

Projektbeskrivelse, tema 3

Titel

Semantik og metadata i byggeprojekter.

Undertitel

Anvendelse af ontologier og metadata til klassificering af information anvendt i byggeprojekter.

Vurdering

Det vurderes at projektet i forhold til dets formål om at forbedre struktureringen af viden indenfor byggeri, har gode muligheder for realisering. Størst vil udbyttet dog være, hvis projektets resultater konkret anvendes i et mere operationelt projekt, som fx projekt om koordineret information, tema 1. Derfor må investeringen i nærværende projekt også ses i forhold til hvad udbyttet i andre projekter forventes at være.

Den største risiko i projektet ligger i, at efterfølgende projekter, der skal udnytte resultaterne fra dette projekt, ikke realiseres, og derved kommer resultaterne til at stå alene uden egentligt operationelt udbytte.

Formål

Projektet har til formål at vurdere mulige metoder til at forbedre struktureringen af den brede og distribuerede viden der eksisterer indenfor de byggetekniske fagdiscipliner. Metoderne udmøntes via informations- og kommunikations teknologier (IKT) i en byggespecifik ontologi, som kan bidrage til, at forskellige typer af teknisk byggeinformation gøres egnede for operationel digital håndtering, f.eks. således at informationer – der handler om samme emne – kædes sammen på en meningsfuld måde, uanset at de stammer fra forskellige domæner. På baggrund af en sådan ontologi skal der udarbejdes et konkret sæt af metadata der kan anvendes til opmærkning af teknisk byggeinformation.

Resultat af projektet vil være udviklede metoder til identificering, herunder søgning, og udnyttelse af viden spredt i byggeprojekter og organisationer. Den udviklede ontologi vil indeholde definitioner på begreber anvendt i byggebranchen og deres indbyrdes relationer, og således udgøre et fælles referencesystem for byggebranchen med hensyn til processer, produkter og den almene byggeinformation. Det er muligt med anvendelse af teknologierne for det semantiske web at håndtere distribuerede data og forbedre struktureringen af information og viden indenfor alle tre videnskategorier der indgår i et byggeprojekts værdikæder. Via ontologier bliver det muligt praktisk og målstyret at håndtere information i intranet og det globale Internet. Ontologierne skal udgøre grundlaget for mere kvalificeret beslutningsstøtte i byggeriets faser, og skabe metoder til opsamling og håndtering af viden og information.

På baggrund af de udviklede ontologier, vil der blive opstillet et konkret forslag til et sæt af metadata til anvendelse af opmærkning af almen byggeinformation. Metadata-sættet skal indeholde nødvendige elementer for tilfredsstillende af identificerede behov og krav fra potentielle brugere. Klassifikationen af information kan ske gennem definition af dokumentklasser og -typer, og klassificeres efter informationernes kontekst og faglige niveau. Selve arbejdsbyrden ved opmærkningen af den byggetekniske information i forhold til det udviklede metadata-sæt bør vurderes og der skal udvikles procedurer for udbydernes håndtering af dette. Vurdering af de økonomiske omkostninger for udbydernes til opmærkning, samt udgifter hvis opgaven ligger i en central organisation, skal foretages.

Projektet har følgende succeskriterier:

- At udarbejde metoder til identificering, strukturering og genanvendelse af begreber og ontologier.
- At støtte kommunikationen og samarbejdet på tværs af fagdiscipliner, baseret på en fælles begrebsforståelse.
- At gensidig nyttiggøre national og international udvikling indenfor området informations- og videnhåndtering. Bl.a. gennem brug af standarder på området fx ISO 82045-5.
- At ontologien skal kunne relateres til øvrige anvendte klassifikationer indenfor byggeri og evt. biblioteksvæsenet.
- At metadatasættet opfylder brugerbehov for opmærkning af alment byggeinformation.
- At ontologi og metadatasæt kan udgøre grundlaget for systemet beskrevet i tema 1

Resultaterne af projektet er af generel natur og forventes at kunne bidrage til formulering og udvikling af domænespecifikke ontologier i andre brancher også.

Effekt

For gennemførelse af projektet er der ikke forudsat gennemførelse af andre projekter først, men effekten af projektet vil først for alvor vise sig, når projektets resultater anvendes i et operationelt system, som fx løsningen skitseret i tema 1. Således vil brugerne af den byggetekniske information kunne få adgang til informationen på en struktureret og målrettet form til rette tid og på rette sted. Resultatet af projektet er således medvirkende til at formålet i tema 1 kan realiseres.

Afgrænsning

Projektet vil afgrænses til at fokusere på information der indgår i byggeprojekters værdikæder fra dedikerede videnleverandører, og til dels information fra materialeproducenter. Struktureringen og klassificeringen af den projektinterne information understøttes i høj grad af arbejdet med klassifikationssystemet under Det Digitale Byggeri.

Den tekniske implementering af projektet vil være understøttet i tema 1. Projektet vil derfor forløbe i to niveauer:

1. Skabelsen og udviklingen af en ontologi for vidensområdet byggeri. Dette indebærer overblik over klassifikationssystemer og anvendte begreber i byggeri samt granskning af nationale og internationale projekter, systemer og øvrig forskning.
2. Udviklingen af et konkret metadatasæt på baggrund af den udviklede ontologi, der tilgodeser krav og behov i byggebranchen for opmærkning af almen byggeinformation.

Den overordnede ontologi, skal udover at tilgodese behovet som opstillet i pkt. 2, også sikre, at dokumenter i bred forstand med relation til byggebranchen kan klassificeres og struktureres.

Budget

Projektets enkelte dele vil kræve en engangsinvestering som tænkes afholdt af de deltagende parter, suppleret af støtte fra finansielle parter med interesser i byggesektorens udvikling. Projektets overordnede tidsramme er 1-2 år, med et estimeret budget i størrelsesordenen 2-5 mio. kr.

Selve driften og vedligeholdelsen af ontologi og metadatasæt vil ligge i organisationen der varetager driften for løsningen beskrevet i tema 1.

Risici

Sikkerheden for realiseringen af de opstillede mål vurderes at være stor, da projektet ikke direkte bygger videre på andre projekter. En risiko i anvendelsen af projektets resultater ligger i, at ingen projekter efterfølgende bruger resultaterne som forudsætning, men snarere opbygger et selvstændigt grundlag. Som selvstændigt projekt og uden efterfølgende projekter, vil der være risiko for at resultaterne aldrig vinder fodfæste i branchen.

Deltagere

- Informationsleverandører
 - Myndigheder (Erhvervs- og Byggestyrelsen, Energistyrelsen, Miljøstyrelsen, Kulturarvsstyrelsen)
 - Forsknings- og uddannelsesinstitutioner (SBI, BYG-DTU, AAU, Kunstakademiets Arkitektskole, Arkitektskolen i Århus).
 - Foreninger, fonde og oplysningsråd (bips, BYG-ERFA, Byggedata, Byggeskadefonden, BvB, Bygningskulturelt Råd, Træbranchens Oplysningsråd, Murerfagets Oplysningsråd, Tagpapbranchens Oplysningsråd m.fl.).

- GTS-institutter (Teknologisk Institut, Dansk Brand- og Sikringsteknisk Institut, DELTA, Dansk Standard)
- Byggeriets organisationer (Dansk Byggeri, PAR, F.R.I., IDA, AAR, BAT-kartellet, Konstruktørforeningen, Teknisk Landsforbund m.fl.).
- Forskningsinstitutioner (Universiteter, Handelshøjskoler, Biblioteksskolen)
- Enkeltvirksomheder.
- Finansielle parter
- Eksterne leverandører til projektet

Det vurderes at have afgørende betydning at projektet gennemføres med bred deltagelse af alle relevante informationsleverandører og under vidtstrakt brugerinddragelse, bl.a. repræsenteret ved byggeriets organisationer.

Til at lede projektet nedsættes en operationel projektgruppe som refererer til en partielt sammensat følgegruppe bestående af repræsentanter for de enkelte hovedgrupper af informationsleverandører, de vigtigste af byggeriets organisationer, samt de medvirkende finansielle parter.

Tidsplan

Projektets tidsramme er 1-2 år.

Aktivitet	Termin
Opnåelse af støttetilsagn til projektet	Maj 2005
Analyse af nationale og internationale projekter og systemer	Sep 2005
Metodevurdering	Okt 2005
Udvikling af ontologi herunder test i mindre skala	Feb 2006
Udvikling af metadatasæt	Apr 2006
Afrapportering mm	Jul 2006

Baggrund

Problemstilling

Byggebranchen har gennem flere år arbejdet med at etablere metoder og systemer for effektiv informationshåndtering, både indenfor virksomheder, i projekter samt mellem videnleverandører og virksomheder. Gennem de seneste år er web-baserede portaler for dokument- og tegningshåndtering (projektwebs) samt systemer for søgning af målrettet information, blevet tilgængelige såvel internt i virksomhederne som eksternt mellem virksomhederne.

Opbygningen af sådanne systemer har øget behovet for strukturering og klassificering af information og viden for bl.a. søgning af information, og dermed også fælles forståelse for anvendte strukturings- og klassifikationsbegreber. Et af problemerne ved brugen af Internettet til søgning af information er søgemaskinernes overflod af hits. Søgmaskinerne søger som oftest efter forekomsten af ord i et dokument, jo flere gentagelser af ordet i dokumentet, jo større hitrate. Problemet er blot, at dette ikke fortæller brugeren noget om, hvad dokumentet egentlig indeholder eller beskriver. Derudover er vigtig information ofte spredt på flere domæner på Internettet og evt. intranet, og de, fra søgemaskinerne returnerede resultatlister, indeholder ikke information om semantiske relationer til andre dokumenter på Inter- eller intranet. Evnen til at definere semantikken af ordene, dvs. at forskellige begreber refererer til det samme objekt eller den samme handling, er afgørende for at få det fulde udbytte af søgningerne.

Behov

Information der anvendes i byggeprojekter kan principielt inddeles i tre kategorier:

1. Projektintern information, der skabes af projektdeltagerne og hvor outputtet fra en projektdeltager er input for øvrige projektdeltagere.
2. Information der hentes fra dedikerede videnleverandører., dvs. almen byggeteknisk information.
3. Information der hentes hos materialeproducenter, dvs. produktanvisninger, monteringsvejledninger etc.

Den projektinterne information og informationen fra materialeproducenter kan naturligt klassificeres i henhold til klassifikationssystemer for byggeri (det forventede DBK). Dette gælder modsat ikke for den almene byggetekniske information, da dette ofte vil have et bredere sigte og ikke udelukkende rette sig mod én be

stemt proces i byggeprocessen. Det er derfor nødvendigt at kunne klassificere denne information på tværs af processer og faser i byggeprojektet.

Med den stadig øgede mængde af information der indgår i et byggeprojekt, følger også et øget behov for strukturering. Behovet eksisterer hos alle parter der indgår i et byggeprojekt gennem alle projekts faser. Dette understøttes af behovsvurderingen i tema 1.

Aktiviteter

Projektet indledes med vurderinger og analyser af eksisterende metoder for strukturering af byggeteknisk information og brugen af ontologier og metadata. Herefter foretages en egentlig detailplanlægning og budgettering af projektets videre forløb, jf. nedenstående pkt. 1-3.

1. Vurdering af metode for strukturering af byggeteknisk information.
2. Analyse af brugen af ontologier og metadata i tilsvarende nationale og internationale projekter og systemer.
3. Detailplanlægning og budgettering af videre projektforbøb.
4. Udvikling af ontologi for vidensområdet almen teknisk byggeinformation, herunder kodificering med teknologier for det semantiske web.
5. Analyse og vurdering af behov for metadatasæt.
6. På baggrund af ontologien at udvikle et metadatasæt til brug for klassifikation af informationer indenfor byggeri.
7. Vurdering af omkostninger for leverandører af teknisk information, herunder vurdering af omkostninger til evt. kontrol og publicering.
8. Kodificering af metadatasæt og udarbejdelse af vejledning i dets brug.

Dokumentation

Ontologier og metadatasæt afleveres kodificeret i en given syntaks, med tilhørende dokumentation for trufne valg og en vejledning i brugen rettet mod de enkelte identificerede brugergrupper. Evaluering af de udviklede produkter og resultater af gennemførte test, tillige med en refleksion over projektet, afleveres som særskilt dokumentation.

Definitioner

Se bilag 1.

Kilder

Se kildefortegnelse i bilag 1.

Bilag 1

Definitioner på begreber og koncepter anvendt i oplægget til temaet om semantik og metadata.

1 Semantik

Definition: Betydningsvidenskab, læren om ordenes betydning og værdi; semantisk: vedr. semantik, betydningsmæssig (Gyldendals Fremmedordbog).

2 Det Semantiske Web

Tim Berners-Lee, der opfandt World Wide Web (WWW) har en vision om et fremtidigt "semantisk web", som kan gøre samarbejdet mellem mennesker og computere lettere. Hans definition lyder (Berners-Lee, 1998):

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

("Det semantiske web er en udvikling af det nuværende web, hvor information gives en veldefineret mening, som muliggør samarbejde mellem mennesker og computere.")

Grundtanken er, at nettet som et hele, kan opbygges mere intelligent og handle mere intuitivt for opfyldelse af brugerens behov. Det semantiske web er således en forestilling om, at man kan udvikle et web, som gør det lettere for mennesker at finde relevant information – og som samtidig gør det lettere for computere at tale sammen. (Kristensen, 2003)

Evnen til at definere semantikken af begreber, er en mangel på det nuværende Internet. Kan der bygges et maskinforståeligt semantisk lag oven på det nuværende Internet, således at computere kan forstå betydningen af ord, som rækker udover ordet i sig selv, vil det være muligt at bringe information til fundamentalt andet niveau (Ahmed et al., 2001, p. 99). Denne definitionsproces skal koble ordet sammen med definitioner, der klargør meningen (Kristensen, 2003). Berners-Lee udtænkte løsninger på, hvordan webudviklere og forfattere til websider kan benytte selvbeskrivende teknikker således, at computere kan forstå en sides kontekst, og dermed udvælge og præsentere hvad brugeren har behov for (Davies, Duke & Stonkus, 2002).

Det semantiske web er at sammenligne med én stor global database, der består af et net af relationer mellem forskellige typer af data. Relationerne muliggør, at en computer kan udføre opgaver, som den ikke direkte vil være i stand til med det net vi kender i dag. (Berners-Lee, 1999, p. 185)

For eksempel som beskrevet i nedenstående scenarium (Kristensen, 2003):

Din søn skal til en serie af behandlinger hos en speciallæge – og du aftaler med din ægtefælle, at I hver især går med ham hver anden gang. Nu starter et kæmpe puslespil, hvor I med hver jeres kalender – og drengens skoleskema, samt lægens sekretær i telefonen forsøger at finde tidspunkter, der kan passe sammen med transportmidler og andre møder og aktiviteter i jeres respektive kalendere. Det semantiske web skal på sigt betjenes af agenter, som skal være intelligente nok til at checke kalendere (hvor alle aftaler så er defineret efter vigtighed – og hvor nogle aftaler kan flyttes rundt) – og finde huller, der matcher de huller, der er i lægens aftaleskema. Derudover kan agenten sikre, at jeres søn ikke forsømmer matematiktimer (fordi matematik er defineret som et fag, han ikke må være fraværende fra, da det er det, han har sværest ved)

Det semantiske web sigter mod at gøre netressourcer tilgængelig for automatiske processer ved tilføjelse af metadata, ontologier og logiske slutningsmuligheder til ressourcernes indhold. World Wide Web Consortium (W3C) arbejder for at fremme det semantiske web som den næste generation af Internettet. RDF¹ er W3C første specifikation for maskinforståelig semantik og metadata. RDF teknologien danner grundlag for yderligere teknologier til specificering af ontologier som RDFS² og OWL³.

3 Metadata

Definition: Metadata er data, der beskriver attributter til en ressource, karakteriserer dennes relationer, samt understøtter dennes genfindning og udnyttelse.

¹ Ressource Description Framework

² RDF Skema

³ Web Ontology Language

Metadata kan anskues som data om data, og beskriver egenskaber om en ressource. Et metadatasæt består af metadataelementer, hver med et navn og en værdi, der beskriver ressourcen, fx en bogs titel, forfatter, forlag etc. Værdien for hvert enkelt metadataelement kan være foruddefinerede i en liste eller kan vælges frit.

4 Ontologi

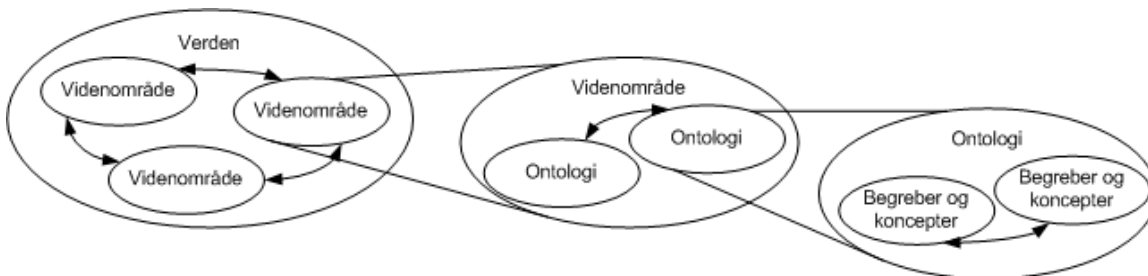
Definition:

- Ontologi: Semantik for begreber.
- I filosofien: "læren om væren".
- I AI (Artificial Intelligence, kunstig intelligens) og WWW-sammenhæng: Dokument eller fil som angiver formelle relationer mellem begreber.
- I udvidet betydning: De begreber og deres relationer som beskriver et fagområde.

En mere præcis definition kan i relation til det semantiske web udtrykkes som:

En eksplicit og formel delt specifikation af hvordan objekter, koncepter, begreber og andre entiteter der forudsættes at eksistere indenfor et vidensområde fremstilles, samt relationerne som binder dem sammen, ses figur 1 (DLI UIUC Glossary). Med *eksplicit* menes at typen af koncepter der benyttes i ontologien og deres brugs begrænsninger, er eksplicit definerede. I det *formelle* ligger at ontologien skal være maskinforståelig, dvs. kodificeret i en syntaks, mens *delt* henviser til at ontologien indeholder konsensual viden – viden der ikke er begrænset til et individ, men accepteret af en større eller mindre gruppe. (Fensel, 2001, p. 11)

En ontologi definerer altså de termer og begreber der benyttes til at beskrive og fremstille et vidensområde. Ontologier benyttes af mennesker, databaser og applikationer, der har behov for at dele domæne information. Et domæne er et specifikt emneområde eller vidensområde som fx medicin, ejendomshandel, byggeri etc. Ontologier indeholder maskinforståelige definitioner af grundlæggende koncepter i domænet og relationerne i mellem dem. De kodificerer viden i et domæne samt viden spredt over flere domæner. På denne måde gør de viden genbrugelig. (Heflin, Volz & Dale, 2002)



Figur 1: Sammenhæng mellem vidensområder, ontologier og begreber.

Selve ordet ontologi har været benyttet til at beskrive artefakter med forskellige grader af struktur. Dette spænder over simple taksonomier (fx Yahoo hierarkiet), til metadatasæts (fx Dublin Core) til mere logiske teorier. Det semantiske web behøver ontologier med en betydende grad af struktur for at kunne specificere beskrivelser for (Heflin et al., 2002):

- Klasser
- Relationer mellem klasser
- Egenskaber for klasser

5 Tesaurus

Definition: En tesaurus er et kontrolleret hierarkisk organiseret vokabular bestående af termer hvor de semantiske relationer mellem termerne vises eksplicit. De semantiske relationer der indgår i en tesaurus kan udtrykkes i en ontologi.

Det kan også betegnes som en liste af synonymer og relaterede ord, så fx en søgemaskine kan finde tilsvarende ord, hvis selve ordet ikke direkte optræder i en tekst. (Sullivan, 2003)

6 Kildefortegnelse

Ahmed, K. et al., 2001, *Professional XML Meta Data*, Wrox Press.

Berners-Lee, T., 1998, *Semantic Web Road map*, W3C.

<http://www.w3.org/DesignIssues/Semantic.html>

Berners-Lee, T., 1999, *Weaving the Web*, Harper San Francisco.

Davies, J., Duke, A., Stonkus, A., 2002, *OntoShare: Using Ontologies for Knowledge Sharing*, BTextact Technologies.

<http://semanticweb2002.aifb.uni-karlsruhe.de/proceedings/Research/davies.pdf>

Fensel, D., 2001, *Ontologies: A Silver bullet for Knowledge Management and Electronic Commerce*, Springer-Verlag Berlin.

Grainger Engineering Library Information Center Digital Library Research Projects, DLI UIUC Glossary.

dli.grainger.uiuc.edu/glossary.htm

Gyldendals Røde Ordbøger; Fremmedord, 1987, Nordisk Forlag A.S, København.

Heflin, J., Volz, R., Dale, J., 2002, *Requirements for a Web Ontology Language*, Working Draft 07 March 2002, W3C.

<http://www.w3.org/TR/2002/WD-webont-req-20020307/#onto-def>

Kristensen, T-M., 2003, *Semnatisk web – bedre samarbejde mellem computere og mennesker?*, Kommunikations Forum.

<http://www.kommunikationsforum.dk/artikler.asp?articleid=10917>

Sullivan, D., 2003, *Search Engine Glossary*, SearchEngineWatch.

<http://searchenginewatch.com/facts/article.php/2156001>