

## **Præhøring af standarder: kommentarer til indlægget Klassifikation - specificering af en forretningsservice for klassifikation v/Palle Aagaard fra IT- og Telestyrelsen**

9. juli 2009

Kære "klassifikations-team"

Jeg er specialestuderende ved Danmarks biblioteksskole, og deltog i præhøringskonferencen for standarder den 24/6 2009 på et wildcard idet jeg sammen med en medstuderende, skriver speciale omkring ESDH emnesager med Odense Kommunes By og Kulturforvaltning som case.

Med udgangspunkt i det udsendte referat og Palle Aagaards, efterfølgende forkortet PA, kommentar ved oplægget, ikke citeret ordret, men noget i retning af "at der ikke har været meget respons på standarden for forretningstjenesten Klassifikation", tillader jeg mig at knytte nogle kommentarer til præsentationsoplægget. Kommentarerne er set med en cand.scient. bibl's briller, sådan en der uddannelsesmæssigt er opdraget til at systematisere og strukturere informationer og data og for mit vedkommende altid med henblik på videndeling og knowledge management.

REFERAT:

### *Klassifikation*

Palle Aagaard fra IT- og Telestyrelsen præsenterede specifikation for forretningsservicen Klassifikation, se de udsendte plancher

I den efterfølgende plenumsession fremkom følgende kommentarer:

- Ambitionsniveauet synes at være højt. Man bør i en standardiseringsproces fokusere på det konkrete forretningsbehov. Der blev hertil svaret, at servicen Klassifikation er tænkt til at kunne håndtere måden man håndterer klassifikationssystemer på, ikke selve klassifikationssystemerne.
- Der blev yderligere spurgt til, om Klassifikation skal kunne rumme alle former for klassifikationssystemer, hvortil der blev svaret, at der i modelleringen arbejdes på en generisk struktur, der kan rumme så mange som muligt.
- Det blev yderligere bemærket, at man nogle steder har problemer med at anvende klassifikationssystemerne korrekt og der efterlystes en vinkel på at understøtte arbejdsgange, der gør det nemmere at bruge sådanne systemer – altså, hvilke muligheder giver en god klassifikation i anvendelsen af sag og dokument-systemer.

---

### STANDARDEN

For mig at se, er Klassifikation en forretningsservice med rigtig stort potentiale i forhold til at kunne smidiggøre arbejdsgange og dermed procesoptimere i sagsbehandlingen.

Ikke ret langt inde i oplægget, gik det op for mig, hvor vigtig forretningsservicen Klassifikation bliver i forhold til genfindning og videndeling, samt ikke mindst i forhold til genbrug af data og informationer. PA gav et eksempel på en detalje i en typisk sagsbehandlingssituation: sagsbehandleren får ret hurtigt i sin arbejdsproces behov for at lave et opslag op imod forretningsservicen Klassifikation for at tjekke informationer eller hente data fra allerede etablerede klassifikationssystemer.

Derfor er det også ærgerligt at der ikke er kommet mere respons. Jeg tror det er en kombination at et meget fagspecifikt emne, nemlig klassifikation og så, at det faktisk er et temmelig abstrakt emne, der skal tænkes på et metaniveau. Samtidig lider standarden måske lidt under, i hvert fald i kommunalt regi, at medarbejdere der sagsbehandler, måske kun har KL's journalplan at sammenligne med. Desværre er denne journalplan et monster af et semantisk værktøj, og man har svært ved at forstå den. I en samlet arbejdsdag er det derfor ikke forståelsen af Journalplanen der prioriteres, man konstaterer blot, at man skal tildele en journalnøgle "fordi det skal man" jf. plenums reaktion omkring et højt ambitionsniveau, og misforståelsen af hvad det egentlig er der skal standardiseres.

I forhold til genbrug af data og informationer synes jeg, der er meget positivt, at der handles på baggrund af en processuel erkendelse af at "opgaveklassifikation" ikke er dækkende for tjenesten. I et objektorienteret perspektiv, viser dette en proces, hvor man erkender at opgaver ikke skal være en del af klassen Klassifikation, men at opgaver skal bestå som en type. Således frigøres Klassifikation og kan bestå som selvstændig overklasse, der kan indeholde diverse typer og objekter. Dette er et vigtigt i opbygningen af klassificere.

## FAGSPECIFIKKE KOMMENTARER TIL DET AT KLASSIFICERE

I forlængelse af slides nummer 4-7, hvor det præsenteres hvad forretningstjenesten skal kunne rumme og hvordan klassifikationssystemer beskrives, må jeg knytte nogle kommentarer, fordi jeg finder terminologien en anelse upræcis set ud fra en bibliotekars optik. Og nu ved jeg jo ikke, om der er tale om tillempet sprogbrug udelukkende til brug for præsentationen for plenum. Men I kan jo kigge på det, ryste på hovedet, og tænke det ved vi da godt :O)

Når man taler om at klassificere ønsker man således at opbygge en fælles forståelsesramme inden for en given kontekst eller domæne. Med opbygning af en fælles forståelsesramme lægger man således restriktioner på det naturlige sprog. Derfor er processen med at udarbejde rammen for fælles forståelse ikke altid nem. Restriktionerne udgør grænserne for semantisk fortolkning i det fælles sprog. Der er tale om grader af restriktiv tilgang dvs. mere eller mindre rigid styring af præsentationen af en fælles mening. Sådanne typer af fælles sprog og meningsfortolkninger kan betegnes som semantiske værktøjer, der fungerer som typer af indekseringssprog. Der skelnes typisk overordnet mellem egentlige klassifikationssystemer og verbale emneindekseringssprog. I klassifikationssystemer vil emner typisk være udtrykt ved symboler i form af notationer med henblik på en fysisk opstilling. Verbale indekseringssprog er udtrykt ved ord i kontrollerede sprog: struktureret eller ustruktureret samt ukontrollerede sprog: naturligt sprog.

Først og fremmest mener jeg, at der bør skelnes mellem niveauer: når jeg ser sliden "lidt klassifikationsfagligt" indeholder den både systemer og hvad der kan betegnes som metadata, samt ingen hensyntagen til ovennævnte begrebsafklaring.

Dette bør holdes adskilt i hvert tilfælde på det analytiske niveau. Tillige bør der være en politik for vedligeholdelse af det semantiske værktøj f.eks. tilføjelse af nye ord "boblere" eller udgående ord.

at klassificere med henblik på fysisk opstilling

- Klassifikationssystemer jf. sliden vil således være: enumerative systemer og facetterede systemer

at klassificere

- Verbale indekseringsprog jf. sliden vil være: lister ; ontologier ; tesauri (flertalsbetegnelsen) ; hierarkiske systemer (her vil jeg foretrække begrebet taksonomier, idet hierarkiske systemer i princippet kan indeholde alt, der ikke er en flad struktur) S
  - som yderligere bemærkning til det at klassificere bør det nævnes at digitalt indhold kan præsenteres i polyhierarkier dvs. man fraviger kongstanken om at ting, materialer mv. kun kan være repræsenteret et sted jf. klassifikationssystemers tankegang. I polyhierarkier er der derfor tale om sideordnede hierarkier der repræsenterer indhold. Man bør her have

metadata

- Metadata jf. sliden vil være: emneord ; klasser ; grupper ; deskriptorer. Det bemærkes at f.eks. "kategori" og "objekt" er fraværende?

Nedenfor har jeg knyttet kommentarer til de nævnte typer af indekseringsprog jf. sliden.

Lister:

Her bør der skelnes mellem en liste af tildelte emneord - ukontrolleret sprog VS en liste af kontrollerede emneord - kontrolleret, men ustruktureret sprog.

Taksonomier:

Her opbygges en hierarkisk træstruktur, hvor man er enedes om at ophæve et emneord til den foretrukne term. En taksonomi siger ikke noget om relationer mellem termer. Til tider ses blot en opremsning af kategorier uden hierarkisk forhold.

Facetterede klassifikationssystemer:

Opbygget af gensidigt udelukkende kategorier vha. enkeltklasser, der sammensættes efter behov. Håndterer klassifikationstværfaglighed. Der er tale om postkoordineret struktur dvs. strukturen og relationerne mellem termer dannes ved query.

Enumerative klassifikationssystemer:

Opbygget af gensidigt udelukkende kategorier vha. enkeltklasser eller sammensatte klasser der repræsenterer begrebet med tilhørende generisk underopdeling. Der er tale om prækoordineret struktur dvs. alle klasser er opregnet på forhånd. Et enumerativt system håndterer ikke tværfaglighed specielt godt.

Tesauri:

Opbygget ved kontrolleret og stramt struktureret sprog. En thesaurus angiver relationer mellem termer, herunder hvilke termer der kan accepteres, og hvilke der ikke kan. Der er altid henvisning fra ikke ønskede termer til accepterede termer.

Tillige en semantisk restriktion på fortolkning af begreber. Repræsentationen af begreber udgøres symbolsk af deskriptorerne U (use) / UF (use for) ; NT (narrower term) / BT (broader term) ; RT (related term) ; SC (scope note).

Ontologier:

Opbygget ved kontrolleret og struktureret sprog der samtidig er repræsenteret i en eller anden grad af formelt sprog f.eks. 1. order logic, der er det mest strikse sprog dvs. størst grad af formelt sprog. Dette præsenterer ontologien i et matematisk sprog anvendeligt for computeralfabetisk læsning dvs. 0'er og 1'taller. Ontologien opbygges typisk med henblik på computerbaseret udveksling af data og informationer. Dette ligger ikke i thesaurusens grundform, der typisk stopper ved den semantiske definition. Hvorimod ontologien tillige er specificeret i et formelt repræsentationssprog.

Ontologiens grundbestanddele består af termer ; sætninger ; regler ; definitioner. Ontologiens symbolske repræsentation afhænger af det formelle sprog der anvendes. At domænet dvs. ontologien repræsenteres i et formelt sprog muliggør automatisering af de arbejdsopgaver, man ønsker ontologien skal medvirke til at løse. Den formelle opbygning af ontologien er baseret på en fælles terminologi: objekter ; egenskaber hos objekter ; relationer mellem objekter. Dette udgør grundlaget for definitioner og restriktioner for den fælles forståelsesramme, det der senere skal kodes. Ontologien klassificerer ud fra et empirisk belæg og definitioner skal være eksplicite. Det man ikke véd medtager man ikke i en ontologi dvs. ting objekter mv. skal være eksisterende. Ved man nyt i morgen opdateres ontologien, men ikke på forhånd, fordi man tror at noget er eller kommer. Samtidig må der ikke forekomme implicit viden i ontologiens definitioner. Ontologien er jf. kravet om eksplicitet uhyre præciserende i sine definitioner og deklamationer. Hvor en thesaurus angiver relationer mellem semantiske begreber med emnemæssig beslægtethed angiver ontologien relationer dels mellem begreber, dels mellem objekter eller mellem semantisk begreb og objekt. Det er således ikke i ontologien kun emnemæssige relationer, der præciseres. Relationer er betinget af det pågældende domæne, man ønsker at repræsentere ved ontologien. Der kan repræsenteres flere domæner dvs. ontologier i samme ontologiske ramme. Der er så tale om at definere et ontologi-bibliotek, der dokumenterer ontologier og deres indbyrdes relationer. Ontologien kan indeholde thesaurusen som hjælperedskab foruden f.eks. en eksplicitering af diverse applikationer, intelligente agenter, mens thesaurusen populært sagt ikke indeholder redskaber ud over sig selv. Ontologien er derfor et fleksibelt værktøj i forhold til thesaurusen.

## AFSLUTTENDE BEMÆRKNINGER TIL OVENSTÅENDE

Der er ingen nemme genveje i konstruktionen af hverken tesauri eller ontologier, begge værktøjer er tidskrævende at opbygge. Begge værktøjer kræver forarbejde i form af identifikation af termer og begreber samt definitioner på accepterede fortolkninger dvs. restriktioner for forståelsesrammen.

Det semantiske niveau for tesaurusen og ontologien ligner derfor hinanden.

Tillige skal der, for ontologiens vedkommende, kodes. Gevinsten for at bruge tid på en sådan manuel intellektuel proces, skal gerne være en ontologi der kan fungere som inter-operationelt sprog mellem applikationer mv. og derigennem automatisere processer. Motivationen for at anvende ontologien som værktøj kan være a) kommunikation mellem mennesker og organisationer, b) inter-operationelitet mellem systemer og c) applikationer og genbrug af komponenter. Med ontologien får man et verbalt indekseringssprog og semantisk værktøj, der i sin form er fleksibel, men stadig uhyre præcis. Dette værktøj kan være et bud på at håndtere den mangfoldighed af semantiske værktøjer, som forretningstjenesten Klassifikation skal kunne rumme f.eks. KLE ; FORM ; Dansk Branchekode ; Standarderne / forretningstjenesterne altså dokumentation på metaniveau ; ontologi-bibliotek altså dokumentation for ontologiens ontologier. Ontologien vil også kunne indeholde intelligente software agenter, der kan muliggøre sømløs interaktion mellem webservices dvs. inter-operationelitet mellem systemer.

I faglitteraturen er der få krydshenvisninger mellem tesauri og ontologier forstået på den måde, at man kan læse side efter side om opbygning og anvendelse af ontologier eller det omvendte altså tesauri, men uden at disse værktøjer umiddelbart forholdes til hinanden. Ontologi bliver ofte anvendt i forskellige betydninger, hvilket medfører upræcis sprogbrug og forvirring omkring ordet og dets anvendelsesmuligheder. Forståelsen af hvad en ontologi er ligger i forlængelse af følgnede begrebsafgrænsning:

" A formal ontology is a controlled vocabulary expressed in an ontology representation language. This language has a grammar for using vocabulary terms to express something meaningful within a specified domain of interest. The grammar contains formal constraints (e.g., specifies what it means to be a well-formed statement, assertion, query, etc.) on how terms in the ontology's controlled vocabulary can be used together." (kilde: <http://www.metamodel.com/article.php?story=20030115211223271>)

Tesauri, kontrolleret sprog, emneordslister mv. opstår typisk ud af den bibliotekariske tradition, hvor man har skulle organisere den fysiske tilstedeværelse af ting, materialer mv. ude på hylderne. Mens ontologien opstår ud af et andet fagligt felt, nemlig den it og computerbaserede tradition, hvor man stadig organiserer fysisk eksisterende ting, men disse er nu repræsenteret digitalt. Det er også i sidst nævnte tradition man ofte finder begrebet taksonomier og en tankegang, der baserer sig på håndtering af det naturlige sprog i form af automatisk indeksering og automatisk kategorisering. Der er tale om to meget forskellige teknikker. Med den it-teknologiske udvikling får man også mulighed for dynamisk indeksering af emneord i form af tag-clouds, der automatisk opdateres efterhånden som brugerne tildeler tags (emneord). Tagclouden kan siges at være en dynamisk liste typisk i ukontrolleret sprog blot i et andet visningsformat: en statistisk baseret sky, der genereres og opdateres på baggrund af brugernes tildeling af termer.

## LITTERATUR

Kowalczyk, R [et.al] (eds.). Service-Oriented Computing: Agents, Semantics and Engineering: García-Sánchez, Francisco [et.al] (2008). "A Knowledge Technologies-Based Multi-agent System for eGovernment Environments". In: *SOCASE, LNCS* vol. 5006, 2008. kap. 6, pp.15-30. Berlin / Heidelberg: Springer-Verlag. (Serie: Lecture Notes in Computer Science).

Metamodel.com. Forumindlæg ved jernst (2003, 15.jan). "What are the differences between a vocabulary, a taxonomy, a thesaurus, an ontology, and a meta-model?". Lokaliseret den 9.juli 2009 på World Wide Web:  
<http://www.metamodel.com/article.php?story=20030115211223271>

Morville, Peter and Rosenfield, Louis (2002). "Information Architecture : for the World Wide Web". USA CA, Sebastopol: O'Reilly Media, 2002. 2.udg.

Uschold, Mike and Gruninger, Michael (1996). "Ontologies: principles, methods and applications". In: *The Knowledge Engineering Review* vol.11/2, 1996. pp. 93-136.

Uspecificeret undervisningsmateriale udleveret ved Danmarks Bibliotekskole til fagområderne vidensorganisation og katalogisering

Med venlig hilsen

Rikke Graff, specialestuderende

mail to: [sportshveps@aarslevnet.dk](mailto:sportshveps@aarslevnet.dk)

mobil: +45 61861687