

PÅ JAGT EFTER VIDEN I STORE TEKSTMÆNGDER  
FRA HALVAUTOMATISK EKSTRAKTION AF INFORMATIONER TIL STRUKTURERET  
VIDEN

by Lotte Weilgaard,

Institute for Design and Communication, SDU

**Abstract**

Viden om fagbegreber forbindes normalt med en faglig og professionel sammenhæng, men selv i hverdags-situationer har vi brug for at sætte os ind i nye emner og forstå centrale begreber, når vi for eksempel skal kommunikere med det offentlige, med vores bank, med vores læge eller andre. På internettet kan man finde uanede mængder af viden, men som regel ikke i struktureret form, som kan hjælpe os til at få et hurtigt overblik over et nyt emne. Den terminologiske metode er en anerkendt metode til at afklare begrebers betydning og gøre viden tilgængelig i struktureret form. I begrebs-systemer sættes begreber i relation til nært beslægtede begreber. Der indsamles oplysninger om begreberne, som i term-baser præsenteres i en konsistent form med entydige definitioner, hvoraf det fremgår, hvad der er særligt kendetegnende for et begreb, og hvordan det adskiller sig fra andre begreber. Endelig registreres de fagudtryk, der dækker over et og samme begreb samme sted, så man ved, at de refererer til det samme begreb. Ofte og især i virksomheders term-baser anbefales et af flere synonymmer som det foretrukne udtryk. Denne måde at strukturere viden på er med til at sikre en entydig og effektiv (faglig) kommunikation og undgå misforståelser. Den terminologiske arbejdsmetode er meget tidskrævende, men ved hjælp af korpusanalyseværktøjer kan man halv-automatisk ekstrahere information om begreber fra større tekstmængder. En metode er at anvende tilbagevendende sproglige mønstre som for eksempel definere som, karakterisere som, inddele i, bestå af, kalde, betegne som søgemønstre til at identificere information om begreber. Søgeresultaterne herfra bearbejdes herefter til struktureret viden. Halvautomatisk ekstrahering er tilgængelig for alle, men der findes i dag også mere avancerede værktøjer, der kan foretage automatisk vidensekstraktion, herunder af såkaldte terminologiske ontologier.