

Tillgänglig textning med hjälp av språkteknologi

Michael Stormbom

Digitaliseringen ger fantastiska möjligheter till digitala tjänster som är genuint tillgängliga för alla, även om stora utmaningar kvarstår. Tjänsteproducenter blir allt mer medvetna om vikten av tillgänglighet, och tillgänglighetsdirektivet har medfört lagstadgade krav på att digitala offentliga tjänster måste vara tillgängliga. Det är möjligt med hjälp av språkteknologier. Ett exempel är undertexter som skapas med hjälp av taligenkänning.

Den pågående digitaliseringen har gett oss nya sätt att producera, förmedla och konsumera digitalt innehåll. Den erbjuder en stor potential att förmedla information och tjänster till en bred och mångfaldig publik på ett inkluderande, tillgängligt och lättbegripligt sätt. Ändå är långt ifrån allt material online åtminstone än så länge genuint tillgängligt för alla. Det är också viktigt att komma ihåg att alla inte heller har tillgång, kunskap och möjlighet att använda sig av digitala tjänster. Det gör att digitaliseringen i själva verket också kan innebära att information blir mera otillgänglig.

Ett konkret exempel på hur information kan göras tillgänglig är textning av video. Undertexter är mycket viktiga för personer med hörselnedsättningar. Det är en stor mängd personer som berörs; enbart i Sverige uppskattar man att över 1,5 miljoner personer har olika grader av hörselnedsättning. (Hörselskadades riksförbund 2020, s. 2). Undertexter är till nytta i alla sammanhang där man inte kan tillgodogöra sig information som förmedlas i form av ljud. De kan vara till stor hjälp för personer som inte behärskar språket eller i situationer där tittaren inte kan höra ljudet.

Europeiska unionen har uppmärksammat otillgängliga digitala tjänster som ett växande problem. Under de senaste åren har EU:s webbtillgänglighetsdirektiv trätt i kraft. Det ställer krav på tillgänglighet i offentliga aktörers digitala tjänster. Direktivet omfattar många olika krav på webbsidor som exempelvis gäller hur de är strukturerade. Ett viktigt krav i direktivet gäller textning/transkribering av video och ljud online. För ljud och video som har publicerats i offentliga tjänster och som är avsedda att finnas publicerade online mer än 14 dagar, är aktörerna skyldiga att också tillhan-

dahålla en motsvarande textversion, vilken i praktiken är en transkription eller undertext. (Europaparlamentet och rådet 2016).

1. Vad är språkteknologier?

Språkteknologi inbegriper olika former av automatisk analys, behandling och produktion av språk, både för text och tal, med hjälp av programvara. Språkteknologier faller ofta under det bredare begreppet artificiell intelligens (AI). För enkelhetens skull kommer jag därför genomgående att använda språkteknologi i den här artikeln.

En specifik form av språkteknologi är så kallad taligenkänning, vars uppgift är att omvandla tal till text. Det finns en lång rad användningsområden för detta, från röststyrning till diktamen och undertextning. Utan att vara medveten om det är du säkert dagligen i kontakt med olika språkteknologier. Bland många tillämpningsområden kan nämnas:

Informationssökning och -hantering. Varje gång du gör en sökning på Google eller via något annat söksystem, finns språkteknologier med i bakgrunden

Stavnings- och grammatikkontrollerna i Microsoft Word och andra ordbehandlare är språkteknologier.

Maskinöversättning, det vill säga automatisk översättning från ett språk till ett annat.

Alla former av **dialogsystem**, som till exempel olika chattbotar.

Röststyrda digitala assistenter som Siri, Alexa eller Google Assistant, bygger på språkteknologier för omvandling av tal till text, tolkning av texten och för att forma svar.

En central aspekt för att system som de här ska fungera bra är att de har tillgång till stora mängder data. Kort uttryckt "lär sig" exempelvis taligenkänningen hur den ska omvandla tal till text genom att exempel med existerande tal och motsvarande text har matats in. På det sättet skapas en modell för vad taligenkänningssystemet ska göra när någon talar. Att samla in och skapa träningsdata är ofta en omständlig och dyr process. För att taligenkänningen ska fungera tillräckligt bra måste dessa data också representera så många olika tänkbara användningsfall som möjligt. Det behövs olika dialekter, olika rösttyper, olika talvolym och -tempo och det måste ha beaktats att individer till exempel kan tala mer eller mindre tydligt. Olika former av nedsättning i talförmåga eller förekomst av bakgrundsljud kan inverka, och de finns därför också med som några av de många aspekter som måste tas i beaktande när man har målet att skapa genuin tillgänglighet.

2. Textning med hjälp av språkteknologier

Språkteknologier kan fungera som direkta hjälpmedel för att göra digitala tjänster och digital information mera tillgängliga för användarna. Ett exempel på sådana hjälpmedel är skärmläsare som läser upp det som står på en skärm för personer med synnedsättningar, ett annat exempel är röststyrning. Språkteknologier har även tillämpningsområden när det gäller själva produktionen av tillgängligt material. Ett konkret exempel är undertexter, där användning av taligenkänning och andra språkteknologier kan göra processen mycket effektivare och snabbare.

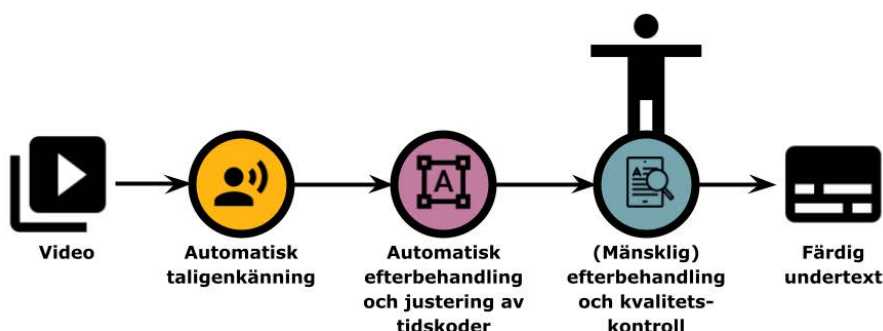
Språkteknologier kan fungera både i system för helautomatisk textning och som verktyg för mänskliga textare som använder den automatiskt producerade texten som en råversion i processen. Då skapar människan den slutgiltiga versionen och kan i arbetet fokusera på kvalitetsaspekter som automatisk teknologi inte kan hantera.

Det finns många fördelar med att använda språkteknologier i produktionsprocessen. För både beställare och producent är kostnadseffektiviteten förvisso en viktig fördel. Användningen av språkteknologier kan också innebära kortare leveranstider, och att det blir möjligt att tillgängliggöra större volymer av material under en kortare tidsperiod än vad som hade varit möjligt enbart med mänskligt arbete. (Koponen 2020, s. 58)

En risk med helautomatisk textning är att även de bästa systemen inte fungerar perfekt och att de därför kan känna igen ord fel. Ifall det inte finns en mänsklig kvalitetskontroll blir då även undertexten felaktig. I bästa fall är misstaken i undertexten harmlösa och påverkar inte förståelsen av innehållet, men i värsta fall kan det få ödesdigra konsekvenser om felaktig information förmedlas, exempelvis i en krissituation. Undertexter av dålig kvalitet kan dessutom ha en negativ inverkan på tittarens uppfattning om hur trovärdig och pålitlig informationen i videon är.

Språket förändras också konstant: nya ord och begrepp kommer till, ord får nya betydelser, olika namn är nyhetsaktuella, nya organisations-, företags- och produktnamn dyker upp. Därför är det också viktigt att kontinuerligt hålla språkteknologierna ajour. Ju fler förändringar i språket som inte beaktats, desto fler sådana förändringar kan till exempel taligenkänningen gradvis och i allt större utsträckning börja känna igen som fel.

Undertexter återger dessutom inte ordagrant det som sägs i ljudspåret på videon. För läsbarhetens skull, och för att man ska hinna läsa undertexten, är det oftast nödvändigt att förkorta och göra omskrivningar. Radbyten ska vara logiska och det får inte finnas för många ord per rad. Detta kan göras automatiskt i en viss utsträckning, men automatiken är än så länge långt ifrån perfekt.



Figur 1. En överblick över textningsprocessen.

3. Människan behövs i textningsprocessen

I många fall kan automatiskt textning vara "tillräckligt bra" och i användarundersökningar ses automatisk textning, potentiellt också av lägre kvalitet än mänsklig textning, som bättre än ingen textning alls (Koponen 2020 s. 60). Det är dock att rekommendera att alltid ha en människa med i textningsprocessen. Automatisk textning kommer – på gott och ont – att bli allt vanligare och mera accepterad av tittarna. Men även om man kan komma långt med automatik finns det inget substitut för mänsklig intuition, kreativitet och kompetens.

Att få information på sitt eget språk är också en viktig tillgänglighetsfråga. I den här artikeln har jag främst fokuserat mig på textning på samma språk som ljudet i videon. Men det finns möjligheter att utnyttja språkteknologier för att översätta existerande undertexter och för att texta direkt till ett annat språk än ljudet i videon; även om det kvarstår mycket forsknings- och utvecklingsarbete är det här ett mycket lovande område (Vázquez et al, 2020 s. 100).

Man bör också komma ihåg att om informationen i videon i sig är otillgänglig och svår att förstå, är det möjligt att också den resulterande undertexten blir otillgänglig. Det är därför viktigt att beakta tillgängligheten från första början i produktionsprocessen.

Det finns såväl utmaningar och möjligheter vad beträffar uppdraget att ta fram genuint tillgängliga digitala tjänster. Språkteknologier kan spela en stor roll i produktionen av tillgängligt material och de kommer att bli allt viktigare i framtiden. De möjliggör både att mera material görs tillgängligt snabbare och sänker tröskeln för att producera tillgängligt material.

Det är trots allt mycket viktigt att alltid sätta tittarna i centrum – det är ändå de som avgör om något är tillgängligt eller inte. Oberoende av om textningen har kommit till genom en fullständigt automatisk lösning, eller om mänskligt

arbete är involverat är informationen aldrig tillgänglig om den inte når fram till sin publik.

Summary

The on-going digitalisation of our society has radically changed our daily lives. We constantly interact with different types of digital services and solutions. While digitalisation has the potential to create a more accessible world, there are also a number of challenges and risks which may result in the opposite. Natural language processing tools for text and speech can be of great help in the production of accessible digital content, and an example of that is the use of speech recognition and other technologies when creating subtitles for online video, both for entirely automatic subtitling and for further post-editing by a human subtitler. It allows for a faster and more cost-effective subtitling process. It is however important always to remember to put the user in the centre when creating content that is truly accessible for all.

Författaren

Michael Stormbom är affärsområdeschef i det finländska språkföretaget Lingsoft Ab. Han arbetar med affärsutveckling kring lösningar baserade på språkteknologier.



Källor

- Koponen, Maarit et al, 2020: User perspectives on developing technology-assisted access services in public broadcasting. I: *User perspectives on developing technology-assisted access services in public broadcasting*. Bridge: Trends and Traditions in Translation and Interpreting Studies 1(2), 47-67. <https://www.bridge.ff.ukf.sk/index.php/bridge/article/view/8/25>
- Vázquez et al, 2020: The University of Helsinki Submission to the IWSLT2020 Offline Speech Translation Task. I: *Proceedings of the 17th International Conference on Spoken Language Translation (IWSLT)*, pages 95–102. <https://www.aclweb.org/anthology/2020.iwslt-1.10.pdf>

Europaparlamentets och rådets direktiv (EU) 2016/2102 av den 26 oktober 2016 om tillgänglighet avseende offentliga myndigheters webbplatser och mobila applikationer. Hämtat 23.11.2021. <https://eur-lex.europa.eu/legal-content/SV/TXT/?uri=CELEX:32016L2102>

Hörselskadades riksförbund, Hörselskadade i siffror 2020. Hämtat 23.11.2021. <https://hrf.se/app/uploads/sites/13/2020/12/Statisik-2020-HRF-och-SCB-Horselbron-och-forskning.pdf>

W3C, Web Content Accessibility Guidelines (WCAG) Overview. Hämtat 23.11.2021. <https://www.w3.org/WAI/standards-guidelines/wcag/>

Nyckelord

tillgänglighet, språkteknologi, taligenkänning, textning, AI