

Workshop om språkteknologi för ett flerspråkigt Europa

09.00 Registrering med kaffe och smörgås

09.15 Välkomstord: Johan Wullt, Europeiska kommissionen i Sverige

09.20 Inledning: LDS workshop i Sverige

09.30 Europeiska kommissionen: The Digital Europe Programme and the Common European Language Data Space. Philippe Gelin, Europeiska kommissionen

Philippe Gelin från EU-kommissionen presenterar European Language Data Space (LDS) och Programmet för ett digitalt Europa (DIGITAL) - ett nytt EU-finansieringsprogram som är inriktat på att föra digital teknik till företag, medborgare och offentliga förvaltningar. *Philippe Gelin är sektorchef för flerspråkighet vid DG/CONNECT på Europeiska kommissionen.*

10.00 European Language Data Space: developing a market for language data and services and benefitting from a joint European effort, Georg Rehm, DFKI and LDS Coordinator

Ett Data Space är ett ramverk som stöder datadelning. Det ger en tydlig struktur för deltagarna att dela, handla och samarbeta om datatillgångar på ett sätt som är förenligt med relevanta lagar och förordningar. European Language Data Space syftar till att bygga en pålitlig, datamarknad inom EU för utbyte språkresurser i den offentliga och i synnerhet i den privata sektorn. *Georg Rehm är forskningsledare på Deutsche Forschungszentrum für Künstliche Intelligenz GmbH (DFKI) och koordinator för European Language Data Space.*

10.30 Paus

10.50 AI och språkteknologi i språkpolitiskt perspektiv. Rickard Domeij, Isot

Språkrådet arbetar på regeringens uppdrag bland annat med att främja utvecklingen av språkteknologi för språken i Sverige, förutom för svenskan också för de nationella minoritetsspråken vilket innebär särskilda utmaningar. Nationellt görs arbetet till stora delar inom Nationella språkbanken som skapar möjligheter att forska i digitala text- och talmaterial med hjälp av verktyg och metoder i gränslandet mellan språkteknologi och AI. Internationellt bedrivs samverkan för att driva på utvecklingen av språkteknologi och AI för alla språk i Norden och Europa, inte minst inom projektet Language Data Space. Konkret handlar det främst om att gemensamt samla in, skapa och dela nödvändiga språk- och teknikresurser. I presentationen får du veta mer om det nationella och internationella arbetet med att utveckla språkteknologi för språken i Sverige och Europa. *Rickard Domeij är doktor i datorlingvistik vid Stockholms universitet och ansvarig för arbetet med språkteknologi på Språkrådet vid Institutet för Språk och folkminnen.*

11.10 GPT för svenska och projektet TrustLLM. Magnus Sahlgren och Danila Petrelli, AI Sweden

AI Sweden leder arbetet med att utveckla storskaliga generativa språkmodeller för det svenska språket. Inom detta arbete har AI Sweden än så länge tränat och släppt det som blev den första stora GPT-modellen för svenska, GPT-SW3. Just nu pågår arbetet med att utveckla nästa generations språkmodell för svenska, som

kommer att vara multimodal och alltså inte enbart kunna förstå text, utan även tal och bilder. AI Sweden är även med i ett antal europeiska initiativ som syftar till att bygga GPT-modeller för en större mängd europeiska språk. Ett av dessa initiativ är TrustLLM, som syftar till att utveckla språkmodeller för de germanska språken. Presentationen ger en överblick över arbetet med storskaliga språkmodeller ur såväl det nationella som det europeiska perspektivet. Vi diskuterar även våra datahanteringsprocesser och utmaningar med att arbeta med data för träning av storskaliga språkmodeller. *Magnus Sablgren är doktor i datorlingvistik och forskningsledare för språkteknologi på AI Sweden. Danila Petrelli har rollen som data manager för språkprojekten på AI Sweden.*

11.40 Diskussion: AI och språkteknologi för språken i Sverige. Hur gör vi?

12.10 Lunch

13.00 KB-labbs språkmodeller och tal-till-textmodeller. Leonora Vesterbacka, KB-labb

Sedan 2019 har KB-labb möjliggjort storskalig forskning på bibliotekets samlingar, men labbet är mest känt inom det svenska AI ekosystemet för sina samlingsbaserade språkmodeller som laddats ned över tre miljoner gånger. Presentationen lyfter fram två pågående projekt: KB-labbs tal-till-textmodeller och ett samarbete med Nasjonalbiblioteket i Norge och Det Kgl. Bibliotek i Danmark för att träna skandinaviska stora språkmodeller. Vidare ges exempel på hur KB-labbs språkmodeller används på svenska myndigheter idag. *Leonora Vesterbacka har en doktorexamen i fysik från ETH Zürich och jobbar på Kungliga Biblioteket som Senior Data Scientist.*

13.20 Hur använder Schibsted AI i sin verksamhet? Agnes Stenbom, Schibsted

Inom Nordens största mediegrupp Schibsted står teknikutveckling högt upp på agendan, inte minst AI-teknik som Schibsted utforskar och använder för ett ökande antal produkter och tjänster. Agnes Stenbom, chef för innovations- och inkluderingslabbet IN/LAB på Schibsted och industridoktorand inom AI och journalistik vid KTH, berättar mer om detta och om varför generativ AI - det vill säga AI-system som kan producera text, bild, ljud och rörligt innehåll - är en intressant teknik för såväl stora mediehus som små publicister. Målet är inte att ersätta reportrar med AI, i stället handlar det om att mänskliga kompetenser ska samverka med AI-tekniken för att skapa de bästa möjliga medieprodukterna för både sam- och framtidens mediekonsumenter. *Utöver sin roll på Schibsted/KTH har Agnes grundat och driver branschnätverket Nordic AI Journalism, som sedan 2020 samlar drygt 450 personer verksamma inom AI på nordiska medier.*

13.40 Språkteknologi; ett redskap i verktygslådan sedan tidigt 90-tal – vardagsmat för branschen. Anne-Marie Colliander Lind, Språkföretagen

Att ha olika typer av språkteknologi i företagets verktygslåda har varit en självklarhet i mer än 30 år för Språkföretagens medlemsföretag. Branschen har hela tiden legat i framkant när det gäller språkteknologi för att säkerställa att dess kunder har fått bästa nyttan av tillgängliga lösningar, t ex för att upprätthålla hög kvalitet, säkerställa konsekvent terminologi eller för att göra leveranser mer tidseffektiv, när verkligheten krävt det. Teknik präglar också många av de interna processerna i ett modernt språkföretag. AI-tekniken är bara ytterligare ett redskap som vi måste lära oss att bemästra. Inte minst för att säkerställa att användandet av AI kan göras på ett säkert sätt där både t ex patientsäkerhet, företagshemligheter och personlig integritet respekteras. En av de största utmaningarna med den nya tekniken är dock att hantera kundernas högt ställda förväntningar på AI. Hör hur Språkföretagens medlemmar ser på hoten och möjligheterna vid ett utökat

samarbete kring insamlandet av språkdata inom EU. *Anne-Marie Colliander Lind är ordförande i Språkföretagen och har varit aktiv inom språkbranschen, I sverige och internationellt, sedan 1990.*

14.00 Paus

14.20 Diskussion: Data för språken i Sverige. Vilka utmaningar finns?

14.50 Avslutning

15.00 Mingel med snacks

15.30 Slut