

Forbedrede annotationer og søgemuligheder for ældre litterær tekst: Arbejdsplan for projektperioden T 9 – T 36

DK-CLARIN WP 2.4-arbejdsplan
Jørg Asmussen og Stefan Iversen

Version 1.0 – 13. oktober 2008

Resumé

Nærværende papir indeholder en arbejdsplan for DK-CLARIN WP 2.4
Ældre litterær tekst for det videre projektforløb T 9 – T 36.

1 Status

Grundlæggende beslutninger for projektet blev truffet, de organisatoriske rammer afstukket og en foreløbig arbejdsplan blev udarbejdet.

2 Arbejdsplan

2.1 Arbejdsfordeling

Hovedformålet med JVJ-Centrets del af denne arbejdsplan er at digitalisere en stor del af Jensens forfatterskab med henblik på opmærkning efter samme standard, som anvendes af ADL. 50 længere bogværker forventes digitaliseret. Efter forhandling af rettigheder købes selve digitaliseringen hos KB. Arbejdet med udvælgelsen af og kontrollen med standen af Jensen-tekster vil blive udført af Stefan Iversen, AU med assistance fra Per Dahl, AU.

KB (ved NN) og DSL (ved NN) sørger for konvertering af ADL og JVJ-materialet til det format, som specificeres af WP 5.1 samt (forbedring af) annotering med metadata på tekstniveau.

Hovedformålet med DSL's del af denne arbejdsplan er at skitsere et prototypisk ortografisk variantleksikon, der i en fuld implementation vil kunne forbedre søgemulighederne i ældre tekst.

DSL står endvidere for administration af denne arbejdsplan.

2.2 Resurser

WP 2.4 råder over 0,9375 mio. kr. Heraf er 20% (187.500 kr.) institutionel medfinansiering. KU-CST's andel er 62.500 kr., hvorefter der resterer 875.000 kr. fordelt på DSL og JVJ-Centret med 375.000 til hver og 125.000 til KB. Denne

arbejdsplan beror indtil videre kun på DSL's, JVJ-Centrets og KB's CLARIN-tilskud uden medfinansieringsandelen og uden den andel, KU-CST står for, altså hhv. 300.000 kr. for DSL og JVJ-Centret samt 100.000 kr. for KB, idet resten betragtes som en projektreserve.

Enheden personmåned (PM) er dimensioneret på samme måde som i WP 2.1, jf. [Asmussen og Halskov, 2008]. JVJ-Centret råder over 8 PM, hvoraf 6 udliciteres til KB, idet selve JVJ-digitaliseringen købes hos KB. Der er således allokeret 2 personmåneders arbejdstid fra JVJ-Centrets side samt 6 personmåneders arbejdstid, købt hos KB. KB råder desuden selv over 2,6 PM og DSL over 8 PM, medfinansieringsandelen i alle tilfælde undtaget.

Oveni lønudgifter er der afsat 12.500 kr. til udstyr/drift for JVJ-Centret, hvoraf den institutionelle egenandel udgør 20%.

2.3 Opgaver

Opgave 1: Forhandling af rettigheder med Gyldendal og arvingerne.

Resurseforbrug: 0,25 personmåneders arbejdstid fra JVJ-Centret.

Aflevering 2.4.1: Kort rapport om afklaringen af rettighedsspørgsmålet. T 10.

Opgave 2: Digitalisering og opmærkning.

Resurseforbrug: 1,75 personmåneders arbejdstid fra JVJ-Centret og 6 personmåneders arbejdstid, købt hos KB til denne opgave.

Aflevering 2.4.2: Tekstsamling bestående af 50 værker af JVJ, digitaliseret og opmærket. T 18.

Opgave 3: Annotering. Det nuværende annoteringsniveau dokumenteres samt ønskelige forbedringer specificeres og udføres i muligt (prototypisk) omfang.

Resurseforbrug: Der afsættes 4 PM hertil (2 fra DSL/JA+NN, 2 fra KB/NN).

Aflevering 2.4.3: Specifikation af ADL/JVJ-annotering med prototypiske eksempler. T 28.

Opgave 4: Konvertering. ADL og JVJ konverteres til WP 5.1-specificeret format.

Resurseforbrug: Der afsættes 1,2 PM hertil (DSL/NN: 0,6; KB/NN: 0,6)

Aflevering: Samling af konverterede ADL- og JVJ-tekster klar til indlæsning i DK-CLARIN's søgesystem. T 28.

Opgave 5: Specifikation af variantleksikon.

Resurseforbrug: 4 PM (DSL/JA+NN).

Aflevering 2.4.5: Dokumentation og prototype. T 28.

Opgave 6: Projektadministration.

Resurseforbrug: 1,4 PM (DSL/JA)

Aflevering 2.4.6: Afsluttende projektrapport. T 36.

Litteratur

[Asmussen og Halskov, 2008] Asmussen, J. og Halskov, J. (2008). Reference-korpus for dansk: T 9-status og arbejdsplan. Rapport, DSL/DSN.