komplett løsning som skulle kunne formidles som en pakke videre til andre. Arbeidsmengden ved å utvikle en komplett løsning som inkluderte KI-funksjonalitet var større enn det vi forventet, og vi oppnådde derfor ikke alle målene vi hadde satt oss. Totalt sett er vi likevel svært fornøyd med prosjektet, og tar med oss erfaringene vi har høstet og erkjennelsene vi har kommet til videre.

Vi har fått innsikt i og erfaring med oppsett av lokale løsninger for bruk av KI, fått et forhold til muligheter som ligger i bruk av teknologien, og lært en del om hvordan mer teknologitunge utviklingsprosjekter bør legges opp. Vi har også fått bekreftet verdien av å ta i bruk løsninger basert på KI for maskinell utøkning av metadata i digitalt arkivmateriale, både i form av en enorm ressurs- og arbeidsbesparelse, men også i form av enormt forbedrede gjenfinningsmuligheter.

En utfordring ved bruk av KI-baserte løsninger er at man til en viss grad mister kontrollen over metadataene som blir trukket ut, og dermed må man forvente og akseptere en viss grad av unøyaktighet. I noen tilfeller kan man leve med dette, mens i andre tilfeller vil maskinelt utøkte metadata med lavere presisjon ikke dekke behovet eller svare til forventningene fra brukerne. Det er viktig å kommunisere svært tydelig til brukere styrken og svakheten til disse mekanismene, forskjellene mellom KI-basert analyse og manuelt påførte metadata, og gi opplæring i hvordan man skal forholde seg til disse metodene og resultatene fra dem.

I forhold til materiellmengdene som i noen tilfeller skal prosesseres vil det ikke være menneskelig overkommelig med en manuell kvalitetsvurdering av maskinelt utøkte metadata. Som tidligere nevnt bruker vårt oppsett ca. fem minutter på å transkribere en time lydopptak. Fra en annen aktør på ABM-feltet fikk vi informasjon om at de beregner femten timer arbeid pr. time lydopptak ved manuell transkribering. Det er et forholdstall på 180:1.

For å ta et eksempel basert på en grov avrunding av vårt materiale, så vil 5000 opptak á en time ta ca. 417 timer å prosessere maskinelt, eller 17,5 døgn med uavbrutt prosessering døgnet rundt. Manuell behandling av det samme vil ta ca. 75000 arbeidstimer, eller ca. 9375 arbeidsdager á åtte timer – eller 46 år á 200 arbeidsdager.

Basert på slike tall blir det klart at man *må* ha måter å få foretatt automatisk vurdering av kvaliteten på de maskinelt genererte metadataene på, og må konstruere systemene som bruker disse metadataene på en slik måte at bruken tolererer svakheter i presisjon og nøyaktighet. For oss, i dette tilfellet, betyr det å bruke maskinelle transkripsjoner til søk og gjenfinning, og så sende selekterte opptak til manuell verifisering og komplettering før tilgjengeliggjøring.

Beskrivelse av plattform

Løsningen består av to (tre) komponenter/servere:

1. Portal – en enkel webserver/databaseserver som presenterer brukergrensesnitt til arkivbrukerne, og lagrer alle data assosiert med løsningen
2. KI-server – maskin med høyere prosesseringskapasitet, og dedikert GPU for å kunne kjøre GPU-assistert KI-prosessering med små modeller
3. Dedikert lagringsområde på arkivserver for oppbevaring av mediefiler som er tilgjengelige i portalen. (Dette må ikke nødvendigvis være på en egen server, men det var hensiktsmessig for oss å sette det opp slik.)

Portalen kjører Debian Linux 12, med Apache webserver, MariaDB databasemotor, og PHP scriptspråk. Alle versjoner av installert programvare er hentet fra default repository til Debian.

KI-server kjører Windows 11 Professional, og whisper.cpp kompilert vha. MS Visual C, med støtte for Nvidia CUDA. (<https://github.com/ggml-org/whisper.cpp>)

Arkivserver kjører Windows Server 2022, og har delt ut et eget område via standard SMB-deling i Windows.

Samhandling mellom komponentene foregår foreløpig gjennom at portalen har tilgang til delte mapper på KI-serveren, plasserer kildefiler på et inputområde for prosessering, og henter resultatfiler fra et outputområde for importering i databasen og indeksering. Portalen har også tilgang til arkivserverens delte område for å hente filer for transkribering og utlevering.

Innlesing av nytt materiale er foreløpig en manuell rutine som kjøres på forespørsel fra kommandolinje på portalen, og der manuelt registrerte metadata tilhørende filene importeres inn fra CSV-filer som en separat operasjon.

Læring

Prosjektet har gitt oss verdifull kunnskap og innsikt på tre hovedområder:

1. Organisering og gjennomføring av prosjekt

Det å skulle omsette en idé til en fungerende kjørende løsning er temmelig omfattende. Vi burde gjort en innledende avgrensning av prosjektet, og lagt opp til en lavere målsetning som skulle utforske teknologien og måter å kvalitetssikre resultatene på, heller enn å sette oss som mål at prosjektet skulle resultere i en ferdig kjørende plattform med all funksjonalitet implementert. Vi burde også i større grad tatt høyde for forsinkelser og hindringer underveis, som har påvirket fremgangen til prosjektet. Et annet viktig aspekt er å ha en avklaring med brukerne på forhånd i forhold til hvilke forventninger de har til teknologiens kapasitet og muligheter. I en løsning som dette, der det er en konkret funksjon som skal utføres, og der brukerne i liten grad har påvirkning på hvordan KI-funksjonaliteten utfører denne funksjonen, er det ikke så vesentlig å avklare brukernes holdninger og forventninger til KI i seg selv, men dersom vi skulle utvidet funksjonaliteten og gitt brukerne mer direkte tilgang til å ha en interaktiv samhandling med KI-tjenesten ville dette vært svært viktig.

2. Teknologi og KI

Teknologien som er utviklet for å ta i bruk KI-baserte løsninger er imponerende, og terskelen for å komme i gang med det er forholdsvis lav, dersom man har en IT-plattform å bygge på. Det kan imidlertid være vanskelig å navigere blant alle mulighetene og alternativene som finnes, og feil valg kan ha stor betydning for både progresjon underveis, og for sluttresultatet i et prosjekt. Vi har også sett at løsninger «ut av boksen» kan ha behov for testing og tilpasninger for å yte optimalt, og at et heterogent datagrunnlag kan føre til betydelige avvik i kvaliteten på resultatet. I forhold til sekvensiell transkripsjon av opptak er ikke ytelseskravene til maskinvareløsning så store, men dersom man skal ha en høytytende plattform med kapasitet for parallellisering av oppgaver vil behovet for kraftigere maskinvare, og kostnaden ved investering i denne, øke nokså raskt. Det samme gjelder dersom man ønsker å ta i bruk andre typer KI-tjenester som benytter andre/større modeller.

3. Utforming av brukervennlige løsninger

Det kan være enkelt, fra et teknologisk ståsted, å skissere en kongruent løsning som skal dekke et behov. Det viser seg imidlertid ofte i praksis at dersom man ikke sørger for tilstrekkelig brukerinput i designfasen av en løsning og underveis mens funksjonaliteten utformes, risikerer man å bomme på det reelle behovet brukerne har. En vanlig problemstilling er at man legger opp til arbeidsflyter som ikke er optimale for brukerne. En annen er at man utelater funksjonalitet som oppleves som grunnleggende nødvendig.

Portalen som er skissert i dette prosjektet har ikke et mangfold av forskjellige funksjoner eller oppgaver som skal dekkes, og således er det noe begrenset risiko for skivebom på dette området, men det er en utfordring å skape gode visninger inn i materialet for arkivarene som skal benytte det. Det er enkelt å tenke at fritekstsøk dekker alle gjenfinningsbehov, men det viser seg ofte at muligheter for å sortere og strukturere og gjøre utvalg basert på allerede registrerte metadata er vel så viktig.

Oppsummering

Prosjektet har resultert i en løsning som dekker de grunnleggende behovene for administrasjon av digitalt audiovisuelt arkivmateriale, inkludert maskinell transkribering og funksjonalitet for fritekstsøk i transkribert tekst, for våre arkivarer. De viktigste resultatene fra prosjektet har vært imidlertid vært økt kompetanse i å bygge løsninger som gir brukere tilgang til avansert teknologibasert funksjonalitet for å analysere og trekke ut metadata fra arkivmateriale, og kunnskap innenfor bruk og begrensninger i KI-baserte løsninger og denne teknologien.

Læringen og forskningen vi har fått gjennomført i dette prosjektet vil være verdifulle bidrag i kommende prosjekter og aktiviteter, og vi vil være tilgjengelige for å dele vår kompetanse og våre erfaringer med andre instanser og institusjoner som også ønsker å ta i bruk ny teknologi for håndtering og analyse av digitalt arkivmateriale.

Vedlegg 1: Funksjonsspesifikasjon, MBDAM

Overordnet beskrivelse av systemet

Prosjektet skal utarbeide en løsning som lar arkivarer utføre administrasjon og gjenfinning av digitalisert audiovisuelt materiale og tilhørende metadata. Systemet skal tilby et intuitivt nettleserbasert brukergrensesnitt. Det skal også bygges grunnleggende funksjonalitet for kobling mot KI-baserte tjenester for maskinell utøkning av metadata.

Systemet bygges som et sett med tjenester eller roller som kan installeres på en eller flere fysiske eller virtuelle datamaskiner, med en SQL-database for systemdata og metadata, og et fillager for oppbevaring av mediefiler.

Systemet definerer et overordnet metadataskjema som lagres i den bakenforliggende SQL-databasen.

Tilgang og tilgangsstyring

Systemet vil i utgangspunktet bli bygd kun for intern bruk på et sikret nett i en organisasjon, og vil ikke implementere sikkerhetstiltak tilpasset tilgang fra internett. Brukerkontoer inkl. passordhash, brukerprofiler og roller er definert i SQL-databasen. Nye brukerkontoer må opprettes av administrator.

Funksjonalitet:

- Brukere må autentisere seg med brukernavn/passord
- Tilganger styres via forhåndsdefinerte roller knyttet til brukerkonto
- Det er i utgangspunktet definert tre roller: Administrator, Arkivar (lese- og skrivetilgang), og Leser (kun søke- og lese-/lytte-/visningstilgang)
- Passord vil ha begrenset gyldighetsperiode, og brukere skal kunne endre eget passord.
- Ved nært forestående utløpsdato skal bruker få varsel om dette i statuslinjen i brukergrensesnittet
- Det knyttes en brukerprofil til hver bruker. Denne kan lagre innstillinger som språkvalg, fargevalg, søkehistorikk, oppsett av visninger inkludert sorteringsvalg, og annet (fullstendige egenskaper er ikke foreløpig definert). Innstillinger lagret i profilen følger pålogging og ikke fysisk klientmaskin.

Basisfunksjonalitet

Systemet skal presentere importert arkivmateriale i en tabellvisning. Visningen skal kunne være sorterbar ved å klikke på kolonneoverskrift.

Brukergrensesnittet er organisert i en klassisk firedeling, med et tittel- og menyfelt øverst, undermeny/kontekstbasert meny i midten til venstre, arbeidsområde i midten til høyre, og et informasjons- og statusfelt nederst. Oppsettet vil kunne tilpasses etter hvilken funksjon som er aktiv.

Funksjonalitet:

- Systemet skal vise et oversiktsbilde i tabellformat over opptakenes metadata, og visningen skal kunne sorteres og filtreres
- Det skal lenkes til det aktuelle opptaket fra hver linje i oversiktsbildet
- Registrerte opptak skal kunne spilles av direkte i brukergrensesnittet
- Der et opptak har en transkripsjon, skal avspilling av opptaket og lesing av transkripsjon kunne skje samtidig i brukergrensesnittet
- Brukerne vil gis mulighet til redigering av noen metadatafelter for det enkelte opptaket
- Redigering av metadata skal/kan skje i et eget skjema
- Det skal tilbys mulighet for fritekstsøk i metadata, med standard funksjonalitet som inkludering/ekskludering av søkeord, jokertegn, og samling av flere ord til søketermer
- Audiovisuelt materiale og tilhørende metadata i CSV-format skal kunne importeres av bruker i brukergrensesnittet
- Materiale og metadata (inkludert transkripsjoner) skal kunne eksporteres for tilgjengeliggjøring
- Bildefiler som viser avfotograferte etiketter eller covere fra fysiske medier tilhørende opptakene skal kunne importeres sammen med mediefilene og vises i brukergrensesnittet
- Hovedmenyen gir tilgang til hovedfunksjonene i løsningen. Ved valg i hovedmeny vises tilhørende undermeny.
- Tilgjengelige valg i hoved- og undermenyer styres av rollen pålogget bruker er tildelt.
- I de tilfeller der valgt hoved- og underfunksjon kan dra nytte av en liste/egen undermeny med innholdsspesifikke valg, vises dette som et eget felt til venstre, nedenfor undermenyen. (Dette kan f.eks. være en liste over arkivnumre, temaer, årstall eller annet.)
- Metadata-skjemaet og brukergrensesnittet skal gi mulighet for at brukerne kan registrere en kvalitetsvurdering av hhv. lyd-/bildekvalitet på det enkelte opptaket, og også en kvalitetsvurdering av maskinpåførte metadata
- Ved eksport av materiale skal også logg for det aktuelle materialet kunne eksporteres, fortrinnsvis som METS-fil i xml-format (dataene for å generere dette hentes fra logg som kontinuerlig føres i SQL-databasen.)

AI-basert utøkning av metadata

Systemet skal ha funksjonalitet for å kunne kommunisere med en tjener som tilbyr AI-funksjonalitet for maskinell utøkning av metadata på lyd- og videofiler. AI-tjeneren skal ha et programgrensesnitt som lar den motta mediedata for analyse, og returnere metadata i det formatet som systemet har forespurt.

Funksjonalitet:

- Brukere av løsningen skal kunne sende ett eller et utvalg lyd- og/eller videoopptak til definerte AI-tjenester direkte fra brukergrensesnittet.
- Første tilgjengelige funksjon er automatisk transkribering (ASR) til norsk bokmål gjennom algoritmen Whisper fra OpenAI, og språkmodeller utarbeidet av Nasjonalbiblioteket, der opptrening på norske dialekter har hatt fokus.
- Transkripsjonene returneres på SRT-format, som ren tekst med tidskoder
- Systemet skal selv gjøre en enkel vurdering av den tekniske kvaliteten på transkriberingene som er utført, og returnere parametere som indikerer kvaliteten sammen med transkripsjonen
- Brukergrensesnittet skal på en intuitiv måte synliggjøre hvilke opptak som er AI-transkribert eller har gjennomgått annen AI-basert metadatautøkning
- Brukergrensesnittet skal indikere den automatisk vurderte kvaliteten på utførte AI-operasjoner, der dette er relevant

Tilleggsfunksjoner for administratorer

- Administrasjon av brukere, roller/tilganger, og brukerpassord
- Visning av aktuelle tabeller i databasen
- Lesing av systemlogger
- Mulighet for å se status på og omstarte deler av systemet
- Funksjoner for å ta og tilbakeføre sikkerhetskopier av systemet

Fremtidig mulig funksjonalitet, ikke i prioritert rekkefølge

- Støtte for andre filformater ved import av metadata enn CSV, f.eks. Excel og JSON.
- Mulighet for masseredigering av metadata på et utvalg filer samtidig
- Mulighet for å gjøre endringer i transkribert tekst direkte i brukergrensesnittet
- Funksjonalitet for å splitte ett opptak i flere deler, der en fil inneholder flere separate opptak
- Mulighet for eksport av et utsnitt av en lyd- eller videofil, med tilhørende metadata
- Mulighet for konvertering av mediefil til annet format ved eksport

- Mulighet for individuelt tilpasset visning inn mot materialet, og mulighet for flere forskjellige visninger, lagret i brukerprofil.
- Liste over søkeord som er benyttet lagres under den enkelte brukers profil ved søk, slik at søkehistorikk kan vises nedenfor undermenyen
- Ved søk i materialet skal bruker med arkivarrolle gis adgang til å legge søkeord til som emneord i metadataene til det enkelte opptaket der søkeordet får treff. Disse emneordene vil være tilgjengelige som hurtigoppslagsord på tvers av alle brukerne.
- Registrering av epostadresse som et felt i brukerprofilen, slik at brukere selv kan løse problemer med glemt passord via endringslenke på epost, og slik at brukere kan få varsler på epost, f.eks. ved ankomst av nytt materiale, eller ved fullførte bestilte AI-oppgaver
- En fjerde rolle, «Gjest», som kan få lese- og søketilgang til et utvalg materiale. (Dette kan benyttes for å gi lesesalsgjester tilgang til utvalgte deler av materialet i løsningen for egne søk.)
- Kobling mot ekstern katalogtjeneste for brukerpålogging, f.eks. Microsofts Active Directory via LDAP.
- Støtte for nynorsk og eventuelle andre språk ved maskinell transkribering, dersom det er ønskelig og det finnes aktuelle språkmodeller.
- Transkriberinger skjer både på bokmål og nynorsk, og automatiske metoder benyttes for å vurdere hvilken av transkriberingene som gir best resultat. Dette må dog ses i forhold til ressursbruk og bærekraft.
- Funksjon for å legge «vannmerker» på opptak som eksporteres, i form av en kort sekvens som spilles av før eller etter opptakene og spesifiserer kilde og annen ønsket informasjon.
- Mulighet for taleridentifikasjon og -gjenkjenning (Speaker Diarization)
- Mulighet for AI-generering av sammendrag av transkripsjoner
- Mulighet for å analysere videoopptak med tanke på sceneanalyse; slik at man kan bruke tekstbaserte søk etter spesifikke scener/hendelser i videoopptakene
- Integrasjon mot APler, f.eks. NextCloud for deling av nedlastingspakker, Digitalarkivet for publisering, Fotoware for deling av materiale på lesesal