

SpråkVis - Språkteknologisk vismansrapport

av Krister Lindén, Kimmo Koskenniemi och Torbjørn Nordgård

Sammanfattning

Nordiska Ministerrådet har beställt en tioårsplan i form av en vismansrapport för att göra de nordiska länderna till en ledande region i språkteknologi. Sex nyckelområden har identifierats: Policy, Resurser, Forskning och utveckling, Utbildning och undervisning, Lagstiftning och Affärsverksamhet, för vilka vi presenterar rekommendationer och en åtgärdsplan i den här vismansrapporten.

Policy: Vi måste sprida insikten att **språkteknologi har en nyckelposition för att bevara och upprätthålla våra språk och vår kultur**. Språkteknologi behövs t.ex. i den digitala infrastrukturen för den humanvetenskapliga och den socialvetenskapliga forskningen. Det är ingen skillnad om språkteknologin har utvecklats akademiskt, med öppen källkod eller kommersiellt, så länge den finns och språkteknologimodulerna är kompatibla och tillgängliga för att bygga stora system och tillämpningar. Små språksamfund kommer inte att få språkteknologi på kommersiella grunder, så de flesta (eller alla) språk i regionen behöver åtminstone en viss mängd offentligt stöd och somliga kommer kanske att vara helt beroende av det. **På nordisk nivå behöver vi komma överens om rekommendationer** för hur vi skall agera på det nationella planet. För att utvärdera situationen för språkspecifika och språkoberoende resurser för språken i regionen, borde **en BLARK-rapport utarbetas där de grundläggande språkresurserna i Norden kartläggs**. Norden behöver hålla sig ajour med utvecklingen inom EU för att inte upprepa redan gjorda insatser och för att fokusera på det specifikt nordiska. Deltagarna i NODALIDA 2005 beslöt grunda en förening för tal- och språkteknologi, som skall kallas [NEALT \(Northern European Association for Language Technology\)](#). En sådan organisation vore idealisk för att koordinera olika initiativ och nätverk.

Resurser: Den mest uppenbara och viktigaste investeringen vore att skapa en lämplig infrastruktur som har tillräckligt med språkteknologiska resurser för relevanta språk i regionen. Resurserna bör kunna användas fritt för såväl forskning och undervisning som för kommersiell produktutveckling. På basen av den utvärdering av situationen som framkommer av BLARK-rapporten bör de **viktigaste korpusarna skapas på nationell nivå med samarbete på nordisk nivå kring utveckling och utbyte av viktiga språkoberoende redskap och metoder**.

Forskning och utveckling: Finansiärer av akademisk forskning bör anamma rekommendationer och regler för språkresurser som skapas (eller har skapats) med allmänna medel. Det borde vara normal praxis att forskare gör språkresurserna tillgängliga för övriga forskare med så fria villkor och licenser som möjligt. Gemensamma gränssnitt och redskap bör skapas i samarbete med både kommersiella och akademiska parter.

Utbildning och undervisning: Mera samarbete behövs kring akademisk utbildning mellan universiteten i den nordiska och baltiska regionen. En tillräcklig mängd specialister med doktors- och kandidatexamen bör behärska de mest avancerade färdigheterna och alla regionens länder och språkgrupper bör delta inklusive minoriteter och små språkgrupper.

Lagstiftning: Nuvarande lagstiftning om kopieringsskydd gör det onödigt svårt och dyrt att samla korpus. Vissa privilegier ges för tillfället åt några nationella bibliotek för att arkivera elektroniska kopior av böcker, tidningar, osv. och ett liknande privilegium behövs för att skapa språkteknologiresurser. Lagstiftningen borde ändras så att det blir möjligt **att samla in text- och talkorpus som används för forskning och utveckling** av språkteknologiredskap. Att använda dylika korpus **bör anses vara förenligt med principerna om kopieringsskydd när återpublicering av korpusen utesluts.**

Affärsverksamhet: Licensvillkoren för språkteknologiresurser måste tillåta och uppmuntra både kommersiell och akademisk användning. Tillämpad forskning på medellång sikt i samarbete mellan universitet och industri bör uppmuntras.

Åtgärdsplan: Målet med rapporten var att identifiera nyckelområden, storleken på finansieringen, berörda parter och former för samarbete. För att förverkliga målen och för att utarbeta mer detaljerade planer och tidsramar för områdena i 10-årsplanen, föreslår vi att resurser allokeras för:

1. etablering av NEALT och dess arbetsutskott
2. mandat för att utarbeta BLARK-rapporter för de nordiska språken
3. nordisk finansiering av samarbete inom språkteknologisk utbildning och undervisning
4. nationell finansiering av tillämpad forskning på medellång sikt i samarbete mellan universitet och industri

När BLARK-rapporterna har färdigställts, bör resurser under NEALTs koordinering allokeras för:

1. nordisk finansiering av språkteknologiska redskap baserade på BLARK-rapporternas rekommendationer
2. nordisk och nationell finansiering av korpus, trädbanker, och lexikon i enlighet med BLARK-rapporterna