

Tekoäly tehostaa videolähetyksen käännöstyötä

18.10.2022

Video on tehokas ja suosittu tapa välittää tietoa tai esimerkiksi osallisuuden kokemusta pitkienkin etäisyyksien taakse. Nykyään helposti ja edullisesti saatavilla olevat tekniset ratkaisut mahdollistavat liikkuvaa ja nk. still-kuvaa sekä puhe- tai muuta ääntä yhdistävien kokonaisuuksien luomisen verrattain edullisesti, nopeasti ja pienellä vaivalla niin suoratoistoa kuin tallentamistakin silmällä pitäen.

Kansainvälistyvässä ja esteettömyyttä korostavassa toimintaympäristössä erääksi resursseja syöväksi tuotantoprosessin vaiheeksi on noussut tarve tekstittää videoilla esitettävät puheet – joko puhutulle kielelle tai käännöksen kautta muille kohdekielille. Tämä tarve koskee sekä SeAMKin omia hankkeita, että hankkeissa mukana olevien yritysten materiaaleja, muun muassa erilaisia ohjeita.

Tehodata-hankkeen eräänä toiminnallisena pilottina päätettiin selvittää, millaisia valmiiksi saatavilla olevia, tekoälyyn pohjautuvia käännös- ja tekstityspalveluita on saatavilla ja miten ne palvelevat pilottikohteen käyttötapausta, jossa

- suorana lähetyksenä välitettävä konsertti ja sen yhteydessä pidettävät puheenvuorot
- käännetään ja tekstitetään kuuntelijan valitsemalle kohdekielelle
- ja tarvittaessa välitetään internetin yli minne päin maapalloa tahansa.

Varsinaisen pilottikohteen ohella tarjontaa peilattiin alustavasti myös SeAMKin sisäisten tekstitys- ja käännöstarpeiden suuntaan.

Tavoiteltu prosessi

Pilottikohteen tarpeita parhaiten palveleva prosessi olisi:

1. Materiaalin kaappaus (kamerat, mikrofonit)
2. Automaattinen tekstitys => .srt tai muu aikaleimallinen tekstimuoto
3. Automaattinen käännös kohdekielelle
4. Lähetys internetiin ja tallennuslaitteelle
5. Asiakas vastaanottaa lähetyksen ja käännöksen tekstinä tai puheena valitsemallaan kielellä
6. Tallenne Youtubeen

Eri käännöskoneet palvelevat eri käyttötarkoituksia

Heti selvityksen alkuvaiheessa todettiin, että hankkeen puitteissa on mielekästä tarkastella ainoastaan valmiiksi saatavilla olevia käännös- ja tekstityspalveluita. Pilottikohde huomioiden painoarvoa annettiin palvelun edullisuudelle tai mieluummin maksuttomuudelle sekä sille, että palvelu on otettavissa käyttöön nopeasti ja sen käyttö on helppoa. Tarkasteltavaksi kysymykseksi muodostui se, että onko olemassa edullista tekoälypohjaista käännös- ja tekstityspalvelua, joka a) tekstittää ja b) kääntää valituille kohdekielille suomenkielistä puhetta 1) suorasta lähetyksestä 2) tallenteesta siten, että tuotos on ymmärrettävissä puhujan tarkoittamalla tavalla.

Alkuperäinen listaus tarkasteltavista palveluista luotiin Google-hakukoneen antamista tuloksista englanninkielisille hakusanoille “automatic voice translation” sekä “automatic transcription”. Tarjonnan monipuolisuuden vuoksi tarkastelusta rajattiin pois erilaiset omien tuotteiden pohjaksi tarkoitetut rajapintapalvelut ja tuloslistan häntäpäähän jääneet tuotteet siten että lopulta käyttötestiin otettiin mukaan:

- Microsoft Translate
- Microsoft Teams litterointi ja live-käännökset
- Google Translate
- Youtube litterointi
- Clevercast
- Screencast-o-Matic

Vertailua olisi ollut kiinnostavaa laajentaa myös Amazonin (Transcribe i.a), IBM:n (IBM i.a) ja esimerkiksi Mozilla-yhteisön (Mozilla i.a) tuottamat litterointi- ja käännösrajapinnat ja toisaalta pienempien toimijoiden näiden pohjalle rakentamia valmiita palveluita. Hankkeen resurssien puitteissa rajaus oli kuitenkin tehtävä melko suppeaksi.

Vertailussa havaittiin, että eri palvelut soveltuvat hieman erilaisiin tarkoituksiin eli että pelkästään litterointi tai kielikäännös eivät riittäneet kuvaamaan tarkasteltavia kyvykkyyksiä, vaan jaottelua kannatti tarkentaa eri palveluiden soveltuvuuksien esille saamiseksi. Päädyimme lopulta seuraavaan jaotteluun:

1. Suora lähetys
    1. Tekstitys suoraan lähetykseen (alkuperäiskielellä)
    2. Litterointi .srt tai muuksi tiedostoksi
    3. Käännös valitulle kohdekielelle tekstinä (asiakkaan toimesta)
    4. Puheen syntetisointi kohdekielelle
    5. Lähetyksen ohjaus esimerkiksi monikameratuotannossa
2. Tallenteesta

1. Litterointi .srt tai muuksi tiedostoksi
2. Käännös valitulle kohdekielelle tekstinä
3. Puheen syntetisointi kohdekielelle

Tarkastelu tehtiin touko-kesäkuussa 2022. Tuolloin mikään tarkastelluista palveluista ei yksinään tarjonnut parasta tai edes kelvollista ratkaisua kuhunkin osa-alueeseen. Taulukkoon 1 on koottu yhteenveto keskeisistä havainnoista tämän jaottelun pohjalta.

Taulukko 1. Keskeiset havainnot tarkasteltujen palveluiden kyvykkyyksistä (0 = ei tukea, 1 = toimii, 2 = toimii hyvin).

Palvelu/kyvykkyys	A1	A2	A3	A4	A5	B1	B2	B3
Microsoft Translate	2[1)]	0	2	2[1)]	0	0	0[4)]	0
Microsoft Teams	2	?[2)]	2	0	0	0[3)]	0	0
Google Translate	0[5),6)]	0[5),6)]	2[5),6)]	1[5)]	0	0[6)]	0[6)]	0
Youtube	0[8)]	0[8)]	0[8)]	0	0	2[7)]	2[7)]	0
Clevercast	1[9)]	0	2[9)]	0	2	0	2[9)]	0
Screencast-o-Matic	0[10)]	0	0	0	0	2	1	0

1) Vain ääniraita ja chat-keskustelu (ei videokuvaa)
2) Ei saatu testattua SeAMKin ympäristössä
3) MS Stream tekee tekstityksen, joka on aiemmin ollut integroitu osaksi Teamsin käyttöprosessia.
4) MS Bing pystyy kääntämään lyhyehköjä tekstejä
5) Google Translate tekstittää myös puhetta mutta sillä ei voi tehdä verkkolähetyksiä. 5000 merkkiä rajoitus.
6) Google Sheets mahdollistaa pidempienkin tekstien kaappaamisen puheesta, jolloin teksti on siirrettävissä Youtubeen tekstitykseksi.
7) alkuperäiskieli englanniksi
8) Youtuben suorien lähetysten toimintoa ei käyty perinpohjaisesti läpi
9) Clevercastin toimittajan mukaan tuki automaattiselle suomenkieliselle tekstitykselle saattaa parantua tulevaisuudessa
10) Screencast-o-matic pystyy tallentamaan suoraa lähetystä ja litteroida sen.

Havainnot, päätelmät ja jatkotyöskentely

Mikään kokeilluista palveluista ei tarjonnut valmista kokonaisratkaisua alkuperäiseen tarpeeseen. On kuitenkin huomattava, että palvelut kehittyvät erittäin nopeasti. Tarkastelu tehtiin touko-kesäkuussa 2022 ja havainnot vanhentuvat lyhyessä ajassa. Lisäksi on jo tiedossa, että tarkasteltujen sovellusten ja palveluiden ulkopuolella on tarjolla monia mahdollisuuksia sekä suorien lähetysten että tallenteiden tekstittämiseen ja kääntämiseen: esimerkiksi Microsoft Stream tarjoaa mahdollisuuden tallenteen automaattiseen tekstittämiseen (Microsoft Stream, i.a.).

Lähimmäksi kokonaisratkaisua pääsivät toisaalta Microsoft Translate (puutteena videokuvan lähettämisen mahdollisuus), Teams (puutteena vaatimus käyttäjätunnuksen olemassaolosta) ja toisaalta Clevercast (puutteena heikko tekstin tuottaminen ja kielen kääntäminen suomen kielestä) (Microsoft Translate, i.a.; Microsoft Teams, i.a.; Clevercast, i.a.). Clevercastin tukipalvelun antamien tietojen mukaan heidän palvelunsa hyödyntää Googlen rajapintaa.

Tarkasteluhetkellä Microsoftin palvelut tuottivat parhaan osuvuuden suomeksi tuotetun puheen tekstitykselle ja käännöksille. Testaus oli varsin pintapuolista ja joka tapauksessa tulokset riippuivat vahvasti käytetystä sanastosta. Etenkin toimialakohtaisen erityissanaston (ja lauserakenteiden) litterointi ja kääntäminen olivat haastava pala sekä Googlen että Microsoftin palveluille.

Screencast-O-Matic on testattujen palveluiden joukossa selkeästi eri tarpeeseen vastaava tuote (Screencast-O-Matic, i.a.). Sen tarkoitus on tuottaa videosta litterointi, joka voidaan tarvittaessa viedä esimerkiksi Youtubeen. Koekäytön perusteella se litteroi varsin hyvin myös suomea.

Clevercast puolestaan pyrkii tarjoamaan valmiin kokonaisratkaisun esimerkiksi monikameratuotannon lähettämiseen (broadcast), tekstittämiseen ja kääntämiseen asiakkaiden valitsemille kohdekielille (Clevercast, i.a). Periaatteessa se vastaa parhaiten pilottikohteen alkuperäistarpeisiin, mutta ongelmana on heikko suorituskyky suomen kielen kohdalla ja toisaalta palvelun varsin kallis hinta.

Pilottiselvityksen aikana kävi ilmi, että vaikka tarkastelun kohteena olleeseen tarpeeseen ei selvää kokonaisratkaisua löytynytkään, on automaattisista tekstitys- ja käännöskoneista merkittävä hyöty myös SeAMKin sisäisessä käytössä. Videomateriaalien esteettömyysvaatimuksista johtuen kaikki videot on tekstitettävä, mikä on erittäin hidasta käsityötä: karkeana yleistyksenä 10 minuutin videon tekstittäminen vaatii noin tunnin työajan ja tämän jälkeen tekstin kääntäminen on oma urakkansa. Tarjolla olevia tekstitysautomaatteja on syytä jatkossa tarkastella tästä näkökulmasta, sillä alustavien havaintojen mukaan esimerkiksi Microsoft Teamsissa ja sen yhteydessä toimivassa Microsoft Streamissa ja toisaalta Microsoft Powerpointin Live-tilassa on varsin hyvät työkalut sekä suoran lähetyksen että tallennetun videon litterointiin ja sitä kautta myös kääntämiseen valitulle kohdekielelle. Syksyn 2022 aikana näitä ominaisuuksia on tuotu Microsoftin tuotteisiin ja niiden hyödynnettävyyttä on syytä tarkastella jatkossa erikseen.

Artikkeli on valmisteltu osana Datasta ketteryyttä ja uutta liiketoimintaa Etelä-Pohjanmaan pk-yrityksiin (TehoData) -hanketta (Seinäjoen ammattikorkeakoulu (SeAMK), i.a.). Samalla haluamme kiittää hankkeen ja tämän artikkelin rahoittamisesta Keski-Suomen ELY-keskusta. Hanke rahoitetaan REACT-EU-välineen määrärahoista osana unionin covid-19-pandemian johdosta toteuttamia toimia.

Tommi Ylimäki
lehtori
SeAMK, TKI

Juha-Matti Arola
projektipäällikkö
SeAMK, TKI

Lähteet:

Amazon Transcribe. (i.a.) https://aws.amazon.com/transcribe/

Clevercast. (i.a). https://www.clevercast.com

Google Translate. (i.a). https://translate.google.com

IBM Watson. (i.a.) https://www.ibm.com/cloud/watson-speech-to-text

Microsoft Stream. (i.a.) https://stream.microsoft.com/

Microsoft Teams. (i.a). https://www.microsoft.com/fi-fi/microsoft-teams/group-chat-software

Microsoft Translate. (i.a) https://translate.it

Mozilla Web Speech. (i.a). https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API/Using_the_Web_Speech_API

Screencast-o-Matic. (i.a). https://screencast-o-matic.com/

Seinäjoen ammattikorkeakoulu (SeAMK). (i.a.). TehoData. https://projektit.app-staging.seamk.hion.dev/tehodata/

Youtube. (i.a). https://www.youtube.com