

Open Information
Informationssøgning i Open Source projekter

Gruppe E4-110

10. juni 2005



10. semester Informatik

Titel:

Open Information
- Informationssøgning i Open Source projekter

Project period:

1. Februar til 10. Juni 2005

Projekt gruppe:

E4-110

Gruppe medlemmer:

Henrik Thidemann Farre
Signe Kristensen
Tomas Norre Mikkelsen
Rasmus Olesen

Vejleder:

Per Madsen

Antal kopier: 8

Antal sider: 134

Bilag: A - H

Synopsis:

Mange Open Source projekter har efterhånden så meget information lagret i forskellige informationskanaler, at det kan være svært at vide hvor, hvilken information ligger, hvis det eksisterer. Denne problemstilling belyses i dette projekt og der opstilles designkrav til et CSCW-system med henblik på effektivisering af informationssøgning i Open Source projekter.

I projektet anvendes arketyper til opstilling af deltagertyper i et Open Source projekt. Efterfølgende laves en spørgeskemaundersøgelse og et interview for at validere arketyperne og for at se, om det er muligt at differentiere dem yderligere. Der analyseres på hvilken informationsøgning søges, samt hvordan informationsøgning udføres. Produktet fra analysen er designkrav til systemet. Designkrav baseret på, hvilke karakteristika de forskellige arketyper har samt en diskussion af eksisterende systemer, og ud fra disse, er deres behov fremanalyseret.

Som et resultat af vores analyse og undersøgelse, viser det sig, at arketypernes karakteristika ikke leder til modstridende krav, men forskellige krav. Disse indgår i designkravene.

Systemet er ikke implementeret, da det ikke er projektets fokus. Da systemet ikke er implementeret er det ikke muligt at sige om systemet vil effektivisere informationssøgningen, men med de designkrav der er opstillet, hvor informationssøgningen samles i et system, antages det, at informationssøgningen effektiviseres.

Forord

Dette speciale er udarbejdet på 10. semester Informatik i perioden 1. februar til 10. juni 2005 af gruppe E4-110 på Aalborg Universitet, Institut for Datalogi.

Vi vil i forbindelse med denne rapport gerne takke de personer, der har deltaget i empiri indsamlingen. Disse personer er anonymiseret og vil derfor ikke blive nævnt med navn. Når personerne citeres skrives deres ID, deres deltagertype samt det projekt, de er tilknyttet. Citaterne er direkte indsat fra empirien. Dette betyder, at citaterne kan indeholde stave- og formuleringsfejl.

Henrik Thidemann Farre

Signe Kristensen

Tomas Norre Mikkelsen

Rasmus Olesen

Indhold

1	Indledning	1
1.1	Initierende problemstilling	5
2	CSCW	7
2.1	CSCW	7
2.2	Informationskanaler	9
2.3	Awareness	14
2.4	Group memory	18
2.5	Opsummering	21
3	Problemstilling	23
3.1	Hypoteser	23
3.2	Problemformulering	25
4	Metode	29
4.1	Arketyper	30
4.2	Design	35
5	Målgruppe	37
5.1	Målgruppebeskrivelse	37
5.2	Kriterier for udvælgelse af projekter	41
5.3	Beskrivelse af projekterne	42
5.4	Problematik omkring informationskanaler	43
6	Spørgeskema	45
6.1	Beskrivelse af spørgeskemaet	45
6.2	Evaluering af spørgeskema	52

6.3 Opsamling på hypoteser	58
7 Identifikation af arketyper	61
7.1 Bestemmelse af arketyper	61
7.2 Prioritering af arketyper	68
8 Interview	69
8.1 Beskrivelse af interview	69
8.2 Evaluering af interview	75
8.3 Opsamling på hypoteser	76
9 Arketyperne	79
9.1 Analyse af arketypernes karakteristika	79
9.2 Arketype 1	81
9.3 Opsummering arketype 1	87
9.4 Arketype 2	88
9.5 Opsummering arketype 2	94
9.6 Arketype 3	95
9.7 Opsummering arketype 3	100
10 Eksisterende systemer	103
10.1 Udvalgelse af eksisterende CSCW systemer	104
10.2 Eksisterende systemer i projekterne	104
10.3 Andre systemer	114
11 Designkrav	117
11.1 Designkrav	117
12 Konklusion	123
12.1 Konklusion	123
13 Refleksion	127
13.1 Arketyper	127
13.2 Den rette løsning	128
13.3 Brugen af eksisterende systemer	129

13.4 Spørgeskema software	129
13.5 Empiri	129
14 Videreudvikling af systemet	131
14.1 Gør det Open Source	131
14.2 Test af systemet	132
Litteratur	134
Figurliste	138
Tabelliste	140
A Mail udsendt på informationskanaler	1
B Spørgeskema	3
C Pilotforsøg	9
D Data fra spørgeskema	13
D.1 Diagrammer for 7 initiale typer	13
D.2 Diagrammer for reviderede typer	16
E Justeringer af svar	23
F Repræsentative deltagere	25
G Interview besvarelser	27
H Spørgeskema besvarelser	35

1 Indledning

“I know more now where to look and realize that MPlayer-devs don’t accept your questions unless you first thuroly research an subject/issue before you ask them anything, ...or you will get your head chopped off ;-/” Spørgeskemarespondent, MPlayer projektet

Det er naturligt for mennesker at samles i grupper for at opnå fælles mål eksempelvis for at producere et fysisk produkt, skrive litteratur eller udvikle software. Der er mange fordele forbundet med at være samlet på samme tid og lokation. Eksempelvis kan folk kommunikere let og hurtigt og her igennem udveksle information. At være samlet på samme lokation kan betyde, at det er let at danne sig et overblik over ting i projektet, eksempelvis hvad der laves, hvem der laver det og hvornår det skal laves. Dette gør, at der er gode muligheder for koordinering af arbejdet, og at man derved undgår unødvendigt spildt arbejde, f.eks. ved at personer i gruppen arbejder med det samme.

Der eksisterer dog tilfælde, hvor det ikke er muligt, fornuftigt eller rentabelt, at arbejdet foregår på samme lokation og på samme tid. I nogle situationer er det fordelagtigt, at arbejde distribueres ud. Det giver f.eks. mulighed for parallel udvikling og produktion af komponenter uafhængig af hinanden. Det virker umiddelbart som en god løsning til f.eks. at udvikle produkter hurtigt og herved opnå økonomiske fordele [Herbsleb og Grinter, 1999]. Ved nærmere undersøgelse er der dog en del problemstillinger, der skal tages højde for. Eksempelvis besværliggøres det at kommunikere, koordinere arbejde samt at dele information i gruppen.

Olson og Olson [2000, s. 1 og 152] og Mockus og Herbsleb [2003, s. 1] beskriver, at samarbejde bliver kompliceret, når det foregår distribueret. Herudover beskriver Mockus og Herbsleb [2003, s. 1] også, at der er et antal specifikke forskelle på distribueret arbejde i forhold til ikke distribueret softwareudvikling. I distribueret udvikling er der ofte mindre kommunikation blandt deltagerne og formidling af vigtig information foregår mindre effektivt. Endvidere beskriver de, at deltagere

som er adskilt af distance, typisk har en mindre følelse af “teamness” altså holdånd eller gruppefølelse. Disse forskelle kan have indflydelse på, at kommunikationen besværliggøres i distribueret udvikling. Mindre kommunikation og formindsket mulighed for at få vigtige informationer kan betyde en større nødvendighed i, at personer i distribueret udvikling selv kan finde den nødvendige information på en let og effektiv måde.

Det kan være svært at finde information omkring hvem, der laver hvad, når man ikke har mulighed for at spørge en person ansigt-til-ansigt. Gutwin et al. [2004, s. 73] beskriver, at meget af den kommunikation samt implicite information, der er tilgængelig i en gruppe, som er samlet på samme lokation, *ikke* er tilgængelig for medlemmer i distribuerede grupper. I ekstreme situationer kan konsekvensen være, at arbejdet bliver nødt til at stoppe, indtil man kan få adgang til en vigtig information fra en person, der befinder sig et andet sted i verden [Herbsleb og Grinter, 1999, s. 68]. Dette kan f.eks. medføre ressourcepild i form af tid og penge. Den ansigt-til-ansigt kommunikation, der eksempelvis foregår i gruppen, er ikke længere mulig, når man arbejder med distribueret udvikling. Derfor kræves en indsats for at opnå denne kommunikation.

Behovet for at koordinere arbejde og dele information viser sig tydeligt i distribueret *software* udvikling. Et typisk stykke software har mange forskellige komponenter, som indbyrdes er afhængige af hinanden, og en enkelt ændring i bare en linje kode kan have store konsekvenser for resten af koden. Dette kan medføre mere arbejde og flere ændringer andre steder i koden. Hvis dette ses i konteksten af, at et stort distribueret projekt kan bestå af mange millioner linjer kode, kan man begynde at forstå, at koordinering og planlægning er en vigtig for projekterne. Disse problemer med samarbejde problematiseres også i artiklen af Olson og Olson [2000]. I artiklen beskriver de blandt andet, hvordan man i et projekt valgte at reorganisere, så man fokuserede på, hvordan man får delt arbejdet op, så en person ikke er afhængig af andre [Olson og Olson, 2000, s. 152].

“Work was reorganized so that people did not have to rely on tight collaboration with a remote team member. For example, the software design effort was reorganized to partition the design work into loosely coupled modules, assigning all the work of each module to one location, the others to the remote location.” [Olson og Olson, 2000, s. 152]

De fleste Open Source projekter udvikles distribueret. Et typisk Open Source projekt består af mange forskellige mennesker fra hele kloden, der arbejder sammen via Internettet. Et eksempel på et distribueret Open Source projekt er Apache projektet som Mockus et al. [2002, s. 310] beskriver som følgende:

“..an extreme case of geographically distributed development, where developers work in arbitrary locations, rarely or never meet face to face, and coordinate their activity almost exclusively by means of email and bulletin boards.” [Mockus et al., 2002, s. 310]

Open Source udviklingsprocessen beskrives ofte som en ustruktureret og kaotisk proces, hvor flere faktorer spiller ind på udviklingsprocessen. Mockus et al. [2002, s. 312] beskriver nogle af faktorerne som følgende:

- *“OSS systems are built by potentially large numbers (i.e., hundreds or even thousands) of volunteers. It is worth noting, however, that currently a number of OSS projects are supported by companies and some participants are not volunteers.*
- *Work is not assigned; people undertake the work they choose to undertake.*
- *There is no explicit system-level design, or even detailed design.*
- *There is no project plan, schedule, or list of deliverables.”*

[Mockus et al., 2002, s. 310]

Som man kan se ud fra ovenstående citat, virker det som om, at det er svært at udvikle Open Source systemer effektivt, da det ikke virker som en struktureret udviklingsproces. Man kan derfor tro, at effektiv koordinering og deling af information næsten er umuligt i et Open Source projekt. Der findes dog alligevel mange Open Source projekter, der klarer sig godt på trods af de opstillede problemer. For at håndtere problemerne benytter projekterne sig af en række informationskanaler til at formidle information og planlægge arbejdet. Disse informationskanaler bruges ikke kun af udviklerne men også af brugerne. Nogle af informationskanalerne er målrettet til udviklerne, som f.eks. versionsstyring samt issuetrackers, mens andre er til mere almen brug, som eksempelvis mailing lister. Et problem med mange af disse informationskanaler er, at de indeholder en del “støj” i form af irrelevante diskussioner og spørgsmål, der er besvaret adskillige gange før. Dette betyder, at der findes en betydelig stor mængde information, hvoraf noget er relevant og andet irrelevant.

Chancen for at den information man søger eksisterer, er større, når mængden af information vokser. Samtidig risikerer man dog også, at det er sværere at finde informationen, idet der er mere information at søge blandt. Der kan f.eks. opstå problemer med, at man finder meget irrelevant information. Søgningen kan desuden forværres, hvis informationen ligger spredt på forskellige informationskanaler, da det kan involvere en væsentlig mængde arbejde at finde informationen, og endnu mere hvis disse har hver deres måde til at søge information. Hvis man eksempelvis søger information på mailing lister, er der en type interface og en måde at finde information på, mens der i issuetracking systemer er et andet interface og en anden måde at søge information på.

Hvis en person, i stedet for at søge efter svar på problemer, ønsker at stille et spørgsmål, har Eric Raymond¹ lavet en guide, der fortæller, hvorledes man bør stille spørgsmålet for at forbedre chancen for et ønsket svar. Et problem heri er, at man, ifølge Eric Raymod, skal undersøge nedenstående, før man stiller et spørgsmål.

“Before asking a technical question by email, or in a newsgroup, or on a website chat board, do the following:

- 1. Try to find an answer by searching the Web.*
- 2. Try to find an answer by reading the manual.*
- 3. Try to find an answer by reading a FAQ.*
- 4. Try to find an answer by inspection or experimentation.*
- 5. Try to find an answer by asking a skilled friend.*
- 6. If you are a programmer, try to find an answer by reading the source code.”*

[Eric Steven Raymond, 2004]

Det fremgår af citatet, at man skal søge efter information mange steder, inden man stiller sit spørgsmål. Det betyder igen, at personen der søger efter en bestemt information, bliver nødt til at arbejde sig igennem meget data, hvilket kræver ressourcer i form af tid og energi. Udover at det kræver tid at søge i mange informationskanaler, kan der også opstå problemer med informationssøgning, hvis man ikke kender til alle informationskanaler. Dette kan betyde, at man ikke finder den information man leder efter.

De eksisterende informationskanaler, der benyttes i Open Source projekter, muliggør samarbejde i form af koordinering og deling af information i en gruppe. På baggrund af de ovenstående problemer kan man så spørge sig selv, om disse informationskanaler er tilstrækkelige, eller om det er muligt at effektivisere arbejdsgangen. Eksempelvis ved at tilbyde et interface, hvorigennem det er muligt at foretage informationssøgningen på tværs af de forskellige informationskanaler, der benyttes i Open Source projekterne.

Gutwin et al. [2004] påpeger i deres artikel “Group Awareness in Distributed Software Development”, at der er behov for et system, der kan samle information fra de forskellige informationskanaler, der benyttes i Open Source projekter og herved lette informationssøgningen. Et andet eksempel på dette er artiklen “Architectures, Coordination, and Distance: Conway’s Law and Beyond” [Herbsleb og Grinter, 1999] der konkluderer følgende:

“Distributed organizations desperately need tools that make it easier to find organizational information, to maintain awareness about the

¹ Medstifter af Open Source Initiative: www.opensource.org

availability of people, and to have more effective cross-site meetings, especially spontaneous ad hoc sessions.” [Herbsleb og Grinter, 1999, s. 70]

1.1 Initierende problemstilling

Indledningen leder os frem til følgende problemstilling: Er det muligt at forbedre samarbejdet i Open Source projekter ved at udvikle et system, der effektiviserer informationssøgningen?

Som problemstillingen viser, kan der være brug for at lave et system til at forbedre samarbejdet i Open Source projekter. Vi antager, at dette system kan forbedre samarbejdet ved at effektivisere informationsøgning i projekterne. Dette er baseret på antagelserne beskrevet i indledningen. Derfor vælger vi, at dette projekt skal fokusere på et system, som bidrager til samarbejde i Open Source projekter.

For at arbejde med områder og problemstillinger der er i distribueret udvikling, samt diskutere hvad disse kan bidrage med til vores problemstilling, beskrives og diskuteres begrebet CSCW samt begreberne awareness og group memory, da disse er begreber indenfor dette område. Dette leder frem til en række hypoteser om, hvad brugere og udviklere i et Open Source projekt oplever, når de søger information.

Hypoteserne bidrager til at opbygge vores problemformulering, hvor vi konkretiserer vores problemstilling.

CSCW 2

I dette kapitel diskuteres forskellige begreber i konteksten af distribueret software udvikling med det formål at få en forståelse af disse. Det første begreb er CSCW og herunder også, hvorledes begrebet anvendes i projektet. Dernæst følger vores forståelse af information og de kanaler, information kan formidles igennem i Open Source projekter. Derefter diskuteres og defineres begreberne awareness og group memory i konteksten af dette projekt.

2.1 CSCW

Som det beskrives i indledningen, er distribueret software udvikling i dette projekt defineret som software, der udvikles af personer, som ikke befinder sig på samme lokation. Vores interesse i distribueret software udvikling fokuserer på Open Source projekter.

For at kunne samarbejde effektivt benytter udviklerne og brugerne i Open Source projekter en række forskellige software systemer. Disse systemer hører til en gruppe af software, som kaldes CSCW.

CSCW står for “Computer Supported Collaborative Work” eller “Computer Supported Cooperative Work”. Det dækker over forståelsen af, hvordan personer i grupper, eksempelvis i distribueret software udvikling, arbejder sammen, og hvorledes deres arbejdsprocesser understøttes af computer applikationer [Telekoooperation, 1997]. I gennemgangen af litteraturen virker det tilfældig om det andet C i CSCW står for Collaborative eller Cooperative. Terzis [1998] differentierer dog mellem disse to. Cooperative bruges om samarbejde i mindre grupper; collaborative anvendes omkring gruppensamarbejdet i organisationer og store projekter, hvor flere personer deltager.

Forskning inden for CSCW beskæftiger sig både med de psykologiske, sociale og organisationelle effekter brugen af computer applikationer har. [Telekooperation, 1997]. Terzis [1998] beskriver CSCW som følgende:

“CSCW is a multidisciplinary research field including computer science, economics, sociology, and psychology. CSCW research focuses on developing new theories and technologies for coordination of groups of people who work together.”

Som det ses af Telekooperation [1997] og Terzis [1998], omhandler CSCW forskellige områder. Dette illustreres i, at der er forskel på de områder de nævner. Telekooperation [1997] nævner organisationelle effekter som område og Terzis [1998] nævner områder som økonomi og computer videnskab.

Netop fordi CSCW er et multidisciplinært forsknings område, findes der flere forskellige specialiserede grene indenfor CSCW. Den gren der arbejdes med i dette projekt er information, da vores problemstilling omhandler søgning af information i Open Source projekter. Den information, som findes i projekterne, defineres i dette projekt som følgende:

- Information som omhandler, hvilket arbejde der bliver udført, af hvem det bliver udført, samt hvor det bliver udført. Denne type af information betegnes “awareness” information, og diskuteres i afsnit 2.3.
- Generel information, der dækker over alle de informationer der er tilknyttet til projektet. Vi vælger at beskrive denne type af information som begrebet “group memory”. Vi benytter dette begreb da det fokuserer på viden i en gruppe, og vi ser et Open Source community som en gruppe der deler viden. Begrebet diskuteres i afsnit 2.4.

De computer applikationer der indgår i CSCW-systemer kaldes *groupware*. Disse er applikationer som eksempelvis chat og videokonference systemer. Terzis og Nixon, s. 8 diskuterer flere forskellige definitioner af hvad groupware er. De konkluderer, at de fleste af definitionerne enten er for snævre eller for brede og at betegnelsen “groupware” betyder forskellige ting for forskellige personer [Terzis og Nixon, s. 8]. De ender op med at anvende definitionen af Greenberg [1991]:

“software that supports and augments group work” [Greenberg, 1991]

Dette er ligeledes den forståelse af groupware der bliver brugt i dette projekt. De informationskanaler der beskrives i afsnit 2.2 falder alle ind under denne definition.

Der er udviklet adskillige CSCW-systemer. Disse indeholder ofte flere forskellige typer groupware og er målrettet til distribueret udvikling. Eksempler på CSCW systemer til distribueret udvikling er Hipikat [Čubranić et al., 2004] og Jazz [Cheng et al., 2003].

2.1.1 Vores anvendelse af CSCW

CSCW er et nøglebegreb i dette projekt, idet projektet handler om samarbejdet i grupper, og hvordan de bruger computer applikationer til at kommunikere og dele information. Det system, som projektet opstiller designkrav til, er et nyt CSCW-system, der skal forbedre samarbejde i Open Source projekter. Systemet skal ikke alene forbedre samarbejdet blandt udviklere men også mellem brugere. Umiddelbart kan det være problematisk at se, hvorledes samarbejde kan forbedres blandt brugere. Vi mener dog at brugere samarbejder med hinanden. Eksempelvis kan en bruger hjælpe en anden bruger med at finde noget information eller løse et problem. Samtidig mener vi også at brugerne samarbejder med projektet og dermed også udviklerne, idet de bidrager til projekts akkumulerede viden. Ye og Kishida [2003] beskriver det således:

“Community members regularly interact with each other for knowledge sharing and collaboration in pursuit of solutions to a common class of problems.” [Ye og Kishida, 2003, s. 419]

Vi mener, at vores system, ved at lette tilgangen til information, vil effektivisere samarbejdet. Som der bliver beskrevet i indledningen, kan der opstå nogle problemer omkring samarbejde i en gruppe, når arbejde udføres distribueret. Disse problemer mener vi vil blive formindsket ved, at lette informationssøgningen for brugere og udviklere i Open Source projekterne. Eksempelvis er man ikke afhængig af at skulle vente på en anden person for at få den information, man har brug for.

De ovenstående typer af informationer er tilgængelige igennem en lang række forskellige *informationskanaler*. I det følgende diskuteres disse, og efterfølgende præsenterer informationstyperne.

2.2 Informationskanaler

Deltagere i Open Source projekter kommunikerer og samarbejder ved hjælp af virtuelle miljøer, som Internettet danner grundlag for [Feller og Fitzgerald, 2002, s. 125]. Disse virtuelle miljøer eksisterer i form af forskellige informationskanaler, som beskrives og diskuteres i dette afsnit.

Vi vælger at differentiere mellem informationskanaler, der er offentligt tilgængelige og, dem der ikke er, samt de informationskanaler der er officielle og dem der ikke er. Begrundelsen for dette er, at alle der ønsker at deltage i et Open Source projekt, vil have mulighed for at kende til disse informationskanaler, da de er officielle, og have adgang til dem da de er offentlige. Vi mener yderligere, at dette valg er nødvendigt, da det netop belyser problemstillingen med at finde information i Open Source projekter. Der findes et utal af informationskanaler, der tilbyder

hjælp og information til Open Source projekter, hver med deres eget community. På grund af denne mangfoldighed kan det være svært og uoverskueligt at finde de rigtige informationskanaler, og det kan være svært at finde rundt i dem.

Kravet om at informationskanalerne skal være offentligt tilgængelige, udelukker offline information, som eksempelvis data lagret på en desktop computer og dokumentation i form af bøger og artikler, som ikke tilbydes via Internettet. Fravalget af ikke offentligt tilgængelige informationskanaler medvirker også til, at informationskanaler som f.eks. personlige e-mails og Instant Messaging, som f.eks. ICQ og MSN, ikke medtages i denne beskrivelse. Gutwin et al. [2004, s. 73] nævner også, at awareness information kun virker, hvis diskussionerne mellem udviklerne foregår i "public channels".

Vi definerer officielle informationskanaler som informationskanaler, der er linket direkte til eller henvist til, fra et projekts hjemmeside. Informationskanaler der ikke overholder dette, betegnes som 3. parts. Et eksempel på dette er <http://www.tek-tips.com/>, der har et forum¹ til at diskutere billede-behandlingsprogrammet The Gimp. Dette forum er der hverken linket eller henvist til fra hjemmesiden, og er derfor ikke en officiel informationskanal ifølge vores definition.

Selvom informationskanaler der ikke er officielle eller ikke offentligt tilgængelige er fravalgt, betyder det ikke, at de er uden værdi for deltagerne i et Open Source projekt. Eksempelvis kan udviklere have brug for information, der ikke er offentligt tilgængelig, f.eks. fra mails. Samtidig kan en bruger stille spørgsmål til andre via informationskanaler som Instant Messaging. Informationskanalen Instant Messaging er også interessant i konteksten af samarbejde, da det er et værktøj der kan indikere brugerens "presence", dvs. om brugeren er online, optaget eller fraværende i øjeblikket. Man kan have en antagelse om, at denne information er nyttigt i en samarbejdssituation, eksempelvis har Cheng et al. [2003] inkorporeret IM i Eclipse udviklingsmiljøet.

Tabel 2.1 viser en oversigt over de mest udbredte informationskanaler i Open Source projekter ifølge Feller og Fitzgerald [2002, s. 93, 125-133]. Desuden er Weblogs også medtaget, hvilket dog ikke er nær så udbredt som de andre informationskanaler. Grunden til at denne informationskanal er medtaget er, at den indeholder konkrete awareness informationer om den person, der har Webloggen.

Til beskrivelsen af informationskanalerne er der opstillet en række begreber, som beskriver forskellige attributter informationskanalerne kan have. De valgte begreber er yderligere valgt ud fra, at de overholder de kriterier, der er opstillet omkring officielle og offentligt tilgængelige informationskanaler.

- En-til-mange: Informationen udsendes fra én afsender og modtages af mere end en modtager. Eksempler på dette er en person, der sender en e-mail til flere modtagere. En-til-en, som f.eks. IM, er i denne sammenhæng ikke interessant, pga. kravene omkring offentligt tilgængelighed.

¹ <http://www.tek-tips.com/threadminder.cfm?pid=451>

- Mange-til-mange: Information afsendes og modtages af flere personer. Et eksempel på dette er en mailing list, hvor der er mange afsendere og mange modtagere [Feller og Fitzgerald, 2002, s. 127].
- Asynkron: Udsendelsen og modtagelsen af informationen foregår ikke på samme tidspunkt. Et eksempel på dette er e-mails, hvor man har muligheden for at læse dem uafhængigt af, hvornår de er blevet udsendt.
- Realtime: Modsat asynkron, skal man ved realtime kommunikation være til stede, når informationen bliver udsendt. Dette er f.eks. tilfældet i chat. Vi vælger at bruge begrebet realtime frem for synkron, da synkron ikke siger tydeligt nok at det er her-og-nu kommunikation, vi henviser til.
- Persistens: Informationen bliver lagret permanent og kan tilgås senere. Et eksempel på dette er e-mail arkiver, hvor e-mails sendt til mailing listen er arkiveret. I princippet kan alle informationskanaler godt være persistente, dog er dette eksempelvis i forbindelse med chat typisk ikke normalt. Gutwin et al. [2004] nævner f.eks.:

“Since text chat is not archived, any developers who are not reading a particular conversation will lose access to that information.” [Gutwin et al., 2004, s. 77]

	En-til-mange	Mange-til-mange	Asynkron	Realtime	Persistens
Website	×		×		×
Fora		×	×		×
Nyhedsgrupper		×	×		×
Mailing list		×	×		×
Chat		×		×	
Versionsstyring		×	×		×
Weblog	×		×		×
Issuetracker		×	×		×

Tabel 2.1: Oversigt over forskellige informationskanaler, der bliver benyttet i Open Source projekter

I det følgende gennemgås hver af informationskanalerne fra tabellen (se tabel 2.1).

Websites

Websites tilbyder en lang række forskellige informationer, eksempelvis dokumentation samt “gateways” til nogle af de andre informationskanaler, som nyhedsgrupper og mailing lists. Et website som f.eks. Sourceforge.net [Sourceforge.net, 2005] tilbyder fora samt web gateway til de mailing lister, der er hosted der. Websites er persistent information. En funktion som nogle websites gør brug af, er en Wiki². En Wiki tillader, at alle personer kan redigere hjemmesiden, og tilbyder en historik over de ændringer, der er foretaget, og gør det muligt at se ældre versioner af websiten.

Alle større Open Source projekter er i besiddelse af et website [Capiluppi et al., 2003]. Med website mener vi en selvstændig informationskanal og ikke en side på enten Sourceforge [Sourceforge.net, 2005] eller freshmeat [Freshmeat.net, 2005]. Eksempelvis, har projektet Xine både information på Sourceforge [Sourceforge.net, 2005] og deres eget website [The xine-Project, 2005].

Fora

Et forum er en webbaseret informationskanal til at diskutere forskellige emner, samt til at udveksle tanker og ideer ved at skrive og besvare indlæg. Fora er typisk emne opdelt, hvori brugere og udviklere har mulighed for at oprette diskussions-tråde. Informationskanalen fora er en af de meget brugte på Internettet [Limanto et al., 2005, s. 978].

I modsætning til f.eks. nyhedsgrupper kan informationen i et debat fora præsenteres i et pænt layout. Webforaet indholder endvidere ofte også en dedikeret søgning, som søger i det specifikke fora [Limanto et al., 2005, s. 978].

Nyhedsgrupper

Nyhedsgrupper er en samling af diskussionsfora. Disse er inddelt efter emner, eksempelvis findes der den danske nyhedsgruppe for Unix systemer: dk.edb.system.unix. Usenet servere, som nyhedsgrupperne ligger på, synkroniserer beskeder mellem hinanden, således at en besked lagt ud på en dansk Usenet server distribueres rundt til servere over hele verden.

Nyhedsgrupperne tilgås typisk gennem et dedikeret program, men det er også muligt at benytte en web gateway som f.eks. Google Groups³. Google Groups arkiverer også nyhedsgrupper groups tilbage fra 1981 [Google, 2005], så disse må regnes som persistente.

² Se <http://wiki.org/wiki.cgi?WhatIsWiki> for mere information omkring Wiki'er

³ <http://groups.google.com>

Populariteten af nyhedsgrupper er faldende, da fokus langsomt er skiftet til mailing lists og websites i de fleste Open Source projekter [Feller og Fitzgerald, 2002, s. 126].

Mailing lister

Mailing lister fungerer ved, at man sender en mail til én e-mail adresse, hvorefter den sendes ud til alle de personer, der abonnerer på listen.

Mailing lists benyttes af stor set alle Open Source projekter. Ofte benyttes flere mailing lists i det samme projekt, f.eks. en liste til brugerne og en til udviklerne. Ved flere af de store Open Source projekter er der meget aktivitet på listerne, og udviklerne benytter mailing listerne som den primære informationskanal [Feller og Fitzgerald, 2002, s. 127].

Mailing lister adskiller sig fra normal e-mail, altså personlig en-til-en eller en-til-mange, ved at være mange-til-mange kommunikation. I en mailing liste har alle adgang til at poste og læse e-mails.

Ud over at størstedelen af alle Open Source projekter har mailing lists, tilbyder projekterne også webbaserede arkiver, der muliggør søgning i listerne [Feller og Fitzgerald, 2002, s. 127].

Chat

Chat kan foregå via et webbaseret chatrum eller et dedikeret chatprogram. I Open Source projekter bruges ofte programmer, der benytter IRC⁴ protokollen, som eksempelvis XChat. Fælles for disse chat programmer er, at personerne er samlet i et virtuelt rum, og beskeder der skrives i dette rum, kan læses af alle i rummet. Det er dog også muligt at fortage én-til-én samtaler på IRC.

Kommunikation, der foregår som chat, er oftest af mindre teknisk karakter. Der tales også om emner, der ikke omhandler Open Source projekter. Chat arkiveres kun sjældent [Gutwin et al., 2004, s. 76].

Versionsstyring

Versionsstyringsystemer benyttes af udviklerne i Open Source projekterne til at håndtere information som f.eks. kildekode, dokumentation og Websites. Kildekoden samles f.eks. i et fælles depot, hvor udviklerne har muligheden for at tilgå den. Systemerne gør det også muligt, at tilgå ældre versioner af filerne og have sideløbende udgaver af den samme fil, såkaldte "branches". Flere udviklere kan også arbejde på kildekoden samtidigt, og systemet kan hjælpe med at flette overlappende ændringer sammen. Yderligere holder systemerne styr på metadata om

⁴ Forkortelse af Internet Relay Chat

kildekoden, eksempelvis hvem der har arbejdet hvor i koden og hvornår. Det primære formål ved versionsstyringssystemer er dog at styre versioner. Eksempler på systemer der f.eks. styrer versioner af kildekode, er CVS og Subversion. CVS er det mest udbredte system til håndtering af kildekode i Open Source projekter. Versionsstyringssystemer bruges i stort set alle projekter [Feller og Fitzgerald, 2002, s. 128]

Weblogs

Weblogs, også kaldet blogs, er et forholdvis nyt fænomen⁵ (i forhold til e-mail og websites). Weblogs har et dagbog lignende format. Personen der har blog'en, skriver jævnligt hvad personen arbejder med. Enkelte Open Source projekter samler relevante blogs, f.eks. fra udviklerne i projektet. Eksempler på dette er <http://planet.mozilla.org> og <http://planet.gnome.org>. Indlæg på blogs er normalt emne inddelt, således kan det pågældende projekt udvælge relevante indlæg og kun vise disse.

Blogs er en-til-mange kommunikation, da der er en afsender. Dog kan man klassificere blogs som en mange-til-mange informationskanal når flere af disse samles et sted. Det er i mange blogs også muligt at kommentere det oprindelige indlæg, og på den måde kan blogs også tilbyde tovejs kommunikation.

Issuetracker

Informationskanaler af typen Issuetracker dækker over en række forskellige systemer, der har det fælles mål, at de holder styr på forskellige dele indenfor et projekt. Dette kan eksempelvis være fejl i software, som man kan holde styr på gennem en såkaldt bugtracker, eksempelvis Bugzilla, eller ønsket funktionalitet, som ligges i såkaldte issuetracker.

Disse systemer regnes for at være mange-til-mange kommunikation, da det normalt er muligt for mange forskellige personer at oprette emner i disse systemer og kommentere på andres indlæg.

Efter gennemgang samt beskrivelse af de valg der foretages angående informationskanalerne, vil det efterfølgende være en beskrivelse og diskussion af den information som disse kanaler tilbyder.

2.3 Awareness

CSCW systemer kan bruges til at fremme awareness i et samarbejde, ved at gøre denne information tilgængelig for gruppens medlemmer igennem et computer sy-

⁵ Ifølge Wikipedia [2005] opstod begrebet "weblog" i 1997, og i 1999 opstod ordet "blog"

stem. Awareness svarer til de danske begreber *opmærksomhed*, *bevidsthed* eller *bevidstgørelse*. Awareness kan bruges om samarbejdssituationer, og det er i denne forbindelse, vi anvender begrebet, idet der i distribueret software udvikling er tale om en samarbejdssituation. Alternativt kan begrebet benyttes i en kontekst, hvor der ikke er tale om samarbejde, f.eks. omkring bevidsthed om en persons omverden. Eksempelvis påpeger Schmidt [2002, s. 288], at begrebet “awareness” ikke altid er direkte forbundet med samarbejde i en gruppe, men sommetider kun periferisk relateret til samarbejde. Vi vælger at støtte op omkring, hvordan man ser awareness i CSCW, på grund af at dette projekt omhandler et CSCW system, som skal hjælpe deltagere i Open Source projekter med at finde information. I de følgende afsnit vil awareness blive diskuteret for at få et dybere indblik i, hvad awareness dækker over, samt hvilken form for awareness, der er bedst egnet til et CSCW system.

For at en distribueret gruppe kan samarbejde om at udvikle et stykke software, er det nødvendigt med bevidsthed om, hvad de forskellige medlemmer i gruppen foretager sig. Gutwin et al. [2004, s. 2] skriver, at awareness er information, der er kritisk for effektivt samarbejde, hvilket er en vigtig del i distribueret softwareudvikling. Eksempelvis for at undgå, at flere personer arbejder på det samme, eller at en person laver en ændring, der har indflydelse på resten af systemet. Til at beskrive den information gruppen skal have bevidsthed om, anvendes begrebet “awareness”. Awareness omhandler således de informationer, man har om andre og de informationer man giver andre i en samarbejdssituation. Informationer om hvem man er, hvad man laver, eller har lavet på et projekt, og hvad man vil lave i fremtiden hører alle under awareness. I dette projekt vælger vi netop at se awareness som information.

2.3.1 Awareness’ mange betydninger

Som nævnt i starten af dette afsnit, så er samarbejde ikke alene nok til at dække awareness’ betydning. Awareness kan have mange forskellige betydninger og definitioner i forskellige henseender [Schmidt, 2002, s. 287]. Schmidt [2002] ser derfor et problem ved awareness. Ordets mange betydninger kan have en forvirrende indvirkning på folks opfattelse af ordet, samt den måde det skal bruges. Ordet kan betyde en række forskellige ting, afhængig af den kontekst det bruges i, som adjektiverne indikerer. I følge Schmidt [2002, s. 287] kan man kun bruge awareness i sammenhænge, hvor ordet bruges om en persons awareness af et eller andet. Det fremgår af følgende citat:

“That is, the term ‘awareness’ is only meaningful if it refers to a person’s awareness of something.” [Schmidt, 2002, s. 287]

Dette er også måden awareness ses i dette projekt. Vi fokuserer på udvikleres og brugeres awareness af, hvad forskellige personer i Open Source projektet har lavet, laver, samt vil lave.

2.3.2 Awareness i konteksten af CSCW systemer

Man skal være opmærksom på, hvad det er for samarbejdssituationer, man forsøger at implementere i sine systemer [Schmidt, 2002, s. 288]. Man kan få svært ved at implementere den information, som man kan tilegne sig ved at møde andre på sin arbejdsplads. Situationer hvor man tilfældigt møder en kollega og spørger ham til råds angående arbejde, vil ikke være muligt at implementere i CSCW systemer. Denne opfattelse har Kjeldskov og Nyvang [2000, s. 47] også. De ser en problemstilling i, at man ikke kan opnå understøttelse af tilfældige og uformelle møder, som de kalder for "*just in time awareness*". Denne form for awareness mener Schmidt [2002, s. 289] er skabt ud fra en social kontekst frem for en direkte samarbejdssituation. Kjeldskov og Nyvang [2000] påpeger endvidere, at man kun kan opnå awareness ved at aktørene selv skaber kontakt med hinanden.

Dette projekt behandler derfor kun awareness information, der er muligt at formidle igennem et CSCW system. I formuleringen af designkravene til vores system ses derfor bort fra formidlingen af eksempelvis non-verbal kommunikation.

Ikke alene kan awareness benyttes i forskellige situationer, men det kan også anvendes på forskellige niveauer. Nedenstående beskrives generel og specifik awareness, som er awareness på to forskellige niveauer.

2.3.3 Specifik og generel awareness

Gutwin et al. [2004] beskriver to former for awareness, der adskiller sig fra hinanden ved abstraktions niveau. Den første form er, hvad Gutwin et al. [2004] kalder *generel* awareness. Denne form for awareness giver det brede overblik over et projekt, eksempelvis hvem er hvem, og hvem laver hvad i projektet. Denne form for awareness kan opnås igennem tekstbaserede (mailing lists, chat) informationskanaler [Gutwin et al., 2004, s. 72]. Gutwin et al. [2004] citerer en NetBSD udvikler, der meget godt beskriver denne form for awareness:

"Since I tend to hop around I really try to keep a 10,000 ft. view of what is going on across the entire system. I watch the commits (but I do not read all of them) and of course developer discussions. Both of those activities are enough to give me a sense of what is going on."
[Gutwin et al., 2004, s. 75]

Kantor et al. [1997] benytter begrebet "Project awareness" til at beskrive dette. Ved project awareness forstås awareness i et større perspektiv. Project awareness drejer

sig om, at de enkelte individer skal have indsigt i, hvad en gruppe som helhed foretager sig, og om større tidsrum som f.eks. uger i stedet for timer. Det gælder altså om, at brugerne skal opnå en bevidsthed om arbejdet i projektet de deltager i.

Den *specifikke* form af awareness som Gutwin et al. [2004] beskriver, omhandler målrettet informationssøgning, hvor generel awareness vedligeholdes kontinuerligt ved f.eks. at følge med på en mailing liste. Specifik awareness er der brug for, når en udvikler eksempelvis vil arbejde et andet sted i koden, end der hvor denne typisk arbejder. Udvikleren leder derfor aktivt efter, hvem der evt. arbejder med koden, har arbejdet eller vil arbejde med koden. Gutwin et al. [2004, s. 78] har identificeret fem kilder, der gør det muligt at finde information omkring, hvem der står for en del af projektet:

1. Maintainer field, altså en bestemt udvikler står nævnt i en fil som den, der er ansvarlig for denne del
2. Repository logs, hvem har arbejdet sidst med filen
3. Issue and bug trackers, hvem er en issue eller bug tilskrevet
4. Spørge andre udviklere
5. Projekts dokumentation

For at adskille de forskellige niveauer som awareness fungerer på, benyttes Gutwin et al. [2004] betegnelser også i dette projekt. Da generel og specifik ligger i hver sin ende af skalaen, bør man dog være opmærksom på, at der selvfølgelig er mange niveauer imellem disse betegnelser. Begge former er nødvendig i distribueret udvikling idet de hjælper til at løse forskellige problemer. Eksempelvis hjælper den generelle awareness med at holde overblik, mens den specifikke form hjælper med koordinering på et meget lavere niveau, når der f.eks. skal arbejdes med et stykke kode.

I de ovenstående afsnit er awareness betydning samt awareness i konteksten af CSCW blevet diskuteret. I det følgende afsnit beskrives awareness i konteksten af Open Source. Dette er vigtigt at beskrive pga. den typiske opfattelse af Open Source udvikling som kaotisk og ustruktureret [Mockus et al., 2002, s. 312].

Awareness i konteksten af Open Source

Gutwin et al. [2004] har undersøgt awareness i tre forskellige Open Source projekter. Awareness opnås ifølge disse undersøgelser igennem tekstbaseret informationskanaler som eksempelvis mailing lister.

“Although there are several sources of information, this awareness is maintained primarily through text-based communication (mailing lists and chat systems). These textual channels have several characteristics that help to support the maintenance of awareness, as long as developers are committed to reading the lists and to making their project communication public.” [Gutwin et al., 2004]

Ifølge forfatteren har disse tekstbaserede informationskanaler flere karakteristika, som kan understøtte awareness, så længe deltagere i projekterne selv følger med ved at læse f.eks. mailing listerne. Hvis dette ses i konteksten af diskussionen i afsnit 2.3.2, så betyder det, at kun en bestemt form for awareness kan formidles igennem disse informationskanaler. Den form som Kjeldskov og Nyvang [2000] betegner som “just in time awareness”, burde derfor ikke være muligt at formidle i Open Source projekter. Dog beskriver Gutwin et al. [2004, s. 77], at chat gør det muligt at opnå denne uplanlagte awareness information. Eksempelvis nævner de følgende:

“... provide exactly the kind of ‘water-cooler awareness’ that has been described for colocated projects” [Gutwin et al., 2004, s. 77]

Vi er dog af den opfattelse, at selv om en vis mængde “just in time awareness” kan opnås igennem chat, formidles meget af den information, eksempelvis non-verbal kommunikation, man opnår ved ansigt-til-ansigt kommunikation ikke igennem chat eller andre CSCW systemer.

2.3.4 Awareness i dette projekt

Vi vælger at forstå awareness i samarbejdssituationen, og som generel og specifik awareness. Vi mener, at det er muligt at implementere disse former for awareness i et CSCW system.

Vi mener også, at awareness er essentiel information, der er nødvendig i Open Source projekter, for at kunne samarbejde både for brugere og udviklere.

Dette projekt har ikke til hensigt at undersøge, om det er muligt, men derimod undersøge hvilke informationer, herunder awareness informationer, deltagerne i Open Source projekter ønsker at kunne søge efter.

Idet vi forstår awareness som information, er der en tæt relation til group memory, som det næste afsnit omhandler. Denne relation vil blive behandlet i det følgende.

2.4 Group memory

Group memory er en betegnelse, der dækker over den samlede information, som eksisterer i en gruppe. Group memory bruges af gruppens medlemmer til at finde

information samt eksempler på, hvordan forskellige ting skal gøres. Et eksempel på dette kan i forbindelse med software udvikling være, hvordan en bestemt funktion skal implementeres, eller hvordan et problem løses.

Afhængig af konteksten benyttes andre betegnelser om dette koncept. F.eks. i forbindelse med en organisation og læring i en organisation, bruges "organizational memory". På et overordnet plan, hvor det handler om viden, der lagres i en social gruppe, bruges betegnelsen "social memory". Berlin et al. [1993, s. 23] benytter følgende definition for group memory i konteksten af CSCW:

"We define a group memory broadly as a common repository of on-line, minimally structured information of persistent value to a group"

Definitionen er, som de selv påpeger, bred. Dette betyder, at mange forskellige systemer falder ind under definitionen. Det største problem ved det er, at en del af disse systemer ikke er designet specielt til at formidle group memory. Dette kan muligvis betyde, at informationerne er svære at genfinde, eksempelvis fordi at søgefunktionaliteten er meget begrænset. Group memory kan derfor også eksistere i forskellige informationskanaler, eksempelvis versionsstyringssystemet til kilde koden, mailing lister, så længe at denne informationskanal gemmer informationerne persistent (se tabel 2.1 for en oversigt af dette).

I dette projekt forstås begrebet i konteksten af CSCW systemer. Altså group memory er den viden, der er blevet formidlet igennem et CSCW system og lagret persistent. Group memory er den del af et CSCW system, der kan ses som en vidensbase. Vi er også af den opfattelse at group memory indholder forskellige informationer, alt efter hvilken informationskanal den er formidlet igennem. Eksempelvis vil versionsstyringssystemer indeholde informationer der er knytte til et artefakt, kilde kode, dokumentation osv.

I konteksten af CSCW systemer og distribueret udvikling kan der opstå nogle problemer i forbindelse med formidlingen af group memory. Čubranić et al. [2004, s. 82] beskriver, hvordan nye deltagere i distribueret software udvikling kan have svært ved at lære af erfarne udviklere, fordi de ikke bare kan gå over til denne og spørge, eller se den erfarne udvikler over skulderen. Hvis udviklerne er samlet på det samme sted, er det både lettere og hurtigere at spørge den person, der er nærmest. Čubranić et al. [2004] påpeger dog også, at situationen ikke er helt håbløs, da meget af den information en ny udvikler har brug for, findes i f.eks. mailing list arkiver eller versionsstyrings systemet til kilde koden. Dog nævner Čubranić et al. [2004, s. 83] en yderligere problemstilling, der ligger sig meget tæt op af den centrale problemstilling i dette projekt:

"However, this information is not easily accessible because of its sheer volume, the lack of tools to search the information effectively, and the difficulty of making connections between logically related items in disparate repositories."

Ud fra den ovenstående diskussion er det tydeligt at let adgang til group memory er essentiel for både udviklere og brugere i Open Source projekter. Det er specielt vigtigt for de Open Source projekter, der har lidt eller ingen dokumentation, idet group memory lagret i mailing lister, fora samt andre informationkanaler er den eneste viden, der findes offentligt tilgængeligt i projektet.

2.4.1 Relation mellem awareness og group memory

Som nævnt i det tidligere afsnit omkring awareness er der en relation mellem awareness og group memory, fordi vi opfatter awareness som information.

For at klarlægge denne relation, benyttes den følgende mail fra en Subversion udvikler til mailing listen for udviklerne som eksempel:

```
"To:dev@subversion.tigris.org
From: <Subversion developer S1 >
Subject: [Patch] links page
- fixed link to cvs2svn page
- added trac
- added submaster
<no follow-up messages >" [Gutwin et al., 2004, s. 75]
```

Ovenstående citat er en mail sent fra en udvikler fra Subversion projektet til de andre udviklere på mailing listen. I denne mail formidles en del awareness information, nemlig hvad udvikleren har *fixed* samt *added*. Og som vi beskriver i afsnit 2.2 omkring informationskanaler, så er mailing lister persistente. Derfor mener vi at awareness informationen går over til at blive del af gruppens akumulerede viden, altså en del af group memory.

2.4.2 Group memory i dette projekt

Vi ser group memory som en essentiel del af et Open Source projekt. Kombinationen af decideret dokumentation, manualer, samt den øvrige information, der er at finde i et Open Source projekts informationskanaler, er en stærk kombination. Dokumentationen og manualerne er praktisk information som kan bruges af alle til at forstå systemet. Den øvrige information er f.eks. diskussioner mellem deltagerne i projektet angående problemer og ideer eller awareness information. Når disse typer information samles som group memory i projektet, komplimenterer de hinanden, og opnår derved en stærk kombination. Der er dog ligeledes et paradoks i at mere og mere information er tilgængelig som group memory. I takt med at mængden af information stiger, vil det efterhånden blive sværere at finde den information man søger. Dog øges chancen for at den information man søger efter eksisterer. Hvis mængden af information i group memoryen er lille, så er det lettere at finde informationen, men igen en mindre chance for, at informationen eksisterer.

Vi er af den opfattelse, at dette paradoks kan løses ved at implementere et system, der effektivt kan søge og finde relevant information blandt store mængder information.

2.5 Opsummering

Igennem de foregående afsnit er opnået en dybere forståelse af begreberne som er centrale i dette projekt. Vi forstår CSCW som et multidisciplinært forsknings, som i konteksten af, at vores projekt omhandler information i form af awareness og group memory. Gennem diskussion af awareness og group memory er vi nået frem til en forståelse af hvilke informationer der er tilgængelige i et distribueret udviklings projekt, samt hvordan begreberne bidrager til at samarbejdet i et sådant projekt forløber effektivt.

I det følgende opstilles en række hypoteser, der tager deres udgangspunkt i diskussionen af de ovennævnte begreber. Efter hypoteserne konkretiseres det problem dette projekt omhandler.

Problemstilling

I dette kapitel opstilles tre hypoteser, som udspringer fra kapitlet omkring CSCW og den initierende problemstilling. Hypoteserne er et udtryk for de områder, vi ser et behov for at belyse. Hypoteserne er et grundlag for vores problemformulering, som efterfølgende beskrives i kapitlet. Afslutningsvis følger en afgrænsning af problemet, hvor de begreber som indgår i problemformuleringen defineres.

3.1 Hypoteser

Hypoteser omhandler, hvilke problemer vi forestiller os, brugere og udviklere i Open Source projekter oplever, når de søger information. Den samlede betegnelse for brugere og udviklere der benyttes fremover er *deltagere*.

Hypoteserne inddeles efter, hvad de omhandler. De tre kategorier er:

- Informationssøgning
- Information
- Komplexitet

3.1.1 Informationssøgning

Den første hypotese lyder:

Hypotese 1 *Formålet med informationssøgning, hvorledes søgningen udføres, problemer angående søgning af information og hvilke informationskanaler der søges i, afhænger af typen af deltager.*

Denne hypotese bygger på en antagelse om, at forskellige typer af deltagere i Open Source projekter, har forskellige formål med at søge information og forskellige måder at finde frem til den ønskede information. Deltagerne søger information både afhængigt af deres rolle i projektet, men også på baggrund af, hvilken situation de befinder sig i. Det vil derfor også være forskellige søgningsproblemer som personer oplever.

Hypotesen dækker også over en forestilling om, at deltagerne bruger forskellige informationskanaler til forskellige opgaver. F.eks. bruges nogle kanaler til at browse efter information, hvor andre bliver brugt til at stille spørgsmål på forskellige relaterede problemstillinger i forhold til projektet. Valget af informationskanal kan også afhænge af hvilken type information deltagerne søger. For eksempel vil nogle informationskanaler måske egne sig bedre til at finde information om ændringer i kode, mens andre kanaler, som f.eks. chat, er gode til off-topic debat.

Endvidere antager vi, at brugerne har forskellige formål og motivation med at søge information om projektet. Nogle deltagere er måske motiveret til at søge information for at forbedre deres programmerings kompetencer, mens andre deltagere i stedet ønsker at følge udviklingen af projektet. Disse forskellige formål kan have betydning for deltageres handlemønstre og hvordan de foretrækker at søge information.

Denne hypotese problematiserer, hvordan og hvorfor deltagere i Open Source projekter søger information. Derudover problematiserer den også, at der kan være forskellige informationssøgnings problemer for forskellige deltagere.

3.1.2 Information

Hypotese 2 *Deltagere har forskellige behov for hvilken slags information, der skal være tilgængelig i et projekt.*

Denne hypotese dækker flere antagelser omkring, hvilken information de forskellige deltagere i Open Source projekter har brug for. Det er samarbejde og koordinering, individuelle behov, samt antal af deltagere, der spiller ind på deltageres behov for information i projekterne.

Endvidere antages det, at jo mere engageret en deltager er, jo mere behov har personen for at vide, hvad de andre deltagere laver. Det vil sige, at udviklere har større behov end brugerne. Hvis personen er udvikler, antager vi således, at personen har behov for at vide, hvad der ændres i projektet, da det kan have påvirkning for det, han selv udvikler på. Yderligere afhænger behovet for awareness også af hvor mange deltagere der er tilknyttet projektet. Antagelsen er, at jo flere personer, jo sværere bliver det for den enkelte deltager at have overblik over, hvad de andre laver.

Deltagerne har af samme grunde også forskellige behov for hvad og hvor meget projektet skal indeholde af group memory.

Denne hypotese hænger typisk sammen med deltagerens formål med at anvende systemet. Hvis deltagerens formål er at finde eksempler på kode, vil personen have behov for, at denne form for group memory er tilgængelig i projektet igennem en af informationskanalerne.

Denne hypotese problematiserer, at det er forskellige former for information, forskellige deltagere i Open Source projekter har behov for.

3.1.3 Komplexitet

Hypotese 3 *Komplexiteten af informationssøgning afhænger af størrelsen af det pågældende projekt, og hvor længe en deltager har været involveret i projektet.*

Størrelsen af et projekt afhænger af antallet af deltagere involveret i projektet samt antallet af informationskanaler der anvendes. Grunden til at størrelsen måles ud fra disse to aspekter er, at vi mener, at det er de to primære kilder til at kompleksiteten af informationssøgning øges. Vores antagelse er således, at kompleksiteten af informationssøgning følger størrelsen af projektet. Dette skyldes, at vi mener, at kompleksiteten øges, idet der bliver produceret mere information, når der er flere personer involveret, eller ved at informationen bliver mere spredt, når der er flere informationskanaler.

Vi antager, at dette typisk opstår når et Open Source projekt har eksisteret i længere tid. Dette skyldes, at projektet formodentlig akkumulerer mere information med tiden, samt at antal af informationskanaler og brugere vokser.

Den sidste antagelse er at jo længere tid, deltagere har været involveret, des bedre kender de til informationskanalerne. Både med hensyn til at søge information i kanalerne, men også hvor de forskellige informationer findes. Derved kan søgningen efter information blive lettere jo længere tid deltageren har været involveret.

I denne hypotese problematiseres således, hvordan størrelsen af et projekt gør deltagerens informationssøgning mere kompleks, men samtidig at deltagerens erfaring kan gøre det lettere at søge information. Hypotesen indeholder derfor to modstridende antagelser.

3.2 Problemformulering

Ud fra de beskrevne hypoteser beskrives projektets problemstilling.

Alle tre hypoteser omhandler en problemstilling i forhold til deltagerne. Kravene til vores system udspringer derfor fra de forskellige deltageres behov. Yderligere er informationssøgning et omdrejningspunkt for hypoteserne, og derfor indgår dette også som et centralt element i vores problemstilling.

Problemstillingen:

“Hvilke designkrav er der til et system, som skal effektivisere informationsøgning og derved samarbejdet blandt deltagere i Open Source projekter?”

Systemet fokuserer på informationsøgning i Open Source projekter. I CSCW afsnittet beskrives, at udveksling af information er vigtig, for at kunne koordinere samarbejde, når der udvikles distribueret. Vores system skal derfor fokusere på forbedring af informationssøgning. Designkravene til systemet, skal således muliggøre effektivisering af informationssøgning i Open Source projekter.

3.2.1 Problemafgrænsning

Dette afsnit uddyber den ovenstående problemstilling ved at definere de begreber, som anvendes i problemstillingen samt andre steder i projektet. På den måde afgrænses projektet, samt fokus for projektet tydeliggøres.

- Ved designkrav fokuseres på funktionalitet og design af systemet. Dette inkluderer, måder systemet skal fungerer på, og funktioner det skal tilbyde. Der tages udgangspunkt i deltagerne i Open Source projekter og herudaf identificeres deres krav til systemet. Der er ikke fokus på tekniske krav, og hvorledes systemet kan implementeres.
- Begrundelsen for at lave designkrav til ét system er, at systemet skal kunne anvendes af alle deltagere i Open Source projekter. Samtidig skal systemet samle informationssøgning til ét sted således det er lettere at foretage informationssøgning.
- Systemet skal være et CSCW system, da der igennem effektivisering af informationsøgning, er tale om forbedring af samarbejde blandt deltagerne i Open Source projekter.
- Med “effektivisere informationssøgning” menes, at systemet skal gøre det lettere for en deltager at søge information på tværs af de informationskanaler, der benyttes i Open Source projekter. Systemets formål er ikke at afvikle de nuværende informationskanaler, men at samle søgningen af information i dem.
- Vi begrænser projektet til at omhandle de informationskanaler, der er beskrevet i afsnit 2.2.
- Målgruppen for vores system er aktive deltagere i Open Source projekter¹, der udvikles distribueret. At deltagerne er aktive betyder, at personerne an-

¹ Projekter der udgives under en licens der overholder Open Source Definitionen, se <http://opensource.org/docs/definition.php>

vender informationskanaler som eksempelvis; *mailing lister*, *fora* samt *issuetracking* systemer. Personerne kan således både være brugere, som f.eks. læser eller spørger til noget angående et projekt, eller det kan være udviklere, der bidrager til den videre udvikling af systemet. Personerne kan endvidere have forskellig erfaring indenfor det projekt de deltager i. Det vil sige, at personerne kan være alt fra nybegyndere til øvede. Systemet skal udvikles så det dækker hele målgruppen. Vi mener, at målgruppen er alle deltagere, der oplever informationssøgningsproblemer, og at alle deltagerne samarbejder. Systemet skal derfor kunne bruges af alle.

I det følgende kapitel beskrives metoden til at besvare problemformuleringen.

4 Metode

I kapitlet beskrives, hvorledes problemformulering besvares. Dette sker i gennem to dele: Første del omhandler identificering og beskrivelse af karakteristika ved deltagere i Open Source projekter og anden del omhandler udarbejdelse af designkrav til vores system. I dette kapitel beskrives således metoder til at udføre disse to dele.

Den første del af projektet, som omhandler deltagernes karakteristika, laves ud fra en modificering af en metode, udviklet af Kujala og Kauppinen [2004]. Den modificerede metode bruges til at opbygge og strukturere projektet ned til anden del af projektet, som er designdelen. Til designdelen beskrives en separat metode omkring, hvordan denne del udføres. I de enkelte kapitler i projektet beskrives hvordan og hvorfor de enkelte dele af projektet udføres.

For at besvare problemstillingen tages udgangspunkt i en undersøgelse og analyse af deltagernes karakteristika ved deres informationssøgning. Disse karakteristika benyttes til at udarbejde en liste af designkrav til vores system. En overvejelse er, om en analyse af eksisterende systemer skal indgå på et tidligt tidspunkt i undersøgelsen. Med eksisterende systemer henvises der til CSCW systemer, der løser lignende problemstillinger, som den der er opstillet i problemformuleringen (se afsnit 3.2). En analyse af eksisterende systemer kan give supplerende ideer til designkravene til vores system. Vi vælger dog i vores undersøgelse at fastholde et fokus på brugerinddragelse og et fokus på deltagernes karakteristika. Dette betyder derfor, at en analyse af eksisterende systemer først inddrages i designdelen, for at bidrage med supplerende viden og ideer til designkravene.

For at fokusere på deltagernes karakteristika vælges en metode ud fra et princip om arketyper. Ideen bag arketyper er at opdele målgruppen i en række repræsentative typer. Der indsamles empiri omkring typerne til at beskrive og validere dem, hvilket sker igennem analyse af typernes karakteristika. Indsamling af empiri sker først igennem en spørgeskemaundersøgelse og herefter ved hjælp af et uddybende

skriftligt interview. De resultater vi får på baggrund af svar fra arketyperne, analyseres til en række opstillede designkrav til vores system.

Udover at identificere arketyper bruges spørgeskemaundersøgelsen og det skriftlige interview også til at validere hypoteserne (se afsnit 3.1 for at konkludere, om vores antagelser holder). De hypoteser der ikke valideres igennem spørgeskemaundersøgelsen, forsøges valideret gennem uddybende spørgsmål i det skriftlige interview.

4.1 Arketyper

Formålet med dette projekt er at udarbejde en liste af designkrav til et system, der skal støtte informationssøgning i Open Source projekter. Dette gøres ud fra en analyse af karakteristika af deltagere i Open Source Projekter. Vores interesse ligger ikke i de tekniske aspekter af et sådant system, men i hvilke designkrav systemet skal overholde for at være brugbart for en stor og varieret målgruppe (se problemafgrænsning afsnit 3.2). På grund af målgruppens størrelse og variation er det kun muligt at involvere en begrænset mængde af målgruppen i udviklingen af et system [Kujala og Kauppinen, 2004]. For at løse denne problemstilling vælger vi at beskrive et antal repræsentative deltagere, såkaldte arketyper, der hver især repræsenterer en del af målgruppen. Ved at anvende arketyper opnår vi information om alle forskellige typers krav. Ud fra disse krav kan der opstilles et design af et system, så det bliver brugbart for alle deltagere, der ønsker at anvende systemet.

Goodwin [2002] bruger en form for arketyper, som han kalder persona. Denne måde at lave en arketype på, fokuserer på at udvælge specifikke personer, som er i målgruppen for systemet. Persona er en beskrivelse af denne enkelte persons måde at opføre sig på, mønstre, mål, kvalifikationer, holdninger, miljø, samt personlige detaljer [Goodwin, 2002, s. 1]. I stedet for en detaljeret liste der beskriver personens opgaver og rolle, er det en mere fortællende beskrivelse af før omtalte karakteristika af personen [Goodwin, 2002]. Vi vælger ikke persona, da det vil give en masse information, som ikke er nødvendig for at besvare vores problemformulering (se afsnit 3.2). I dette projekt vil det netop være en detaljeret beskrivelse af deltagerens karakteristika der omhandler informationssøgning og ikke f.eks. en fortællende beskrivelse af deltagerens handlemønstre i deres dagligdag. Kujala og Kauppinen [2004] beskriver også et problem med persona, som vi også ser i forbindelse med vores undersøgelse. Kujala og Kauppinen [2004, s. 297] beskriver persona som en præcis beskrivelse af en hypotetisk bruger og dennes mål, som skal repræsentere hele gruppen. Dette mener Kujala og Kauppinen [2004] kan være interessant i den sammenhæng hvor man ikke skal sælge produktet til hele målgruppen, som beskrevet i følgende citat:

“For example, Cooper reports that the Dodge Ram pick up was a best seller even if 80% of people in a focus group hated it“ [Kujala og Kauppinen, 2004, s. 298]

Men dette er ikke altid formålet, som f.eks. når et systems succes afhænger af en stor mængde af forskellige bruger grupper [Kujala og Kauppinen, 2004, s. 298]. Som Kujala og Kauppinen [2004] beskriver det, vil man have brug for at designe til en stor varieret målgruppe, når man designer til organisationer i stedet for individualister (se efterfølgende citat).

“In particular, varied user groups need to be considered in cases where the system is designed for organizations and not for individuals.” [Kujala og Kauppinen, 2004, s. 298]

Da dette projekt også fokuserer på organisationer, i dette tilfælde Open Source projekter, samt at systemet skal anvendes af alle deltagere, er det nødvendigt at undersøge hele målgruppen i stedet for enkelte individer som persona gør det.

Der hvor vi ser, at persona og den måde Kujala og Kauppinen [2004] bruger arketyper, adskiller sig fra hinanden, er ved at persona er afbildninger af typiske brugere i målgruppen, hvor den form for arketyper Kujala og Kauppinen [2004] bruger, er repræsentative typer for hele målgruppen. Kujala og Kauppinen [2004] beskriver Cooper [1999] persona som:

“His solution is to narrow the design target to a single group of users (persona or user archetype) even if it represents a minority of users” [Kujala og Kauppinen, 2004, s. 298]

Da målgruppen til vores system er alle deltagere i et Open Source projekt, mener vi ikke, at den metode Cooper [1999] har udviklet, er velegnet som metode til at identificere deltageres krav til vores system. Hvorimod Kujala og Kauppinen [2004] måde at se arketyper på, kan bruges når man skal designe til hele grupper, og giver mulighed for at analysere krav fra repræsentative typer i hele målgruppen.

Måden hvorpå arketyper skal forstås i dette projekt, er ud fra forståelsen som Kujala og Kauppinen [2004] bruger. Deres forståelse stemmer overens med måden, hvorpå vi vælger at besvare vores problemformulering, da vi vil identificere deltagerne i Open Source projekter og beskrive deres krav til information og informationssøgning.

En arketype er i dette projekt en deltager i et Open Source projekt, der repræsenterer en større del af målgruppen. I stedet for blot at benytte en unuanceret beskrivelse af målgruppen, benyttes arketyper til at give en detaljeret beskrivelse af et udsnit af deltagerne og deres karakteristika. Ved at opstille disse arketyper kan der eksempelvis identificeres om der er nogle af typerne, der har modstridende

karaktistika, som skal tages højde for i designdelen. Ved at vægte de forskellige typer, eksempelvis ud fra hvor stor en procentdel de udgør af målgruppen, eller ud fra hvor meget de kommer til at bruge systemet, kan man drage beslutninger om, hvorledes systemet skal designes.

Der ligger dog en problemstilling i dette. Hvis de arketyper man udvælger *ikke* er repræsentative, vil man ikke ud fra de opstillede karakteristika opfylde alle delta- gers krav i et design af systemet. For at undgå dette bør man opbygge arketyperne iterativt, og således kontinuerligt validere og uddybe typerne [Kujala og Kauppi- nen, 2004]. Man skal også holde sig for øje, hvor stor en del af målgruppen arke- typen repræsenterer, for at undgå krav som størstedelen af målgruppen måske ikke har brug for, eller som i værste fald kan have en negativ indflydelse på resten af målgruppen.

På grund af ovenstående problemstilling vælges i dette projekt en detaljeret initial opdeling af typerne i målgruppen. Dette gøres for så vidt muligt at undgå, at re- sultaterne ikke afspejler en enkelt type i målgruppen. Denne detaljerede opdeling sikrer, at fokus hverken ligger på brugerne eller udviklerne. Der forsøges at undgå, at ingen af de enkelte types krav bliver klassificeret vigtigere end andre, for der- ved at undgå en skævvridning af resultaterne fra spørgeskemaundersøgelsen og det skriftlige interview.

4.1.1 Metode til arketyper

Kujala og Kauppinen [2004] har på baggrund af metoder udviklet af Hackos og Redish [1998] samt Noyes og Baber [1999] udviklet en metode, der både inde- holder identifikation og udvælgelse af repræsentative brugere, som i dette projekt defineres som arketyper. Kujala og Kauppinen [2004] har opbygget metoden af de følgende 5 trin:

1. *“Brainstorm a preliminary list of users*
2. *Describe the main user characteristics (including market size).*
3. *Describe main user groups and prioritize them*
4. *Select typical and representative users from the groups.*
5. *Gather information from the users and redesign the user group descriptions according to the new information gathered.”*

[Kujala og Kauppinen, 2004]

Den ovennævnte metodes proces med at identificere og udvælge arketyper, skal ifølge Kujala og Kauppinen [2004, s. 298] foregå iterativ. Metoden ligger op til, at der skal indsamles ny information om typerne for herefter et redesigne dem.

Vi vælger at tage udgangspunkt i den ovenstående metode og modificere denne til vores projekt, da der er aspekter ved ovenstående metode, vi ikke mener er

velegnede til vores projekt. Disse aspekter er beskrevet i gennemgangen af vores metode.

Vores metode til at definere og udvælge arketyper samt beskrivelse af deres krav, består af de følgende trin:

1. Beskrivelse af initiale typer på baggrund af eksisterende undersøgelser
2. Udvælgelse og beskrivelse af arketyper på baggrund af empiri
3. Indsamle information og redesign af beskrivelse af arketyper

Ligesom Kujala og Kauppinen [2004] metode, bygger metoden i dette projekt også på en iterativ proces. Processen i dette projekt går igennem tre iterationer, hvor forskellig empiri er baggrund for hver iteration. Den første iteration bygger på en undersøgelse af typer i Open Source projekter, anden iteration bygger på en spørgeskemaundersøgelse, og sidste iteration bygger også på spørgeskemaundersøgelse, men med supplerende empiri fra et skriftligt interview.

4.1.2 Beskrivelse af initiale typer på baggrund af eksisterende undersøgelser

Det første trin i Kujala og Kauppinen [2004] metode er at opstille en foreløbig liste af brugere på baggrund af en brainstorm. Ifølge Kujala og Kauppinen [2004, s. 298] skal denne brainstorm laves ud fra viden om organisationen, f.eks. viden omkring forskelligheder blandt de forskellige typer i organisationen. Kujala og Kauppinen [2004] konkluderer dog senere at:

“Thus, it may not be trustworthy to brainstorm and describe user groups...” [Kujala og Kauppinen, 2004, s. 302]

Kujala og Kauppinen [2004] beskriver også, at man bør tage udgangspunkt i “field studies” eller anden empiri for at beskrive sine typer. Endvidere nævner Kujala og Kauppinen [2004], at der kan indsamles forskellige typer af information, som eksempelvis statistikker og markedsanalyser, hvilket kan bidrage til at udforme arketyper. Vi vælger at basere vores initiale typer på eksisterende empiri lavet ud fra feltstudier (se initiale arketyper afsnit 5.1.1) frem for en brainstorm. Vi mener, at dette giver et solidt fundament til videreudvikling af typerne i stedet for en brainstorm af ideer, der vurderes subjektivt. De næste to iterationer handler om denne videreudvikling af de initiale typer.

4.1.3 Udvælgelse og beskrivelse af arketyper på baggrund af empiri

På dette trin, som er anden iteration i vores metode, udvælges og beskrives karakteristika for hver af de initiale typer. Dette danner grundlaget for valg af arketyper.

Dette trin svarer til trin tre i den metode Kujala og Kauppinen [2004] har opstillet. Kujala og Kauppinen [2004] har opstillet forslag til karakteristika, eksempelvis alder, køn og uddannelse. Kujala og Kauppinen [2004] deler disse karakteristika ind i forskellige grupper, eksempelvis: “Personal-”, “Task related-” og “Geographic-” and “Social characteristics”. Det er vigtigt at vælge hvilke karakteristika, der skal bruges til at differentiere arketyperne med. Kujala og Kauppinen [2004, s. 299] beskriver, at den form for karakteristika man skal medtage, afhænger af produktet eller typen af system. Som eksempel bruger de:

“However, the personal characteristics are most relevant in consumer products, which are used during leisure time.” [Kujala og Kauppinen, 2004, s. 299].

Vi vælger at fokusere på arbejdsrelaterede karakteristika, da projektets mål er at undersøge krav til et system, som har til formål at effektivisere informationssøgning og herved en arbejdsrelateret opgave, i Open Source projekter. Ifølge Kujala og Kauppinen [2004] skal disse karakteristika hjælpe med til at identificere de grupper, som de forskellige brugere danner. For at identificere hvorledes deltagerne forholder sig til informationssøgning, awareness samt group memory, udsendes et spørgeskema (se afsnit 6.1) til en række Open Source projekter. Analyse af data fra spørgeskemaundersøgelsen benyttes til at identificere og opbygge en beskrivelse af Arketyperne (se bestemmelse af arketyper afsnit 7.1). Arketyperne identificeres ved hjælp af beskrivelsen af de initiale typer (se målgruppebeskrivelse afsnit 5.1), samt de områder hvor de initiale typer deler flere karakteristika. De deltagere der deler flere karakteristika, grupperes i én arketype. Arketyperne beskrives ud fra de karakteristika, hvor de differentierer sig. I næste trin valideres arketyperne ud fra et interview, hvorefter de bliver beskrevet hver især, ud fra hvilke svar de har tilkendegivet i både spørgeskemaundersøgelsen, og i det skriftlige interview.

4.1.4 Indsamle information og redesign af beskrivelse, af arketyper

Den sidste iteration går i dybden med de udvalgte typer ved at indsamle information igennem et interview (se afsnit 8.1). Denne tredje iteration svarer til trin fire og fem i den metode Kujala og Kauppinen [2004] har opstillet. Vi udvælger de deltagere, som skal benyttes til interview, hvor deltagerne skal være repræsentative i forhold til den arketype, de er grupperet under. Dette svarer til trin 4 i metoden af Kujala og Kauppinen [2004]. Herefter samler vi information igennem et skriftligt interview, og redesigner arketyperne, hvilket svarer til trin fem.

Interviewet skal være med til at validere beskrivelsen af arketyperne. Dette betyder, at hvis resultaterne fra interviewet indikerer, at den opdeling der er nået frem til igennem spørgeskemaet, ikke er korrekt, så skal arketyperne redesignes. Hvis resultaterne fra interviewet stemme overens med de resultater, der er opnået igennem spørgeskemaet, kan disse bruges til at give en dybere beskrivelse af arketyperne.

Interviewet skal ses som en kvalitativ dataindsamlingsmetode, som bruges til at uddybe hypoteser, som ikke er besvaret gennem spørgeskemaet, samt til at uddybe de karakteristika, hvor arketyperne har ligheder eller differentierer sig. Til interviewet udvælges enkelte af respondenterne fra spørgeskemaundersøgelsen på baggrund af de svar de har givet.

Beskrivelsen af arketyper består af karakteristika, som denne sidste iteration identificerer, og som herefter analyseres over til en liste af krav til vores system. Dette er samtidig afslutningen på projektets første del, som har til opgave at liste deltagere i Open Source projekters krav til vores system.

4.2 Design

Denne del af projektet har til formål at udarbejde designkrav ud fra de opstillede karakteristika for arketyperne. For at udarbejde designkrav ud fra arketypernes karakteristika, vælges det at supplere med en analyse af eksisterende systemer som skal bidrage med design ideer til systemet. Analysen af eksisterende systemer indeholder en diskussion af positive og negative aspekter ved eksisterende systemer, der kan anvendes til udarbejdelsen af vores designkrav. Ud fra arketypernes karakteristika og analysen af eksisterende systemer opstilles således en række designkrav til systemet. I beskrivelsen af designkravene diskuteres også, hvorledes arketypernes eventuelle modstridende krav løses.

Målgruppe 5

Dette kapitel dækker trin 1, “Beskrivelse af initiale typer på baggrund af eksisterende undersøgelse”, i specificeringen af arketyper i forhold til vores metode. De initiale arketyper beskrives, og der opstilles en række kriterier til udvælgelsen af de Open Source projekter, som skal indgå i empiriindsamlingen. Yderligere beskrives disse udvalgte projekter og hvorledes de opfylder kriterierne.

5.1 Målgruppebeskrivelse

Som beskrevet i problemafgrænsningen (se afsnit 3.2.1) henvender vores system sig til aktive deltagere i Open Source projekter. I empiriindsamlingen, både spørgeskemaet og interviewet, henvender vi os også til disse deltagere, idet vi således får information fra potentielle brugere. Vi vælger at arbejde med en mere detaljeret opdeling, end bruger og udviklere, til empiriindsamlingen. Dette valg er foretaget ud fra artiklen af Ye og Kishida [2003, s. 419-420], som beskriver at typerne typisk i Open Source projekter ikke er skarpt opdelt mellem udviklere og brugere, som det ofte er i proprietær software udvikling. I Open Source er der en flydende overgang mellem bruger og udvikler [Ye og Kishida, 2003, s. 419-420]. Ved derfor at vælge en detaljeret opdeling præciseres de enkelte deltagere. Hvilken type en deltager er i et Open Source projekt, afhænger direkte af, hvor meget denne bidrager til projektet. Deltagerne kan selv påvirke hvilken type af deltager de er, ved eksempelvis at blive mere involveret i projektet. I det følgende beskrives denne opdeling.

5.1.1 Initiale arketyper

Som der beskrives i metoden (se kapitel 4), vælges der i dette projekt at bruge en fordefineret opdeling af typer til brug i dette projekts undersøgelser. De initiale typer som vi vælger at arbejde med, er ud fra en undersøgelse beskrevet i en ar-

tikel af Ye og Kishida [2003]. I stedet for at benytte en brainstorm til den initiale opdeling af typerne (se metoden kapitel 4), vælges denne opdeling af typerne, da det antages, at det giver et solidt fundament til at videreudvikle de ønskede arketyper på. Ye og Kishida [2003] opdeling er baseret på en undersøgelse af fire Open Source projekter, som er beskrevet i artiklen af Nakakoji et al. [2002]. Nakakoji et al. [2002] har gennem analyse af forskelligt empiri, som eksempelvis CVS, har fundet frem til en opdeling af deltagere i Open Source projekter, til otte forskellige typer. Argumentet for specifikt at tage udgangspunkt i denne undersøgelse, er pga. måden Ye og Kishida [2003] adskiller typerne på. Ye og Kishida [2003, s. 420] adskiller typer ud fra, hvor meget indflydelse disse har i projektet. Hvor meget indflydelse deltagerne har, hænger sammen med hvilke arbejdsopgaver de varetager i projektet, som Nakakoji et al. [2002] f.eks. har fundet igennem analyse af CVS loggen.

Ye og Kishida [2003] opdeling af typer går fra de brugere som ikke er aktive i projektet, til den person som leder og styre projektets evolution. Resultaterne af Ye og Kishida [2003] undersøgelse er en klassificering af deltagere i Open Source projekter i 8 kategorier. Denne opdeling vurderes til at have en bredere tilgang til differentieringen af deltagerne i Open Source projekter, end f.eks. den opdeling Feller og Fitzgerald [2002] benytter. Feller og Fitzgerald [2002, s. 107] deler deltagerne i Open Source domænet, ind i fire grupper. Opdelingen skal ikke ses som en opdeling i typer i samme forstand som den Ye og Kishida [2003] opstiller, men derimod som involverede parter i et Open Source projekt. Gruppering som Feller og Fitzgerald [2002, s. 107] arbejder med er; *developers, user communities, commercial organizations* og *non-commercial organizations*. Den måde som de opdeler deltagerne på, vurderes ikke som hensigtsmæssig til vores undersøgelse. Begrundelsen for dette bunder i, at den opdeling af grupperne som Feller og Fitzgerald [2002] har, ikke er beskrevet som gensidig afhængige. Det vil sige, at grupperne kan overlape hinanden ved f.eks. at en deltager både kan tilhøre gruppen user, som bruger af systemet, samt være en del af gruppen developers, som udvikler på systemet. Vi vælger i vores undersøgelse at differentiere mellem bruger af systemet og udvikler af systemet, hvilket gør, at opdelingen som Ye og Kishida [2003] benytter kan bruges. Von Krogh et al. [2003, s. 1220] opdeler også typerne med fokus på at differentiere mellem udviklere og bruger typer. Denne opdeling er som følgende; *list participants, joiner, newcomer* og *developer*, og vurderes også ud fra, hvordan man bidrager til et Open Source projekt, og derved opnår forskellige roller. Udover Feller og Fitzgerald [2002], benytter Von Krogh et al. [2003] som beskrevet en klassificering af deltagerne ud fra fire kategorier, hvor Ye og Kishida [2003] benytter en opdeling af otte typer. Dette bidrager også til at opdelingen af Ye og Kishida [2003] vælges. Kujala og Kauppinen [2004, s. 302] (artikel der ligger til grund for vores metode, se kapitel 4) konkluderer også, at det er vigtigt at have en bred opdeling i starten af analysen, da der er en tendens til at undervurdere brugergruppens mangfoldighed:

“... practitioners tend to underestimate the diversity of user groups in the beginning” [Kujala og Kauppinen, 2004, s. 302]

Kujala og Kauppinen [2004, s. 299] beskriver også, at de karakteristika man bør fokusere på i beskrivelsen af ens målgruppe, afhænger af typen af system [Kujala og Kauppinen, 2004, s. 299]. I dette projekt beskrives typerne ud fra deres arbejdsopgaver, idet vores system skal omhandle informationssøgning og altså arbejdsopgaver.

Det følgende er en beskrivelse af Ye og Kishida [2003] 8 initiale typer:

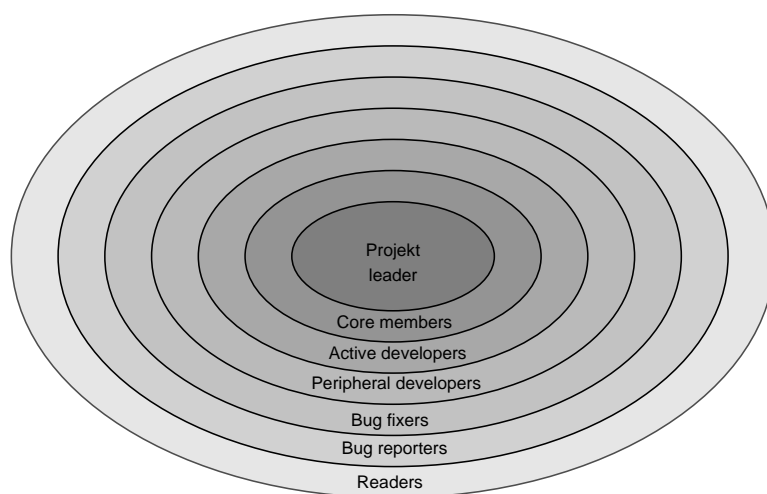
- **Project Leader:** Person der styrer hvilken retning det pågældende projektet skal gå i, og oftest også den person som startede projektet og opstillede visionerne for projektet.
- **Core Member:** Styrer og guider udviklingen af projektet. Core members er ofte dem, som har deltaget aktivt i projektet i længere tid, og som har bidraget betydeligt til udviklingen af systemet. Betegnelsen *maintainers* bruges også om denne gruppe.
- **Active Developer:** Bidrager regelmæssigt med nye funktioner og bug fixes, og er en af de største bidragsydere til projektet.
- **Peripheral Developer:** Bidrager af og til med ny funktionalitet og bug fixes. Men primært med bug fixes, så forskellen mellem peripheral developer og bug fixer kan være meget lille.
- **Bug Fixer:** Retter bugs de selv har fundet, eller som andre har beskrevet. For at gøre dette behøver de kun sætte sig ind i det kode som bug'en berører.
- **Bug Reporter:** Disse deltagere anvender systemer, og når de finder fejl, laver de en bug report. De retter ikke selv fejlen og kigger oftest derfor heller ikke i koden.
- **Reader:** Er aktive brugere af systemet. Ud over dette forsøger disse også at forstå, hvorledes systemet fungerer ved at læse kildekoden. Readerne gør dette eksempelvis for at lære at programmere, andre gør det for få en forståelse af systemet, som de således kan benytte til implementation af et andet system.
- **Passive User:** En bruger som anvender systemet, men ikke benytter projektets informationskanler.

[Ye og Kishida, 2003, s. 420]

Disse typer er ikke alle nødvendigvis repræsenteret i ethvert Open Source projekt [Ye og Kishida, 2003]. Ligeledes er fordeling af typerne heller ikke den samme i

alle projekter. Ye og Kishida [2003, s. 420] nævner i den forbindelse, at passive users udgør op til 99% af alle brugere i Apache projektet. Der er også et skarpt fald fra readers til core members, idet de fleste Open Source projekter er udviklet af få udviklere [Ye og Kishida, 2003, s. 420]. Betegnelserne opstillet af Ye og Kishida [2003] er heller ikke nødvendigvis de samme som Open Source projekterne benytter selv, eksempelvis som beskrevet under core member, kan maintainers også bruges til at beskrive denne type.

De typer af deltagere som spørgeskemaet og interviewet henvender sig til, er de syv første. Grunden til, at deltagere af typen passive user ikke er med i målgruppen er, at denne type ikke anvender informationskanalerne i Open Source projekterne, og derfor ikke er del af vores målgruppe.

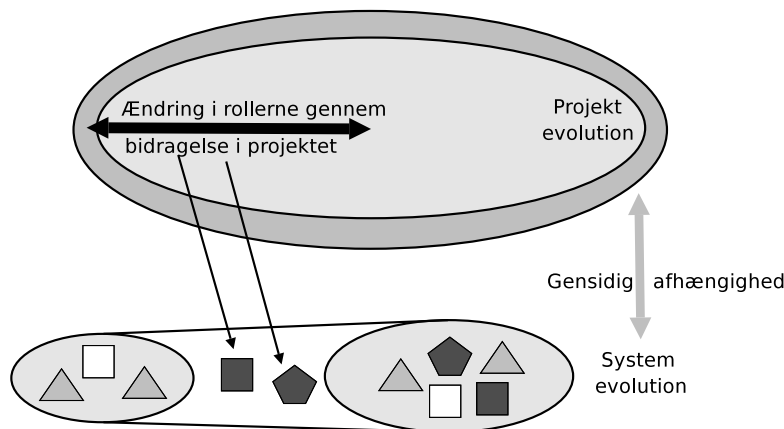


Figur 5.1: Figur over de initiale deltagertyper i et Open Source projekt.

Figur 5.1 viser en oversigt over deltagerne i et Open Source projekt. Figuren har oprindelse i artiklen af Ye og Kishida [2003, s. 421], men er tilpasset vores målgruppe, ved at typen passive user ikke er inkluderet. Den måde hvorpå de forskellige roller opstår, er igennem deltageres personlige interesser i at bruge og/eller udvikle systemet. Derfor er det op til deltagerne selv og ud fra deres interesse i projektet, at de påtager sig forskellige roller, og ikke ved at rollen bliver tildelt af andre [Ye og Kishida, 2003, s. 420]. Figur 5.1 illustrere der forskellige roller, hvor den rolle der er tættest mod midten, har den største radius af indflydelse.

De roller og den associerede indflydelse deltager påtager sig i Open Source projekter, bliver realiseret igennem bidrag til projektet. Derved er rollerne ikke fast defineret, og deltagerne kan påtage sig større eller mindre roller alt efter deres ambitions niveau. Dette gør, at rollerne i Open Source projekter bliver flydende. Dette er illustreret i figur 5.2, hvor man kan se, at rollerne ændres i takt med, at deltagerne bidrager til projekterne. Figuren er modificeret i forhold til den originale figur fra Ye og Kishida [2003], hvor der kun er placeret en pil ind mod midten. Pilen kan gå

begge veje, da rollerne kan ændres i begge retninger. Eksempelvis kan en “Active developer” af forskellige årsager, være nødsaget til at bidrage mindre til projektet, og derfor vil hans rolle ændre sig ud mod den yderste ring. Figuren viser også, at systemet i et Open Source projekt og dets omgivelser er gensidigt afhængige af hinanden, hvilket sker igennem bidrag til projektet af de forskellige roller.



Figur 5.2: Illustration over hvordan roller i Open Source projekter skifter, når deltagerne bidrager til projektet

At rollerne er flydende i Open Source projekter, som beskrevet ovenfor, giver endnu et argument for, at systemet skal fungere for alle slags deltagere i projektet. Som beskrevet i den initierende problemstilling opstiller vi nogle krav til et system, der skal hjælpe med at forbedre samarbejdet i Open Source projekter. Yderligere har vi argumenteret for, at både brugere og udviklere samarbejder i et projekt. Dette giver endnu et argument for at starte med en bred defineret opdeling af de initiale typer for at få betydningen af et dynamisk skiftende miljø med i undersøgelsen.

Efter denne gennemgang af valg og definering af initiale arketyper, defineres i det følgende kapitel en række kriterier for de Open Source projekter, som indgår i empiriindsamlingen.

5.2 Kriterier for udvælgelse af projekter

Til at udvælge de projekter, vi ønsker der skal deltage i spørgeskemaundersøgelsen samt interviewet, opstilles der i det følgende en række kriterier, som projekterne skal opfylde. Målet med dette er at sikre, at de udvalgte projekter falder inden for målgruppen (se problemformulering, afsnit 3.2).

Projekterne vi vælger skal være Open Source projekter, da det er i forbindelse med disse, at vi arbejder med et CSCW system. Yderligere er det et kriterium at projekterne skal udvikles distribueret.

Et andet kriterium er, at projekterne skal have en vis størrelse og aktivitet. Ved størrelse menes både antallet af deltagere og antallet af informationskanaler i projektet,

samt mængden af aktivitet i projektet. Kriteriet bygger på vores tredje hypotese (se afsnit 3.1) om, at komplekse projekter har flere problemer med informationssøgning. Dette valg betyder dog, at vi ikke kan sammenligne store projekter med små projekter, som kunne sige noget om problemstillingen kun gælder i store projekter.

Et kriterium der også skal gælde er, at projekterne skal have flere officielle informationskanaler (se afsnit 2.2 omkring informationskanaler) tilknyttet. Hvis det er muligt, skal projekterne også have forskellige informationskanaler. Som tabel 5.2 viser, er der en lille spredning i, hvilke informationkanaler de valgte projekter anvender. Med dette kriterium forsøger vi at differentiere de forskellige informationskanaler ved f.eks. at se, om der er nogle, der er mere anvendelige til visse formål end andre.

For at overholde kravet omkring at der skal være officielle informationskanaler tilknyttet, er det nødvendigt, at Open Source projekterne har deres eget website, hvorfra de henviser til de informationskanaler som de benytter.

Yderligere er et kriterium for udvælgelsen af projekterne, at systemet bruges af en bred gruppe af deltagere, så der er en stor mulighed for at projektet har alle slags deltagere repræsenteret. Når vi opstiller dette kriterium, skyldes det, at målgruppen (se afsnit 5.1) både indeholder udviklere og brugere.

5.3 Beskrivelse af projekterne

Vi har udvalgt følgende tre Open Source projekter: *Gaim*, *MPlayer* og *The Gimp*. Alle tre projekter opfylder de kriterier, der er opstillet i det forgående afsnit. Hvorledes de opfylder disse kriterier gennemgås i det følgende.

Det første kriterium som projekterne skal opfylde er, at det er Open Source. Dette kriterium er opfyldt af alle projekterne, idet de er udgivet under license som er godkendt af Open Source Initiative [2005], se tabel 5.1. Kriteriet, der omhandler at Open Source projekterne skal udvikles distribueret, kan vi ikke sige om er opfyldt, før spørgeskemaets spørgsmål 3 er besvaret.

Gaim	GNU Public License
MPlayer	Lesser GNU Public License
The Gimp	GNU Public License

Tabel 5.1: Licenserne som Open Source projekterne er udgivet under.

Kriteriumet om aktiviteten er svært at vurdere, da det er svært at få overblik over communitiets størrelse på grund af, at informationen er spredt ud på forskellige informationskanaler. Ved at kigge på Sourceforge.net [2005] og Freshmeat.net [2005] kan vi se, at Open Source projekterne er meget populære. Ligeledes kan vi se, at projektet Gaim, som ligger på Sourceforge.net [2005], har meget aktivitet i

	Gaim	MPlayer	The Gimp
Website	×	×	×
Forum	×		
Nyhedsgrupper			(×)
Mailing Lists	×	×	×
Chat	×	×	(×)
Versionsstyring	×	×	×
Bug- og Issuetracking	×	×	×

Tabel 5.2: Tabel over informationskanaler i Open Source projekterne. Ved felter markeret med (×) ses beskrivelse af: Problematik omkring informationskanaler.

deres informationskanaler. F.eks. har Gaim fået tilføjet 148 bugs i deres bugtracking system i perioden fra 1. til 20. april 2005 [SourceForge.net, 2005]. The Gimp og MPlayer, som ikke er placeret på Sourceforge.net [2005], har ligeledes meget aktivitet i deres informationskanaler. I perioden fra 1. til 14. marts 2005 er der på MPlayers mailing liste, *mplayer-users*, sendt 325 e-mails. Også projektet The Gimp har meget aktivitet i deres community. Idet der er tilføjet 190 nye bugs til projektets bugtrackingsystem i perioden fra 1. til 20. april 2005.

Et andet kriterium som projekterne skal opfylde er, at der skal være officielle informationskanaler tilknyttet projekterne. Projekternes officielle informationskanaler, kan ses i tabel 5.2.

Projekterne som er udvalgt er alle programmer som er alment brugt, og falder de derfor ind under dette kriterium. Programmet Gaim er en Instant Messaging client, MPlayer er en multimedia afspiller, og The Gimp er et billedredigeringsprogram.

5.4 Problematik omkring informationskanaler

At deltagerne i Open Source projekterne kan opleve det problematisk at finde den information de søger, kan bekræftes ud fra de erfaringer som opnås i dette projekts undersøgelse af de tre Open Source projekter, Gaim, The Gimp og MPlayer. Erfaringen viser, at det ikke altid er lige let at gennemskue, hvilke informationskanaler projekterne tilbyder.

Eksempelvis har der i The Gimp projektet været problemer med at finde ud af, hvilke informationskanaler projektet gør brug af. Som vist i tabel 5.2, linker The Gimp projektet ikke officielt til Chat og Nyhedsgrupper.

The Gimp projektet linker på deres hjemmeside til <http://ircd.gimp.org>. Linket kalder de: "The Official site of GIMPNet", og siden indeholder ikke noget information om dette link, og det fremgår ikke tydeligt om det er en irc-kanal. Vi har selv forsøgt at tilgå serveren ircd.gimp.org gennem en irc-client, og på denne

måde kan vi se, at det er en irc-kanal. Søger man f.eks. på Google finder man så frem til wiki.gimp.org/gimp/irc, som er et subdomæne. Dette domæne linkes ikke til gennem hovedsiden, gimp.org, og kan derfor være svær at finde, hvis man ikke kender til den. Denne side, wiki.gimp.org, indeholder information om både irc-kanaler og nyhedsgrupper, som The Gimp projektet gør brug af.

Som der beskrives i afsnittet omkring informationskanaler (se afsnit 2.2), vælges der i dette projekt, at undersøgelserne kun medtager de officielle informationskanaler, som defineres ved at de er linket direkte til fra Open Source projekternes websites. Et problem der opstår ved dette er, at et projekt som f.eks. The Gimp også bruger subdomænes, f.eks. wiki.gimp.org. The Gimp projektet beskriver, at de gør brug af mailing lister, bugtrackers og versions, men ikke noget om chat og nyhedsgrupper.

Hvis projekterne ikke selv tydeliggør, hvordan og hvor forskellig informationen formidles, som eksemplet ovenfor med The Gimp, må man antage, at det også er svært for deltagerne i projektet at finde ud af, hvor man skal søge en specifik information. Denne problemstilling giver endnu en anledning til vores undersøgelse og til at opstille designkrav til et CSCW system. Hvis en deltager kun skal søge et sted, og det er op til projektet at angive hvilke informationskanaler som systemet skal søge i, behøver brugeren ikke have viden om, hvilke kanaler søgningen foretages i, så længe værktøjet søger i de informationskanaler, projektet stiller til rådighed.

Spørgeskema

Dette kapitel beskriver, hvorledes spørgeskemaet er opbygget. Herunder formålet med de enkelte spørgsmål og relationen til hypoteserne. Spørgeskemaet bidrager til at differentiere arketyperne. Herefter følger en opsamling på hvordan og hvilke hypoteser, der valideres gennem spørgeskemaet. Svarene fra spørgeskemaet er at finde i bilag H, og søjlediagrammer over dem er at finde i bilag D.

6.1 Beskrivelse af spørgeskemaet

Spørgeskemaet benyttes til både at få valideret hypoteserne samt at få defineret en opdeling af deltagertyper. Det skal også danne grundlaget for beskrivelsen af arketyperne og designkrav til vores system. Spørgsmålene i spørgeskemaet tager udgangspunkt i relationen mellem de initiale typer samt vores hypoteser. Vi vil belyse typerne igennem de problemer de oplever, og hvorledes de udfører informationssøgning, og derfor tages udgangspunkt i hypoteserne. Hypoteserne omhandler problemer, vi mener, deltagerne oplever, og disse belyses igennem spørgeskemaet. Ved at validere hypoteserne igennem spørgeskemaet undersøges om deltagerne oplever disse problemer. Hvis dette er tilfældet bygges videre på hypoteserne og således indgår de i processen til at finde designkrav.

Spørgeskemaet henvender sig til vores målgruppe (se afsnit 5.1). Målet er at få fat på de syv forskellige typer af deltagere i de udvalgte projekter. Mailen, der udsendes, er derfor målrettet alle deltagere og postet flere steder, i håb om at flest muligt ser linket til spørgeskemaet. I bilag A kan mailen samt hvor den blev udsendt ses.

Spørgeskemaet tager udgangspunkt i det specifikke projekt, som deltagerne er involveret i. Dette er valgt, da det er problemstillinger mellem deltagerne og det pågældende projekt, der er interessant. Dette betyder dog, at hvis deltagerne har erfaring med informationssøgning fra andre Open Source projekter, fremgår det ikke af besvarelsen af spørgeskemaet.

Følgende beskriver de områder spørgeskemaet omhandler. Under hvert område er spørgsmål til det pågældende område opstillet. Formålet med spørgsmålene forklares samt hvorledes de bidrager til at validere hypoteserne (se afsnit 3.1.3). Spørgeskemaet er også at finde i bilag B.

Til mange af spørgsmålene i spørgeskemaet følger et "Why"-spørgsmål. Formålet med disse er at afdække hvilken motivation, der ligger bag svaret af det foregående spørgsmål.

Da det er i Open Source projekter, systemet skal anvendes, er der en afgrænset mængde af forskellige informationskanaler, som alle er internet baseret kommunikations teknologier. I denne rapport er informationskanalerne begrænset til dem, der anvendes i de Open Source projekter, der bruges som cases. Til de spørgsmål, hvor deltagerne kan vælge mellem forskellige informationskanaler, listes de informationskanaler, der er beskrevet i afsnit 2.2. Weblogs er fravalgt da ingen af projekterne anvender denne informationskanal. På listen over mulige informationskanaler er det desuden muligt for deltagerne at udfylde et tekstfelt, hvis de bruger en informationskanal som ikke er på listen f.eks. Google.

Inden spørgeskemaet udsendes, udføres et pilotforsøg. Resultaterne af forsøget findes i bilag C. Feedback fra forsøgspersonerne benyttes til at rette fejl samt formuleringer i spørgeskemaet, inden det sendes ud til de udvalgte projekter.

6.1.1 Introduktion

Spørgeskemaet indleder med en forklaring af formålet med spørgeskemaet, samt at alle svar vil blive behandlet fortroligt.

Indledningen lyder:

"The purpose of this survey is to study how people search and find information in different open source projects.

All answers will be kept confidential, only group members from the university project group will have access to them.

Continue on to the survey

(If you have any questions write us)"

6.1.2 Personligt

De første spørgsmål i skemaet angår deltagerens personlige information. Der spørges ind til dette for at få en idé om hvem deltageren er, om projektet udvikles distribueret, og om deltageren må kontaktes til et uddybende interview.

1. Age

2. *Gender*

- *Male*
- *Female*

3. *Where are you from?*

- *(liste af lande)*

Formål: At afgøre om projektet er distribueret.

4. *E-mail address (Only supply your e-mail address if we may contact you later, to conduct a e-mail interview)*

Formål: Deltagerne kan oplyse deres e-mail, hvis vi må kontakte dem med henblik på det uddybende interview.

5. *Where did you first hear about this survey?*

- *Mailing list*
- *Forum*
- *Chat (e.g. IRC)*
- *Homepage*
- *Other*

Formål: Svaret på dette spørgsmål kan indikere, hvilke informationskanaler deltagerne søger information i.

6.1.3 Relation til projekt

Disse spørgsmål handler om deltagerens relation til projektet; hvor længe og hvor meget deltageren er involveret i projektet. Formålet er at finde ud af, hvilken type deltager, som har besvaret spørgeskemaet. Dette benyttes til at opdele deltagerne efter, hvilken arketype de tilhører. Svarene fra spørgsmål 7 og 8 anvendes til at validere sidste del af hypotese 3 (se afsnit 3.1), som afhænger af, hvor længe deltageren har været involveret i det pågældende projekt.

6. *In which way are you involved in this project?*

- *Svarmulighederne er de 8 typer af deltagere, som er beskrevet i afsnittet Målgruppebeskrivelse 5.1.*

Formål: Ved at stille dette spørgsmål, opnås at der kan skelnes mellem typen af deltagerne. Dette benyttes både til at opdele deltagerne i arketyper, men også til at besvare hypoteser. Oplysningen om deltagertype er en nødvendighed for at kunne besvare både hypotese 3 og 2.

7. *How long time have you been using this project?*

- *less than 1/2 a year*
- *between 1/2 and 1 year*
- *between 1 and 2 years*
- *between 2 and 3 years*
- *more than 3 years*

Formål: Ved at få svar på dette spørgsmål kan deltagerne yderligere differentieres efter, hvor længe de har været involveret i projekterne. Intervallet er valgt ud fra en vurdering af, hvornår en deltager er ny, og hvornår en deltager har deltaget længe i et projekt.

8. *How long have you been reading/posting on this project's...*

(1: never, 2: less than 1 year, 3: between 1 and 2 years, 4: between 2 and 3 years, 5: more than 3 years)

- *Forum*
- *Mailinglist*
- *Chat (e.g. IRC)*
- *Newsgroup*
- *Code repository (e.g. CVS)*
- *Bug and issuetracker (e.g. Bugzilla)*

9. *How often do you read/post on this project's...*

(1: never, 2: monthly, 3: weekly, 4: daily, 5: several times daily)

- *Forum*
- *Mailinglist*
- *Chat (e.g. IRC)*
- *Newsgroup*
- *Code repository (e.g. CVS)*
- *Bug and issuetracker (e.g. Bugzilla)*

Formål: De to spørgsmål skal afdække, hvilke informationskanaler i det pågældende projekt den enkelte deltager bruger, og hvor ofte denne bruger dem. Svaret til disse to spørgsmål giver sammen med spørgsmål 7 et detaljeret billede af, hvor erfaren og aktiv deltageren er.

6.1.4 Informationssøgning

Den følgende gruppe af spørgsmål skal afdække, hvilke informationskanaler deltagerne anvender, hvilken information de søger, og hvilke problemer de oplever ved informations søgning.

10. *When you are searching for information, where do you look first?*

- *Forum*
- *Mailing list*
- *Chat (e.g. IRC)*
- *Newsgroup*
- *Code repository (e.g. CVS)*
- *Bug and issuetracker (e.g. Bugzilla)*
- *Other*

11. *If you did not find the information the first place you looked, where would you then look?*

- *Forum*
- *Mailing list*
- *Chat (e.g. IRC)*
- *Newsgroup*
- *Code repository (e.g. CVS)*
- *Bug and issuetracker (e.g. Bugzilla)*
- *Other*

Formål: At belyse hvilken informationskanal forskellige typer af deltagere foretrækker som deres primære informationskanal, samt hvilken informationskanal de herefter foretrækker. Hvis svarene fra de to spørgsmål holdes sammen med typen af deltager, kan det give overblik over, hvem som bruger hvilke kanaler. Resultatet benyttes til at validere hypotese 1.

12. *Do you experience that you have to search several sources (e.g. mailing lists, news groups and so on) to find the information you are looking for?*

- *Yes*
- *No*

Formål: Svaret kan indikere om deltagerne oplever problemer med at finde information, og derfor bliver nødt til at søge i forskellige informationskanaler.

13. *When you are searching for information, which sources do you search: "official" sources (e.g. directly related to the project) or 3. party (e.g. Google, a general news group, or another person) or both?*

- *Official*
- *3. Party*

Formål: Spørgsmålet kan belyse, hvorvidt deltagerne foretrækker projektets officielle informationskanaler eller andre. Hvis deltagerne overvejende bruger 3. party, kan det indikere, at de informationskanaler projektet tilbyder ikke lever op til deltagerens krav.

14. *Why do you search in the sources you have selected above?*
15. *Do you experience it to be difficult/easy to find the information you need in this project? (1: Difficult - 5: Easy)*

- *Your experience*

Formål: Dette spørgsmål skal belyse, om deltageren har problemer med at søge information i projektet. Spørgsmålet hjælper også med at besvare første del af hypotese 3, omkring at informationssøgningen afhænger af størrelsen af det pågældende projekt.

16. *Why do you experience it to be difficult/easy to find the information?*
17. *Which two kinds of information are you usually looking for?*

Formål: At finde ud af hvilken type information deltagerne typisk leder efter. Dette kan relateres til typen af deltager og sige noget om sammenhængen mellem forskellige typer af deltagere og den type information de søger. Dette er en del af hypotese 2. Vi vælger ikke at give forslag til typer af information for ikke at fastlægge bestemte kategorier for deltageren.

18. *Where are you looking for those two kinds of information?*

Formål: Dette spørgsmål uddyber ovenstående spørgsmål, idet spørgsmålet koncentrerer sig om de informationskanaler, hvor deltagerne søger de to typer information. Ud fra relationen mellem de to spørgsmål, kan konkluderes, hvilke kanaler der benyttes til at søge efter den information, der primært søges der.

6.1.5 Awareness

Den følgende sektion af spørgsmål skal afdække deltagerens behov og ønske om at kunne se awareness information om andre deltagere i projektet.

19. *How interested are you in knowing what people... (1: Not at all - 5: Very much)*

- *currently are working on in this project?*
- *have been working on in this project?*
- *are going to work on in this project?*

Formål: At afklare hvor interesseret deltageren er i awareness information. I kombination med spørgsmålet omkring typen af deltager er det muligt, at se om forskellige typer af deltagere har et forskelligt behov for awareness. Dette skal bruges til at validere hypotese 2

20. *If you want to know who has modified an artifact (e.g. source code, documentation and so on) where do you look?*

- *Forum*
- *Mailing list*
- *Chat (e.g. IRC)*
- *Newsgroup*
- *Code repository (e.g. CVS)*
- *Bug and issuetrackers (e.g. Bugzilla)*
- *Other*

Formål: Ud fra besvarelserne til dette spørgsmål kan konkluderes, hvilke informationskanaler deltagerne benytter for at finde awareness information om, hvem der har ændret i et artefakt.

6.1.6 Erfaring

Efterfølgende er to spørgsmål, der handler om deltagerens erfaring med at finde information, og om det har ændret sig, mens de har deltaget i projektet.

21. *Did you experience it difficult/easy to find information... (1: Difficult - 5: Easy)*

- *when you joined this project?*
- *now?*

22. *Why do you experience it to be more difficult/easier now?*

Formål: Spørgsmål 21 og 22 afdækker, om der er et problem med at finde information. Deltagerne skal både svare på, hvorledes de ser det på nuværende tidspunkt, men også hvorledes det var, da de var nye på projektet. Efterfølgende er det muligt at uddybe, hvorfor de finde det nemmere eller sværere

på nuværende tidspunkt. Derved skal spørgsmålene hjælpe til at validere hypotese 3, der omhandler relationen mellem hvor længe en deltager har været i et projekt, og hvor svært denne oplever informationssøgning.

6.1.7 En- eller tovejskommunikation

De resterende spørgsmål omhandler, deltagerens ønsker for at melde tilbage eksempelvis ved at spørge eller bidrage med information.

23. *If you have the choice, do you prefer to ask a question to get the information you need, or do you prefer to search for it? (if you prefer different approaches depending on the information you are searching for, please write it in question 24)*

- *Ask a question*
- *Search information*

Formål: Dette spørgsmål handler om, hvorvidt deltageren foretrækker at stille et spørgsmål, eller selv søger efter de informationer denne ønsker. På den måde fremgår det, om deltagerne foretrækker en- eller tovejs kommunikation til at finde information.

24. *Why do you prefer the above?*

25. *If you have found the information you searched for, would you like to have the opportunity to post a followup message (e.g. you need more information, or you would like to add some information)?*

Formål: Skal vise om deltageren ønsker at kunne bidrage med information, f.eks. ved at stille uddybende spørgsmål.

26. *Why would you like to have that opportunity or not?*

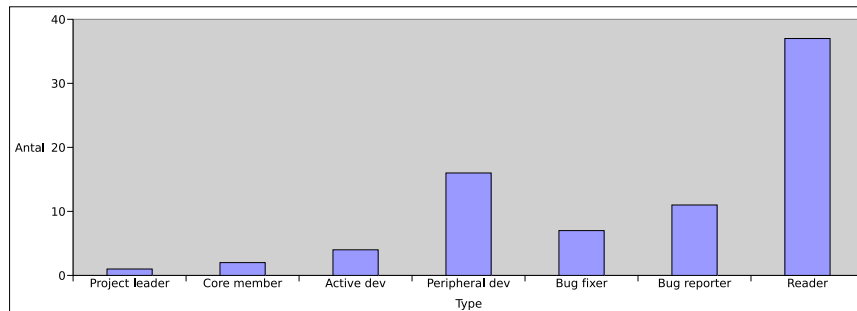
6.2 Evaluering af spørgeskema

I det følgende afsnit evalueres spørgeskemaundersøgelsen, både med hensyn til resultater og fejlkilder, samt hvorledes vi forholder os til disse.

6.2.1 Respondenter

Deltagerne er fordelt på alle syv typer af deltagere defineret i målgruppebeskrivelse (se afsnit 5.1). Dette illustreres af søjlediagrammet, se figur 6.1. Deltagernes involvering spænder fra under $\frac{1}{2}$ år til mere end tre år, og niveauet af deltagererfaring er forskellig indenfor projekterne. Disse resultater viser, at målet med at nå ud til

forskellige deltagere i projekterne er opnået. Yderligere viser det også, at der er opnået et repræsentativt udsnit af deltagere som har svaret, fra projekterne.



Figur 6.1: Fordeling af deltagere

6.2.2 Distribueret udvikling

Som det beskrives i kriterier for udvælgelsen af Open Source projekterne, opstilles det kriterium, at Open Source projekterne skal være udviklet distribueret. Dette kriterium er nu opfyldt, da spørgeskemaet viser, at deltagerne er placeret rundt i verdenen. F.eks. er de 24 respondenter fra MPlayer projektet fordelt på 12 forskellige lande.

6.2.3 Passiv user

I beskrivelse af målgruppen (se afsnit 5.1.1) fravælger vi “Passive user” som en del af vores målgruppe. Derfor er det en fejl, at passive user er en svar mulighed i spørgeskemaet i spørgsmål 6. Passive user og readers adskiller sig ved, at readers er aktive i projekterne, hvilket passive user ikke er. Dette betyder, at hvis en deltager har svaret passive user i spørgsmål 6, er denne faktisk en reader, da deltageren har fulgt med i den informationskanal, hvor linket til spørgeskemaet blev posted. Vi vælger derfor, at de deltagere der har svaret passive user, tilføjes til gruppen med readers. I resten af projektet er det de resterende syv typer der benyttes.

6.2.4 Hvad er information

Enkelte respondenter har tilkendegivet, at de er i tvivl om, hvad der menes med information (se nedenstående citat). Dette påpeges også af én af testpiloterne. For ikke at påvirke deltagerne med vores egen definition, er det fravalgt at beskrive, hvad vi mener med information (se bilag afsnit C).

“I just wanted to comment on a little area of confusion I did have with the survey. In the survey you asked how we look for information. I was unsure what you meant by these questions – what information in particular. Did you just mean information that was related to a bug report being filed, information related to an open source product (but not necessarily related to a bug or enhancement request), or information related to anything I want to find (not necessarily related to open source software).” Privat mail fra en Gaim reader

Vi kan dog se ud fra besvarelsene, at deltagerne har forstået, hvad der menes med information, se eksempelvis afsnit 7.1.1 omkring hvilken information deltagerne søger. Vi vælger derfor ikke at se dette som noget problem.

6.2.5 Ny forståelse af awareness

I afsnittet omkring awareness (se afsnit 2.3) beskrives det, at awareness opfattes som information. I spørgeskemaet er der stillet et spørgsmål (spg 19), der omhandler, hvor interesserede deltagerne er i, hvad folk; har lavet, laver nu og vil lave. Altså awareness information i forhold til hvad personer laver. I spørgsmål 17, der omhandler hvilke to typer information deltagerne typisk leder efter, har en del af respondenterne tilkendegivet, at de er interesseret i en form for awareness, der ikke er direkte forbundet til en person. Eksempler på denne form for awareness information er, hvilken retning projektet bevæger sig i, nye releases af softwaren samt hvilke nye funktioner denne indeholder. I citat er et eksempel på dette.

“Project direction. Current work being done.” 81, bug reporter, The Gimp

Diskussionen af awareness i afsnit 2.3 fokuserer på, hvad personer laver, og i generel awareness (se afsnit 2.3.3) fokuseres der på awareness informationen omkring personer (herefter betegnet person awareness).

Vi mener, at denne nye forståelse af awareness, på samme måde som person awareness, kan benyttes på både det specifikke og generelle niveau af awareness. Den fokuserer ikke på personen, men andre områder i projektet. Eksempelvis softwaren, hvor der findes awareness information omkring hvornår nye releases udkommer, se nedenstående citat:

“Bug fixes and new releases” 202, reader, Gaim

I resten af projektet bruges den nye forståelse af awareness information, som indebærer både den gamle forståelse af person awareness samt den nye forståelse af awareness.

6.2.6 For specifik spørgsmål omkring awareness

Spørgsmål 20 (If you want to know who has modified an artifact (e.g. source code, documentation and so on) where do you look?) fokuserer kun på, hvilken informationskanal deltagerne bruger, når de skal finde ud af hvem, der *har* ændret noget. Spørgsmålet burde også have belyst, om deltagerne benytter forskellige informationskanaler, alt efter om de leder efter, hvad der er blevet lavet, bliver lavet, og skal laves.

Det har den betydning, at vi kun kan sige noget om, hvilken informationskanal deltagerne benytter, når de vil vide, hvem der har ændret noget.

6.2.7 Krav til systemet

Spørgsmål 25 bidrager ikke med information til besvarelse af hypoteser. I stedet giver svaret et krav til vores system. Svaret på spørgsmålet er, at de fleste deltagere ønsker at kunne svare tilbage efter en informationssøgning. Den primære grund til dette er, at de ønsker at kunne give feedback eller yderligere information, som kan hjælpe andre, der søger samme viden. Den yderligere information kan både være uddybelse af den information, som allerede er tilgængelig eller supplerende med deltagerens egen version, hvis de er uenige i det, som står der. Resultatet giver det krav til systemet, at det skal være muligt at svare tilbage efter informationssøgninger.

6.2.8 Manglende svarmulighed

En deltager beskriver, at han mangler svarmuligheder i spørgsmål 17 (se citat nedenunder) til at kunne beskrive, hvilke typer information han leder efter. Som der er beskrevet i formålet med spørgsmål 17, vælger vi ikke at give nogle eksempler på typer af information, for ikke at fastlægge kategorier af informationer.

“out of what categories?” 138, core member, Gaim

De udvalgte projekter har ikke de samme informationskanaler tilknyttet, som eksempelvis MPlayer der ikke har et officielt forum (se tabel 5.2). Problemet er, at det i spørgeskemaet ikke er muligt at svare “ikke eksisterende” i spørgsmålene angående, hvor længe og hvor ofte de anvender de forskellige informationskanaler (spørgsmål 8 og 9). Det resulterede i, at deltagerne i MPlayer projektet har været nødsaget til at vælge “Never” i disse to spørgsmål. For at få brugbare resultater vælger vi at sortere disse svar fra, når vi laver statistikker over resultaterne på disse to spørgsmål. I bilag E giver figur E.1 en oversigt af hvilke svar er blevet fjernet.

En af deltagerne gør i en kommentar selv opmærksom på problemet. Kommentaren ses i følgende citat:

“The reason for not looking at someplace may be that that place doesn’t exist, or it could be that that place is inconvenient. Very different reasons, none show any preference because the questions and answer options are inadequate.” 153, peripheral developer, The Gimp

Selvom alle projekterne således ikke har tilknyttet alle informationskanaler, kan vi alligevel benytte søjlediagrammerne. Søjlerne vil repræsentere deltagerne, og arketyperne skal bestå af disse. Det er således ikke et problem, at nogle af informationskanalerne måske er vægtet mindre, idet ikke alle projekter har en sådan kanal tilknyttet.

En måde vi kunne have undgået problemet på, var ved at have en n/a (not available), som deltagerne kunne vælge, hvis de mente at en specifik informationskanal ikke findes i et projekt. Dette kunne f.eks. også have vist hvis X antal readers havde markeret mailing liste som n/a, så er der måske et problem i projektet, at folk ikke kan finde mailing listen.

Yderligere mangler der i flere af spørgsmålene, 8, 9, 10, 11 samt 20, “website” som valgmulighed. Spørgsmål 8 og 9 er dog specielle tilfælde, hvor vi spørger efter, hvor længe/ofte deltagerne har læst og *postet* på de forskellige informationskanaler, og da vi opfatter websites som envejs kommunikation, er denne svarmulighed fravalgt i de spørgsmål. Dog burde spørgsmålene være delt op, idet en deltager ofte godt kan læse på projekts website, men aldrig postet på denne informationskanal. Dette betyder, at vi ikke har nogle data omkring, hvor ofte deltagerne benytter websites. I spørgsmål 10, 11 og 20 har deltagerne haft mulighed for at vælge “other” og skrive website som svar, hvilket 7 deltagere også har gjort. Vi mener dog, at hvis website havde været en mulighed, ville flere havde valgt dette. Derfor vil vi tage højde for dette i analyse af de eksisterende systemer i kapitel 10.

6.2.9 Erfaring

Vores spørgeskemaundersøgelse behandler deltagerens opfattelse af informationsøgningen inden for det specifikke Open Source projekt, de er involveret i. Der bliver ikke taget højde for den påvirkning. En deltagers erfaringer fra eksempelvis deltagelse i et andet Open Source projekt kan have, på de svar denne har givet. F.eks. kan en deltager i Gaim, af typen reader, godt have været eksempelvis core member i et andet Open Source projekt. Den viden og erfaring denne deltager har fra tidligere, tager vi ikke med i vores analyse og behandling af data. Dette betyder, at en person som kun har været med i projektet et halvt år, måske svarer, at denne har rigtigt let ved at finde information. Vi kan ikke sige om, om det er fordi information er let tilgængelig eller om det er fordi denne har nogle erfaringer fra tidligere, som der trækkes på. Kort sagt, vi behandler kun situationen indenfor det konkrete projekt uden relation til deltagerens tidligere erfaringer.

Det er ligeledes derfor at vi vælger at stille spørgsmål specifikt til problemstillinger omkring informationsøgning i projektet. I det følgende afsnit evalueres dette.

6.2.10 Specifikt til projekt eller generelle spørgsmål

En deltager påpeger, at vores spørgsmål svinger mellem at være specifik til det projekt vi har udsendt spørgeskemaet til, og generelt, så dette kan have ledt til forvirring hos deltagerne.

“you seem to oscillate back and forth in this survey between my open source activity in general and activity in gaim specifically” 138, core member, Gaim

15. **Do you experience it to be difficult/easy to find the information you need in this project? (1: Difficult - 5: Easy)**

16. **Why do you experience it to difficult/easy to find the information?**

Figur 6.2: Spørgsmål 15 og 16 fra spørgeskemaet.

Dette lader dog ikke til at være et gennemgående problem, da mange af besvarelsenerne retter sig specifikt til projektet. F.eks. i spørgsmål 15, se figur 6.2, skriver vi *this project*, og da spørgsmål 16 er en opfølgning til spørgsmål 15, er det ikke unødvendigt at sige projektet igen, når vi i spørgsmål 16 henviser til “information”.

Hvis vi kigger på de svar vi har fået, så er mange direkte til det projekt, vi har sendt spørgeskemaet til. F.eks. har en deltager svaret følgende til spørgsmål 16 (Why do you experience it to difficult/easy to find the information?):

“Lots of knowledgable users. (Gimp is one of the most widely used OSS project, so there are many users, in general.) The project is also mature, so there is a lot of information.” 149, reader, The Gimp

Her henviser deltageren direkte til det projekt, denne er involveret i. Derfor kan det også konkluderes, at deltagerne har forstået spørgsmålene korrekt, med hensyn til om vi mener general eller specifik, ud fra de svar de har givet.

6.2.11 Informationskanaler skemaet er postet på

Spørgeskemaundersøgelsen er sendt ud på informationskanalerne, mailing list, fora, og chat, som Open Source projekterne gør brug af, se tabel 5.2. Vi har ikke sendt til eksempelvis nyhedsgruppen, som The Gimp bruger, da vi ikke ser den som en officiel informationskanal i projektet.

I og med at vi har postet på overstående informationskanaler, giver det også nogle relative indlysende svar flere steder i spørgeskemaet. Eksempelvis hvis en deltager

har set spørgeskemaet på mailing listen, bruger han sandsynligvis denne informationskanalen som enten den primære eller sekundære informationskanal. Desuden er spørgsmål 5 (Where did you first hear about this survey?) ikke til meget nytte. Derfor benyttes svarene fra dette spørgsmål ikke i den videre analyse.

6.2.12 Relation mellem spørgsmål 17 og 18

Formålet med spørgsmål 18 (*Where are you looking for those two kinds of information?*) er at belyse relationen mellem typen af information, og hvor denne søges. Dette er dog ikke muligt, da deltagerne i besvarelserne af spørgsmål 18 ikke har skrevet, hvilken information de søger i de forskellige informationskanaler. Enkelte deltagere har ikke besvaret spørgsmål 18, andre oplyser kun én informationskanal, mens andre oplyser flere end to informationskanaler. Derfor kan det ikke med sikkerhed siges noget om, hvor deltagerne finder de informationer, de har oplyst i spørgsmål 17. Fejlen kunne løses ved at redesigne spørgeskemaet, således at besvarelserne fra spørgsmål 17 blev brugt i 18, og deltagerne så bare skulle sætte kryds ud for, hvilken informationskanal de søger informationen i.

På grund af dette problem kan vi ikke sige noget omkring relationen mellem information og informationskanal.

6.3 Opsamling på hypoteser

I dette afsnit konkluderes hvilke og hvordan resultaterne fra spørgeskemaet kan validere dele af hypoteserne.

Hypotese 1 *Formålet med informationssøgning, hvorledes søgningen udføres, problemer angående søgning af information og hvilke informationskanaler der søges i, afhænger af deltageren.*

Af resultaterne fremgår det både hvilke former for information deltagerne søger, og hvilken type deltager, de selv karakteriserer sig som. F.eks. søger readers typisk efter information om nye features. Til sammen giver resultaterne et billede af, hvad deltagernes roller er i projektet, men ikke nødvendigvis hvad deres formål med deltagelsen er. Der er derfor behov for at spørge ind til dette i interviewet. Hypotesen nævner, at formålet afhænger af deltageren. Vi mener derfor, at dette er et punkt, der kan være med til at differentiere mellem typer af deltagere.

Ud fra svarene på spørgeskemaet fremgår det, hvorledes deltagerne søger information, samt hvilke informationskanaler de søger i. Det fremgår ikke, hvilke problemer de finder i informationssøgning, og dette er således endnu et område, der skal afdækkes ved hjælp af interviewet.

Deltagerne vurderer gennemsnitligt på en skala fra 1-5 (svært-let), at finde information ligger på 3,6. Vi konkluderer derfor, at de synes det er rimelig let at finde information. Begrundelserne for deres svar er, at der findes meget information, når

det er store projekter, og at der typisk er andre, der tidligere har spurgt efter samme og svaret derfor allerede findes som information i projektet. Trods det svarer 55% samtidig, at de skal søge mere end et sted, for at finde den information de søger. Vi mener derfor, at søgningen kan optimeres, således de kun skal søge et sted. Nogle af deltagere påpeger, at de har svaret middel, da de nogle gange finder det nemt at finde information og andre gange svært. I interviewet skal dette derfor uddybes, så vi kan se, hvilken information som er let/svært at finde.

Hypotese 2 *Deltagere har forskellige behov for hvilken slags information, der skal være tilgængelig i et projekt.*

Resultaterne fra spørgeskemaet viser en klar forskel i, hvilke informationer de forskellige deltagere søger i projektet. Nogle søger eksempelvis efter tutorials og andre information om nye releases. Fra differentieringen kan vi udlede deltagerens forskellige krav, til hvilken slags information de ønsker skal være tilgængelige i projektet. Vi kan derfor konkludere, at denne hypotese er valid. Der er derfor ikke behov for at uddybe den i interviewet.

Hypotese 3 *Kompleksiteten af informationssøgning afhænger af størrelsen af det pågældende projekt, og hvor længe en deltager har været involveret i projektet.*

Yderligere ses det af resultaterne, hvorledes deltagerne mener, at det bliver lettere at søge information, jo længere tid de har været involveret i projektet. Nogle af deltagerens uddybende svar i spørgeskemaet giver et muligt svar på dette. En deltager giver f.eks. dette svar:

“I’ve got over the initial learning curve.” 210, reader, The Gimp

Det er således erfaring, der gør, at de ikke finder, at det bliver sværere at finde information. Dette strider imod antagelsen om, at det bliver sværere, fordi projektets kompleksitet øges, som tiden går. I interviewet stilles uddybende spørgsmål til dette for at finde deltagerens holdning til, om kompleksiteten påvirker, hvor svært det er at søge efter information. Da vi vurderer, at alle tre udvalgte projekter er store, kan vi derfor ikke validere, om der er forskel afhængigt af projektets størrelse.

7 Identifikation af arketyper

Dette kapitel repræsenterer trin nummer 2, “Udvælgelse og beskrivelse af arketyper på baggrund af empiri”, i specificeringen af arketyper i forhold til vores metode. Dette trin baseres på besvarelserne af spørgeskemaundersøgelsen. I kapitlet opdeles deltagerne i tre arketype, og det beskrives, hvorledes de differentierer sig.

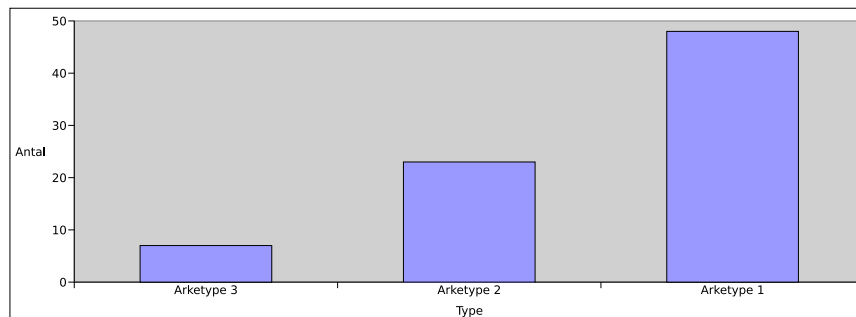
7.1 Bestemmelse af arketyper

Ved at behandle resultaterne fra spørgeskemaet fremgår det, hvilke forskelle og ligheder, der er blandt deltagertyperne. Dette leder op til, at arketyperne kan præciseres nærmere og begrænses til færre typer. Hvor den initiale beskrivelse af typerne er en nuanceret og bred opdeling, ønsker vi at gruppere disse initiale typer ud fra deres ligheder, og derved opnå færre typer. I grupperingen fokuseres på, hvilke karakteristika der samler og differentierer typerne. Ud fra forskellene og lighederne blandt de første syv deltagertyper vælger vi at inddele arketyperne i tre grupper.

Opdelingen af typerne er som følgende: (se også figur 7.1)

- Arketype 1: Readers og Bug reporters, 48 deltagere
- Arketype 2: Bug fixers og Peripheral developers, 23 deltagere
- Arketype 3: Active developers, Core members og Project leaders, 7 deltagere

De områder hvor typerne grupperer sig ved: hvilken information de søger, deres behov for awareness, tiden de har været involveret i projektet og deres valg af in-



Figur 7.1: Fordeling af arketyper efter gruppering

formationskanal. I det følgende beskrives, hvorledes typerne grupperer sig indenfor disse områder.

7.1.1 Hvilken information søger de efter

Ved at kombinere svarene fra spørgsmål 6 (In which way are you involved in this project?) og 17 (Which two kinds of information are you usually looking for?) i spørgeskemaet, gives der et billede af, hvilken type information de forskellige typer af deltagere søger efter. Da deltagerne har mulighed for at svare hvad som helst i spørgsmål 17, er det nødvendigt at klassificere deres svar i kategorier, således at det er muligt at opstille disse i et diagram. Vi vælger at klassificere deltageres svar i de følgende fire kategorier. De kategorier som de forskellige informationer er inddelt under, vil også være de inddelinger informationerne er inddelt under i resten af projektet.

- Type 1, Teknisk dokumentation: Dækker over teknisk information som eksempelvis API dokumentation, programmerings dokumentation samt kode eksempler. API kan også ses som en manual, men af teknisk karakter, og derfor er den klassificeret som denne type.
- Type 2, Forbedringer: Denne type er information som eksempelvis bug reports og fixes samt forbedringer (enhancement). Alt sammen information, der omhandler forbedringer af softwaren på et teknisk niveau.
- Type 3, Awareness: Denne kategori indeholder annonceringer af eksempelvis nye features i softwaren, fremtidige planer for projekt samt anden awareness information. Denne kategori adskiller sig fra de andre kategorier ved at være abstrakt. I eksemplet med annoncering af nye features, vil selve den nye feature falde ind under type 2, men idet det er en *annoncering*, klassificeres dette som awareness information. Selvom denne kategori er abstrakt, mener vi, at denne kategori kan sidestilles med de tre andre typer, da vi ser awareness som information.

- Type 4, Manualer: Kategorien består af løsninger til problemer, HOWTOs, tutorialer og dokumentation, der beskriver brugen af softwaren, samt hvordan problemer løses. Det skal ikke ses som problemer med en bug, men en hjælp til hvordan et stykke software anvendes.

I det følgende afsnit gives eksempler på, hvorledes deltageres svar klassificeres i de fire kategorier.

Klassificeringen af svar

Som det beskrives i ovenstående afsnit, laves en klassificering af de typer af information, som deltagerne søger efter. I dette afsnit gives eksempler på, hvorledes deltageres svar klassificeres.

I spørgeskemaet giver en af respondenterne følgende svar:

“I am spending a lot of time with Bugzilla, fixing bugs or simply maintaining bug reports. I also try to surveillance user forums and mailing-list in order to get an idea of what the users want and what problems they are faced with.” 137, project leader, The Gimp

Ovenstående citat er klassificeret under to kategorier, nemlig *forbedring* og *awareness*. Første del af citatet: “I am spending a lot of time with Bugzilla, fixing bugs or simply maintaining bug reports” er klassificeret som typen *forbedring*, da deltageren gennem Bugzilla forsøger at udbedre fejl i projektet. Og anden halvdel: “I also try to surveillance user forums and mailing-list in order to get an idea of what the users want and what problems they are faced with.” klassificeres som typen *awareness*, da deltageren vil være bevidst om, hvad deltagerne i projektet ønsker.

“Plug-in API, old mails” 170, core Member, The Gimp

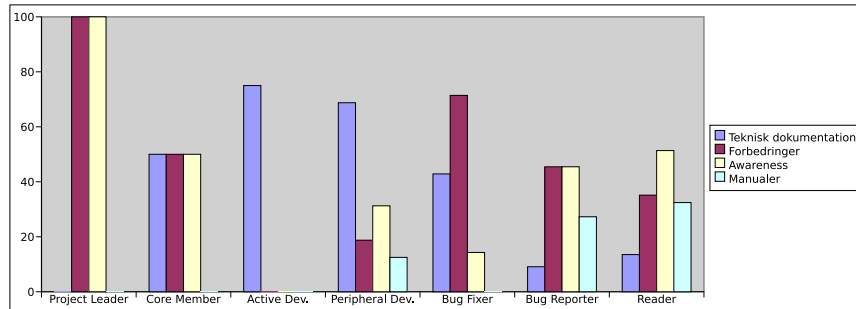
Første del af ovenstående citat: “plug-in API” er klassificeret under typen *teknisk dokumentation*. Anden del af citatet: “old mails” er sværere at klassificere, da det ikke siger noget om, hvilken type information han søger efter i de gamle e-mails. Denne del af citatet har vi derfor valgt ikke at klassificere.

I det følgende diskuteres, hvorledes de forskellige deltager fordeler sig, i forhold til den information de søger.

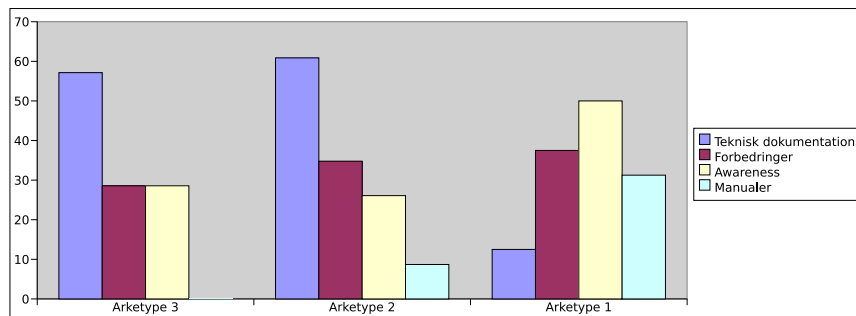
Fordeling af typer i forhold til informationstype

Deltagerne differentierer sig på dette punkt ved, at arketype 3 primært søger efter teknisk dokumentation, og at det kun arketype 1 og 2, der søger efter manualer.

Dette kan ses ud fra søjlediagram 7.2, som viser de 7 initiale typer, samt hvilken information de oftest søger efter. I søjlediagram 7.3 ses samme information, men her er deltagerne opdelt i de 3 arketyper.



Figur 7.2: Søjlediagrammet viser hvilken type af information, de 7 typer leder efter. Se afsnit "beregning" for hvorledes søjlediagrammet er beregnet



Figur 7.3: Søjlediagrammet viser hvilken type af information, de 3 arketyper leder efter. Se afsnit "beregning" for hvorledes søjlediagrammet er beregnet

En deltagers svar kan klassificeres som flere typer af information. Derfor beskrives det i næste afsnit, hvorledes diagrammerne beregnes.

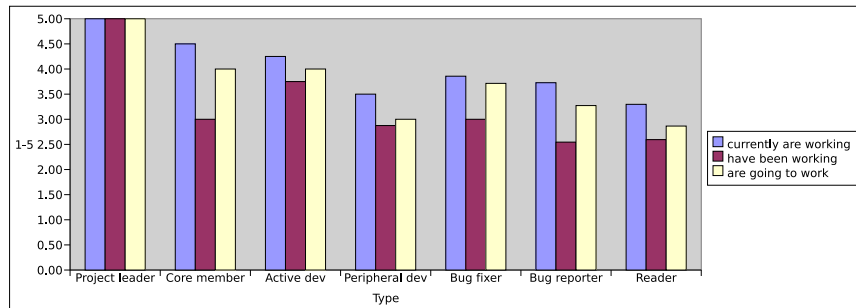
Beregning

Diagrammerne over hvilken information arketyperne søger, summer ikke til 100%. Dette skyldes, at respondenterne har mulighed for at give mere end et svar, og derfor kan deres svar klassificeres i mere end en kategori.

Eksempelvis er der i arketype 3 (se søjlediagram 7.3) 4 ud af 7 som har svaret, at de søger efter *teknisk dokumentation*, 2 ud af 7 søger efter *forbedringer*, 2 ud af 7 har svaret *awareness*, og ingen har svaret, at de søger efter *manualer*. Diagrammet skal derfor forstås således, at 57% af arketype 3 søger efter teknisk dokumentation, men samtidig kan de godt søge anden information, og på den måde være repræsenteret mere end en gang.

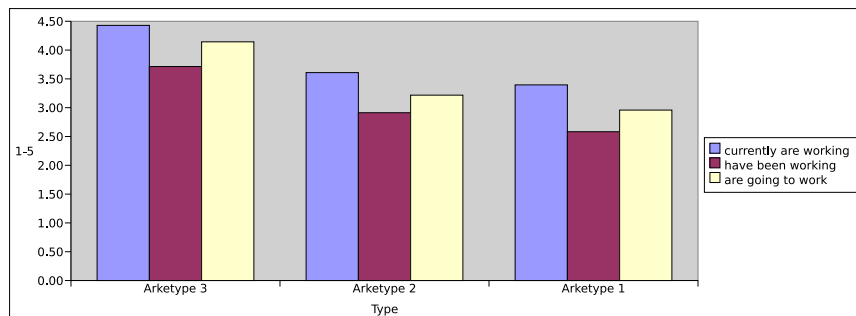
7.1.2 Awareness

Søjlediagram 7.4 viser, hvorledes de forskellige typer af deltagere vurderer relevansen af at kunne få awareness information. Skalaen går fra 1, hvilket betyder, at de finder det irrelevant og op til 5, som betyder at de finder det meget vigtigt. Diagrammet viser, at søjlerne er højest ved project leader og faldende mod readers, som ligger lavest.



Figur 7.4: Relation mellem interesse i awareness og deltager type. (1: Ikke interesseret - 5: Meget interesseret)

Ud fra søjlediagram 7.4 kan det konkluderes, at jo mere involveret en deltager er i projektet (og derved placeret tættere mod midten ifølge figur 5.1), jo vigtigere og større er behovet for awareness. I søjlediagram 7.5 kan ses, hvorledes fordeling er i forhold til de tre reviderede typer.

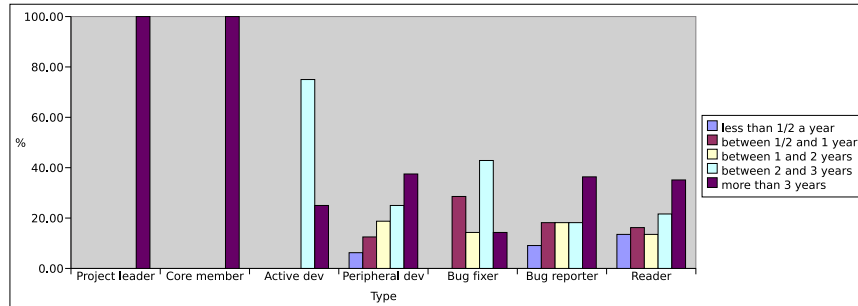


Figur 7.5: Relation mellem interesse i awareness og reviderede arketyper. (1: Ikke interesseret - 5: Meget interesseret)

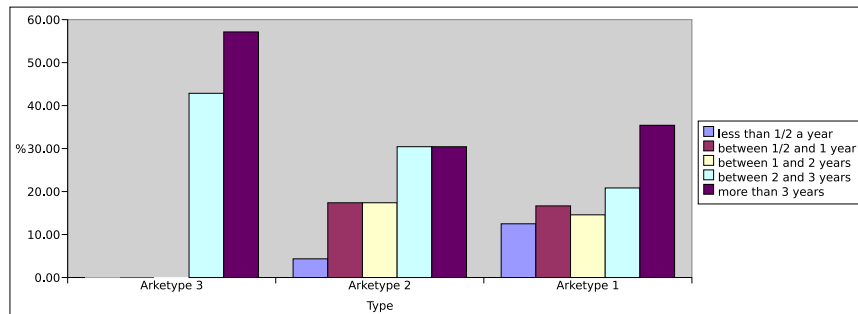
7.1.3 Tid involveret i projektet

Søjlediagram 7.6 illustrerer, hvor længe de forskellige typer har været involveret i projekterne. Det ses, at arketype 3 i gennemsnit har deltaget længere tid i projekterne end de to andre typer (se søjlediagram 7.7). Således differentierer arketype 3 sig på dette område.

Arketype 1 er spredt med hensyn til, hvor længe deltagerne har været med i projektet. De to andre arketyper har derimod en fordeling, hvor deltagelses perioden



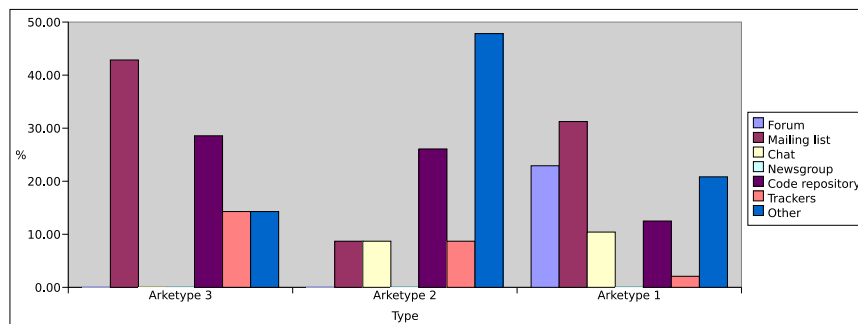
Figur 7.6: Hvor længe de initiale typer har været involveret i projekterne



Figur 7.7: Hvor længe de reviderede typer har været involveret i projekterne

er længere. I arketype 2 har 60% været med i mere end 2 år, og i arketype 3 har alle været med i 2 år eller mere. Dette giver en indikation af, at jo længere tid man har været med i projektet, jo mere bidrager man med, og jo mere involveret er man.

7.1.4 Valg af informationskanal



Figur 7.8: Hvor leder typerne først efter information

Søjlediagram 7.8 viser, hvilke af informationskanalerne arketyperne vælger at søge i først. Det ses at de deltagere, som hører til arketype 3, holder sig til færre informationskanaler end de andre typer. Deltagerne i arketype 1 har det til fælles, at de spreder sig over flest informationskanaler. Der er også forskel i, hvilke in-

formationskanaler typerne foretrækker. I arketype 3 anvender deltagerne primært mailing list eller versionsstyringssystemer (Code repository). Arketype 2 benytter også mailing lister, men overvejende andre kanaler end de officielle. Arketype 1 deler sig over tre kanaler, som hovedsageligt anvendes. Dette er forum, mailing list og tredje parts.

7.1.5 Opsumering

For at opsummere på arketyperne og deres karakteristika opstilles de som punkter:

- Arketype 1
 - Type af information: awareness og forbedringer, den arketype der søger mest efter manualer
 - Tid involveret i projektet: 1-2 år
 - Primær informationskanal: mailing list, forum og other. Karakteristisk er også, at de anvender alle informationskanalerne undtagen newsgroup.
 - Awareness om hvad andre arbejder med: 3,4
 - Awareness om hvad andre har arbejdet med: 2,58
 - Awareness om hvad andre skal til at arbejde med: 2,96
- Arketype 2
 - Type af information: teknisk dokumentation
 - Tid involveret i projektet: 2-3 år
 - Primær informationskanal: forum, 3. parts som Google
 - Awareness om hvad andre arbejder med: 3,61
 - Awareness om hvad andre har arbejdet med: 2,91
 - Awareness om hvad andre skal til at arbejde med: 3,22
- Arketype 3
 - Type af information: teknisk dokumentation
 - Tid involveret i projektet: mere end 2 år
 - Primær informationskanal: mailing list, code repository
 - Awareness om hvad andre arbejder med: 4,43
 - Awareness om hvad andre har arbejdet med: 3,71
 - Awareness om hvad andre skal til at arbejde med: 4,14

7.2 Prioritering af arketyper

Som Kujala og Kauppinen [2004] beskriver, kan en matrice anvendes til at vægte deltagerne og herudfra hvilken funktionalitet systemet skal indeholde. Dette er brugbart, hvis der skal udvikles et system, hvor en bred målgruppe skal bruge samme system. Typisk kan dette være fordelagtigt i organisationer, hvor brugerne ofte har faste roller, og hvor det er nemt at sige, hvor mange brugere der er af hver rolle.

I Open Source projekter hvor rollerne er flydende og skiftende (se målgruppe beskrivelsen afsnit 5.1), er det problematisk at anvende denne matrice. Der er ikke et klart overblik over, hvor mange deltagere der er i hver gruppe, og dette kan også ændre sig afhængigt af de skiftende roller, deltagerne kan have.

I problemformuleringen (se afsnit 3.2) beskrives, at vores system designes til alle typer deltagere. Vi vælger at vægte alle typerne ligeligt. Spørgeskemaundersøgelsen tegner et billede af, at der er flest readers og færrest project leaders. Stadig vægter vi ikke readers højere, da project leader til gengæld har en central rolle for projektet.

Interview

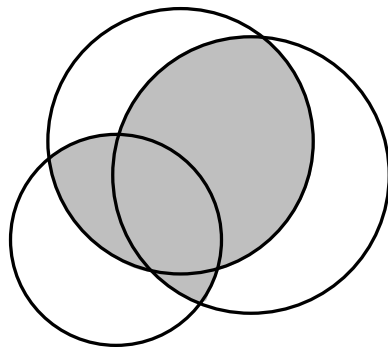
Dette kapitel omhandler indsamling af information og repræsenterer således første del af trin 3, “Indsamle information og redesign af beskrivelse af arketyper”, i vores metode. Kapitlet beskriver de spørgsmål, der indgår i det interview, og som sendes ud til repræsentative deltagere fra hver arketype. Først beskrives formålet med interviewet, derefter opbygningen og til sidst de enkelte spørgsmål samt deres formål.

8.1 Beskrivelse af interview

Formålet med interviewet er både at differentiere arketyperne yderligere, samt uddybe områder hvor de er ens. Ligeledes benyttes interviewet til at validere de hypoteser, der ikke er blevet dette igennem spørgeskemaet.

Formålet med at kigge på de områder, hvor deltagerne differentierer sig er, at finde de ydre punkter hvor arketyperne adskiller sig, som der er beskrevet i identifikation af arketyper (se afsnit 7.1). Ved at stille uddybende interview spørgsmål til disse områder, kan man f.eks. finde de områder, hvor arketyperne har modstridende krav, som er spændende i forhold til vurdering og udvikling af krav og design af systemet. Figur 8.1 illustrerer, hvorledes arketyperne har visse krav tilfælles, det grå område, og de krav deltagerne ikke har tilfælles, de hvide områder. Det der ligger udenfor cirklerne, er uden for målgruppen. Det er netop i de hvide områder, vi antager, at de modstridende krav ligger. Arketypernes modstridende krav kan være forbundet med mange forskellige aspekter. Eksempelvis kan man antage, at arketype 3 er mere teknisk pga. deres involvering i Open Source projektet, og derfor har mere tekniske krav til systemet.

Der udvælges en række deltagere fra spørgeskemaet, der er repræsentative for den arketype, de er klassificeret som, til at besvare spørgsmål i interview. I spørgeskemaet kan deltagerne frivilligt oplyse deres e-mail, hvorved de samtidigt giver os lov



Figur 8.1: Figuren illustrere hvor arketyperne differentiere sig, samt hvor de er ens

til at kontakte dem for at udføre interviewet. Ud af de 78 deltagere, der har udfyldt spørgeskemaet, har 30 deltagere oplyst deres e-mail adresse. Det er derfor i blandt disse 30, at der udvælges repræsentative deltagere til interview. Da der kun kan vælges repræsentative deltagere ud fra de 30, der har tilkendegivet deres e-mail, og ikke alle 78 deltagere der har svaret på spørgeskemaundersøgelsen, risikeres, at det ikke er muligt kun at vælge de mest repræsentative deltagere.

Interviewet udføres som et skriftligt interview via e-mail, da dette er en effektiv måde at kommunikere med personer, der er fordelt over hele verden. Fordi det er et skriftligt interview, begrænses mængden af spørgsmål, idet et skriftligt interview kræver mere arbejde af respondenteren end et mundtligt interview gør.

I det følgende beskrives hvilke deltagere, der er udvalgt, samt hvorledes disse er udvalgt.

8.1.1 Respondenter til interviewet

Ud fra svarene i spørgeskemaet (se afsnit 6.1) udvælges ni personer, tre personer fra hver af arketyperne. I beskrivelsen af arketyperne (se afsnit 7.1.5) opstilles en række karakteristika, som arketyperne besidder. Disse karakteristika ses som kriterier, som deltagerne skal opfylde for at være repræsentative for deres arketype. Da det kun er tre deltagere, som opfylder de fire kriterier, accepteres deltagere, som opfylder tre ud af fire kriterier.

Personerne der udvælges kan ses i tabel 8.1. I tabellen sættes \times ved de kriterier, som deltagerne opfylder. Den kolonne der hedder awareness, dækker over de tre underspørgsmål i spørgsmål 19. For at få \times i denne kolonne, skal deltagerne passe med gennemsnittet på to af de tre. I bilag F ses tabeller over deltagerens svar på de 4 kriterier.

Følgende afsnit beskriver de spørgsmål, som stilles til deltagerne. Idet der stilles spørgsmål til arketyperne og ikke deltagerne specifikt, bruges de samme spørgsmål til alle arketyperne. Nogle spørgsmål er dog tilpasset deltagerne, da de spørger di-

ID	Type information	Tid	Primær info-kanal	Awareness
Arketype 1				
108	×		×	×
180	×		×	×
188	×	×	×	×
Arketype 2				
107	×		×	×
114	×	×		×
171	×		×	×
Arketype 3				
207		×	×	×
121	×	×	×	×
204	×	×	×	×

Tabel 8.1: Personer som udvælges fra de tre arketyper

rekte til de svar, de giver i spørgeskemaet. I spørgsmålene er teksten, der står mellem “<>”, tilpasset den deltager interviewet udsendes til. Således kan deltageren uddybe det svar, personen har givet i spørgeskemaet.

8.1.2 Introduktion

E-mailen, der udsendes til deltagerne, indleder således:

“First of all we would like to thank you for participating in our survey about communication in Open Source. You provided us with your e-mail address so we could contact you later, which we are doing now :) We would like you to answer the following questions in depth. If you don’t feel like participating in this interview anyway, we would appreciate it if you let us know. The following questions revolve around finding information in the <Projekt >project.”

Formålet med introduktionen er at gøre det klart for deltagerne, at de skal give detaljerede svar til spørgsmålene. Desuden bedes deltagerne skrive tilbage, hvis de ikke har lyst til at deltage i interviewet. Dette giver overblik over, hvor mange der er villige til at svare, således andre deltagere kan udvælges, hvis der er en lav svarprocent. Desuden oplyses deltagerne om, at de skal besvare spørgsmålene ud

fra det projekt, de deltager i, da der var forvirring omkring dette i forbindelse med spørgeskemaet (se afsnit 6.2.10)

8.1.3 Informationsøgning

De første tre spørgsmål i interviewet omhandler, hvorledes deltagerne søger efter information.

1. *Which kinds of information do you consider easy to find? Or difficult to find? To answer this question you can describe situations in which you found it easy/difficult to find information. Please also state what you think the reason is?*

Formål: Ud fra besvarelserne af dette spørgsmål kan udledes, hvilken type information, der er svær eller let at finde, og hvorfor det forholder sig således. Svarene fra spørgeskemaet giver ikke et entydigt svar på dette, idet der i spørgeskemaet ikke differentieres mellem forskellige informationstyper, når der spørges til, hvor svært det er at finde information. I spørgeskemaet spørges der efter, hvilke to typer information de leder efter (spg 17), samt om deltagerne generelt synes, det er let at finde information (spg 15). Vi kan derfor ikke konkludere, om der er forskel på at finde forskellige typer af information. Dette illustreres også af den følgende kommentar til spørgsmål 15:

“15 is a frustrating question because I want to say sometimes 1 and sometimes 5, but not in the middle. (When the information is there, it’s usually not too hard to find, but there are exceptions)”
153, Peripheral developer, The Gimp

Da denne relation ikke belyses tydeligt i spørgeskemaet, vælger vi at følge op på det i interviewet, da dette skal bidrage til at besvare hypotese 1.

Dette spørgsmål kan bidrage til at designe vores system, idet det belyser hvilken information, der er svær at finde. Systemet skal derfor have funktioner til at finde denne type af information.

En motivation til dette spørgsmål er også en undren over deltageres svar i spørgsmål 12 og 17. Mere end halvdelen af deltagerne siger i spørgsmål 17, at det er let at finde den information, de søger efter. Samtidig oplyser deltagerne, at de skal lede mere end et sted for at finde den information de søger (spg 12). Vi undrer os over, at deltagerne synes, det er let at finde information, på trods af at de skal søge mere end et sted for at finde den. Derfor undersøges, hvornår de finder det svært at finde information.

2. *In the survey you answered that you primarily use the <informationskanal fra spørgsmål 10 > to search for information. What do you like about <informationskanal fra spørgsmål 10 >? Is it for example the possibility to ask others, the design or is it because it's widely used or something else?*

Formål: Ud fra spørgsmål 10 i spørgeskemaet, der lyder “when you are searching information where do you look first”, fremgår det hvilke informationskanaler, deltagerne foretrækker at søge i. Der kan dog ikke udledes, hvad der gør, at de benytter denne informationskanal som det første sted, de søger. Formålet med dette spørgsmål er derfor at belyse hvilke funktioner, muligheder eller noget helt tredje der gør, at deltagerne benytter den informationskanal, de har oplyst i spørgeskemaet.

Hvis en respondent svarer “Other” i spørgsmål 10, i spørgeskemaet, benyttes svaret fra spørgsmål 11 i stedet. Spørgsmålet lyder: “If you did not find the information the first place you looked, where would you then look?”. Begrundelsen for dette ligger i, at det er informationskanalerne beskrevet i afsnit 2.2, der er interessante i forhold til problemstillingen i dette projekt, og ligeledes disse som undersøges og ikke andre.

Svarene på spørgsmålet giver viden omkring, hvilke ting de forskellige arketyper kan lide ved de forskellige informationskanaler. Dette kan give ideer til hvilke funktioner, der skal være til rådighed i vores system.

3. *We asked you in the survey which two kinds of information you usually look for are looking for, and you answered: <deres svar til spørgsmål 17 >. What is your motivation for searching this information? Is it for example because you: have problems using the software/want to improve your programming skills/want to learn new things/are interested in open source/or something entirely different?*

Formål: Formålet med dette spørgsmål er at undersøge, hvilken motivation deltagerne har for at søge informationer, og hvilke muligheder der skal være for at finde de informationer, de søger. F.eks. hvis deltagerne gerne vil have mulighed for at forbedre sine kode evner, skal det være nemt at finde information omkring koden.

Endvidere skal spørgsmålet også være med til at give et billede af arketypen. Dette udspringer fra hypotese 1:

Hypotese 1 *Formålet med informationssøgning, hvorledes søgningen udføres, problemer angående søgning af information og hvilke informationskanaler der søges i, afhænger af typen af deltager.*

Vi har identificeret, at deltagerne også adskiller sig i hvilken to typer af information de typisk leder efter (spg 17 i spørgeskemaet). Vi mener dog, at det i dette tilfælde ikke er nødvendigt at uddybe dette mere. I stedet benyttes dette

spørgsmål til at se, om arketyperne differentierer sig ved at have forskellig motivation for at søge information.

Ved at få et indblik i deltagernes motivation kan designet af systemet tilpasses til at støtte denne motivation optimalt.

8.1.4 Awareness

Det følgende spørgsmål belyser, hvorledes arketyperne differentiere sig i forhold til awareness.

4. *In your answers to the following questions in the survey you estimated your need for information to be <værdier fra deres svar i spørgsmål 19 >.*

- *How interested are you in knowing what people... (1: Not at all - 5: Very much)*
 - *currently are working on in this project?*
 - *have been working on in this project?*
 - *are going to work on in this project?*

Please elaborate on why you chose this? You can give examples of when you needed this information, and why you needed it.

Formål: Spørgsmålet gør det muligt at uddybe differentiering af deltagernes behov for awareness. Desuden belyses i hvilke situationer, der er behov for awareness. Dette giver en forståelse af, i hvilken kontekst der er behov for awareness, eksempelvis om det kun er når en udvikler skal kode noget nyt, at denne har behov for awareness.

8.1.5 Komplexitet

Næste spørgsmål omhandler, hvad deltagerne mener, har indflydelse på deres informationsøgning.

5. *Do you think that factors like the amount of users, information, and information channels has an influence on how it is to find information in this project?*

Please describe why you think these factors have or don't have influence?

Are there other factors that influence finding information? Please describe which factors and why?

Formål: I hypotese 3 antages, at det bliver sværere at finde information, når et Open Source projekt bliver mere komplekst. Men i spørgeskemaet har deltagerne svaret, at det faktisk er blevet nemmere. Begrundelsen, deltagerne giver for dette, er erfaring. For at få en forståelse af, om erfaring er nok til at modvirke problemer med informationssøgning, når projektet bliver mere komplekst, spørges ind til dette.

8.2 Evaluering af interview

I dette afsnit evalueres interviewet, i forhold til problemstillinger ved interviewet samt feedback modtaget af deltagerne i interviewet.

8.2.1 Redesign af arketyper

Som beskrevet i metoden skal interviewet være med til at validere beskrivelsen af arketyperne. Hvis resultaterne af interviewet indikere at inddelingen af arketyperne ikke er korrekt skal arketyperne redesignes. Ud fra resultatet af interviewet (se bilag G) kan det ses at de udvalgte repræsentative arketype stemmer overens med vores inddeling. Der er dog kun et lille antal arketyper der har besvaret interviewet, denne problemstilling beskrives i det følgende.

8.2.2 Udsendelse af interview

Det beskrives i afsnit 8.1.1, at interviewet sendes til de ni personer, som opfylder de kriterier der opstilles. Idet det kun er to personer, som besvarer interviewet indenfor en kortere periode, vælges det at sende interviewet ud til alle de personer, der oplyser deres e-mail i spørgeskemaet. På denne måde mindskes risikoen for få tilbagemeldinger.

Hvor de første ni personer er repræsentative på flere punkter, er de øvrige vi sender ud til kun repræsentative på nogle punkter. De enkelte svar skal derfor anvendes sammen for at samle en repræsentativ deltager.

8.2.3 Uddybning af spørgeskema

Problemstillingen vi behandler i spørgeskemaet (spg 12) "Do you experience that you have to search several sources (e.g. mailing lists, news groups and so on) to find the information you are looking for?" burde vi have belyst i interviewet, da dette omhandler den centrale problemstilling i projektet; at deltagerne i Open Source projekter skal søge i flere forskellige informationskanaler, for at finde det de søger.

Da vi ikke har belyst denne problemstilling, kan vi ikke konkludere *hvorfor* deltagerne skal søge flere steder. Vi kan dog have nogle antagelser omkring dette base-

ret på spørgsmål 1 i interviewet. De informationer deltagerne oplyser som svære at finde, kan også være de informationer, deltageren bliver nødt til at lede flere steder efter. Yderligere må vi antage, at deltagerne søger flere steder, hvis de ikke kan finde informationen det første sted de søger.

8.2.4 Generiske spørgsmål

De mails der er udsendt til deltagerne, er opbygget ens, men er tilpasset den enkelte deltager ud fra, hvad denne har svaret i spørgeskemaet. Til enkelte af spørgsmålene opstilles en række eksempler på svarmuligheder, således at deltageren lettere kan svare på spørgsmålet. Kombinationen af en standard mail, hvor deltagerens svar er sat ind betyder, at enkelte af de eksempler vi giver, ikke passer på det deltageren har svaret. En deltager har påpeget dette, se nedenstående citat:

“You’ve processed the responses to the multiple choice questions into a form letter (which is clear when you talk about code repositories and “the possible to ask others”) and asked me to elaborate on each of them in detail.” E-mail, Peripheral developer, Gaim

Vi mener dog ikke, at det er et problem, da eksemplerne netop er eksempler, og ikke det deltageren skal vælge mellem i sit svar.

8.3 Opsamling på hypoteser

Dette afsnit konkluderer på, hvorledes hypoteserne valideres igennem besvarelsen af interviewet. Der er fokus på hypotese 1 og 3, da hypotese 2 er valideret igennem spørgeskemaet. På grund af den lave bevarelse af interviewet kan hypoteserne ikke valideres med sikkerhed. Besvarelsene på interviewet er at finde i bilag G

Hypotese 1 *Formålet med informationssøgning, hvorledes søgningen udføres, problemer angående søgning af information og hvilke informationskanaler der søges i, afhænger af deltageren.*

Formålet med interview spørgsmål 1 er at få mere information, der kan være med til at validere hypotese 1 yderligere. Det kan ses ud af besvarelsene til spørgsmål 1, at deltagerne beskriver forskellige problemer med at finde information, samt hvilke informationer der er svære og lette at finde. Dette kan indikere, at vores antagelse i hypotese 1 holder. I afsnit 9.1 beskrives de 3 arketyper måde at foretage informationssøgning, og hvordan de oplever dette.

Hypotese 3 *Kompleksiteten af informationssøgning afhænger af størrelsen af det pågældende projekt, og hvor længe en deltager har været involveret i projektet.*

I opsamlingen på hypotese 3 efter spørgeskemaet (se afsnit 6.3) beskrives, at det yderligere skal uddybes om deltagerne er af den opfattelse, at et projekts størrelse

har indflydelse på, hvor svært det er at finde information. Ud fra besvarelserne af interviewet kan det konkluderes, at det har en indflydelse. Eksempelvis svarer deltagerne:

“Yes. :)” 149, reader, The Gimp

“Yes, of course. Is there any project where it doesn’t in some way or other?” 126, reader, Gaim

Flere af deltagerne påpeger positive aspekter ved at projektet vokser. De mener at flere brugere resulterer i bedre dokumentation og bedre svar:

“As a general rule, I think that the more users=more questions, More questions->better documentation so that people don’t have to ask so many questions.” 124, bug reporter, The Gimp

“On the other hand if you have many educated, users that at least know how to read a manual and how to look for already answered question before asking their questions, that would help tracking issues faster.” 107, peripheral, MPlayer

Ud af dette kan der derfor på hypotese 3 konkluderes, at størrelsen af projektet giver de fordele, i forhold til at søge information, at der bliver flere deltagere at spørge samt en bedre dokumentation. Samtidig mindskes kompleksiteten af at søge jo længere tid deltageren har været involveret, da deltageren får erfaring (se beskrivelse i afsnit 6.3). Dermed er hypotese 3 valideret.

Arketyperne

Dette kapitel identificerer karakteristika ved de tre arketyper og redesigner derved beskrivelsen af arketyperne. Dette er anden del af trin 3, "Indsamle information og redesign af beskrivelse af arketyper" i vores metode. Identificeringen af karakteristika sker via en analyse af deltagernes svar i spørgeskemaet og interviewet. Først beskrives hvorledes analysen er opbygget og herefter følger analyse af hver af de tre arketyper.

9.1 Analyse af arketypernes karakteristika

Til analyse af arketypernes karakteristika tages udgangspunkt i spørgeskemaet (se afsnit 6.1) og interviewet (se afsnit 8.1). Spørgsmålene fra begge undersøgelser klassificeres i en række kategorier, som beskrives i det følgende. I hver kategori opstilles spørgsmål, som svarene fra spørgeskemaet og interviewet bidrager til besvarelsen af.

De spørgsmål fra spørgeskemaet og interviewet, der udvælges, bidrager alle med informationer om karakteristika ved arketyperne. Der er kun fokus på dem som er task-related. Dette betyder, at der eksempelvis ikke fokuseres på de spørgsmål, der belyser arketypernes alder, nationalitet osv.

De tre kategorier er:

- Informationskanaler
 - *Hvor søger arketypen information?* Dette spørgsmål belyser hvilke informationskanaler arketyperne benytter samt hvad de foretrækker ved kanalerne.
Spørgsmål 10, 11, 13, 18, 20 fra spørgeskemaet samt spørgsmål 2 fra interviewet

- *Hvor meget bruger arketyperne informationskanalerne?* Dette spørgsmål benyttes til at få en forståelse af hvor aktiv arketyperne er, igennem hvor ofte de benytter informationskanalerne.

Spørgsmål 8, 9 fra spørgeskemaet.

- Informationssøgning

- *Hvordan oplever arketyperne informationssøgning?*

Spørgsmål 12, 15, 16, 21, 22 fra spørgeskemaet, spørgsmål 1, 5 fra interviewet.

Dette spørgsmål benyttes til at identificere hvorledes arketyperne oplever informationssøgningen, eksempelvis hvilke problemer de oplever, samt hvad arketyperne oplever gør informationssøgningen kompleks?

- *Hvordan foretrækker arketyperne, at informationssøgningen udføres?* Dette spørgsmål omhandler den måde hvorledes arketyperne søger efter information. Dette indebærer hvilke informationskanaler arketyperne foretrækker samt hvorfor det forholder sig således.

Spørgsmål 14, 23, 24, 25, 26 fra spørgeskemaet samt spørgsmål 2 fra interviewet.

- Information

- *Hvilken information søger arketyperne?* Dette spørgsmål belyser hvilke information arketyperne søger, samt hvorfor de søger efter denne information.

Spørgsmål 17, 19 fra spørgeskemaet, samt spørgsmål 1, 3, 4 fra interviewet.

Analysen er opbygget således, at de tre arketype analyseres for hver af de tre kategorier. Efter hver arketype følger en opsummering af arketyperens karakteristika.

9.1.1 Udvalgelse af svar

Som det beskrives i afsnittet omkring udvælgelse af respondenter til interviewet (se afsnit 8.1.1), er der kun kommet svar fra en af de ni personer, som anses for at være repræsentative for deres arketype. Dette betyder, at der er brugt respondenter, som ikke er repræsentative. Disse deltagers svar anvendes kun i de tilfælde, hvor de opfylder kriterierne for arketyperne for det pågældende spørgsmål. Dvs. at hvis en deltager ikke er blandt de ni repræsentative deltagere men har svaret mailing liste, og dette er karakteristisk for arketyperne, anvendes deltagerens svar.

9.2 Arketype 1

9.2.1 Hvor søger arketypen information?

Af søjlediagram D.14 fremgår det, at arketype 1 søger i alle informationskanaler undtagen nyhedsgrupper. De fleste vælger mailing lister, fora eller andre kanaler, eksempelvis Google eller projektets website, som det første sted, de søger. Derefter er det varieret, hvilken informationskanal de vælger som andet valg. Deres andet valg kan ses på søjlediagram D.15. Grunden til at ingen af deltagerne anvender nyhedsgrupper skyldes, at ingen af projekterne, deltagerne er involveret i, officielt har tilknyttet nyhedsgrupper.

I interviewet uddyber deltagerne, hvad de kan lide ved den kanal, de foretrækker. Da arketype 1 er karakteriseret ved at sprede sig over alle kanaler (undtagen nyhedsgrupper), kan alle deltagernes svar bruges. Der er tre deltagere fra arketype 1, der har svaret. Svarene lyder:

“(Chat (e.g. IRC)): I like chat because it’s realtime and also gives me the opportunity to give back to the community by helping others. Even if I’m not participating and just watching a chat, or reading chat messages from earlier in the day, I feel like I’m part of the active community, rather than just a detached end-user.” 149, reader, The Gimp, interview

“(Forum): A lot of it is probably force of habit - I’m for some reason more comfortable with forums than for instance mailing lists or IRC. I think that forums probably represent a good middle ground there, easy to post and reply without having to fire up an external program (as with mail) at the same time searchable and persistent (not volatile like IRC). So, for any project that provides them, I tend to use the forums first and foremost if it is not apparent that some other channel is much more efficient.” 126, reader, Gaim, Interview

“(Mailing list): Less spam than newsgroups + more focused group.” 124, bug reporter, The Gimp, interview

De tre deltagere fokuserer på forskellige fordele ved informationskanalerne. Deltager nr. 149 vælger chat, fordi den er realtime, men også fordi han kan føle sig som en del af projektet og hjælpe andre ved at følge med på chatten. Deltager nr. 126 foretrækker derimod fora, da han derved undgår at skulle anvende et eksternt program, og fordi fora er persistente samt giver mulighed for at søge. Yderligere siger han, at en del af grunden er at han er komfortabel med fora. Den sidste deltager anvender primært mailing lister, idet de er fokuserede og der ikke er for meget spam. De tre deltagere har derfor ikke noget til fælles om, hvorfor de foretrækker en kanal frem for andre. Dette stemmer overens med at der er stor spredning

i, hvilke informationskanaler arketype 1 anvender. Deltagerne har altså forskellige egenskaber de foretrækker og vælger derfor forskellige kanaler.

Flertallet (58%) af deltagere i arketype 1 benytter både bruger de officielle samt tredje parts kanaler (se søjlediagram D.17). Ca. en fjerdedel siger, at de kun bruger de officielle, og få siger de kun anvender tredje parts kanaler.

Endvidere er det karakteristisk for arketype 1, at de primært leder i versionsstyringssystemer, hvis de vil vide, hvem der har ændret i et artefakt. Næst efter følger mailing lister og issuetrackers (se søjlediagram D.22)

9.2.2 Hvor meget bruger arketyperne informationskanalerne?

Arketype 1 er den gruppe, der har været involveret i projekterne kortest tid (se søjlediagram D.11). Dette afspejles også i, at de kun har brugt informationskanalerne i maksimum 1 år (se søjlediagram D.12). Det er kun mailing lister de har brugt i mellem 1-2 år.

Af søjlediagram D.13 ses, at den informationskanal arketype 1 oftest bruger er mailing lister. Denne anvender de ugentligt. Fora bruger de månedligt og de resterende kanaler ligger midt i mellem, at de aldrig bliver brugt, eller at de bruges månedligt.

9.2.3 Hvordan oplever arketyperne informationssøgning?

Ud fra svarene på spørgsmål 12, hvor deltagerne bliver spurgt, om de skal søge i flere informationskanaler for at finde den information de søger, svarer 56% af arketyper 1 "yes" (se søjlediagram D.16).

Det er meget forskelligt, om deltagerne i arketype 1 synes, det er let at finde den information, de leder efter (se diagram D.19). Dog svarer over halvdelen (62%) 4-5 på en skal fra 1-5 hvor 5 er "easy". Af de deltagere, der i arketype 1 synes det er let at finde den information, de leder efter, er der flest, der begrundet det med, at de modtager hjælp fra andre deltagere, enten i form af svar eller andre der har oplevet lignende problemer. Samtidig er der også mange, der synes, det er let fordi der er mange informationer i projektet. Dette fremgår af følgende citater:

"because they are efficient in giving me the results i need. the active user community is often there to respond or may have already experienced/resolved the problem." 180, reader, Gaim

"They are pretty good at having updated info and at answering questions." 126, reader, Gaim

En deltager uddyber i interviewet hvorfor det er let at finde information. For det første beskriver deltageren, at det generelt er let at finde information, på grund af

dokumentation og FAQs. Yderligere forklarer deltageren det med, at det er venlige personer der findes i fora (se citat).

“Well, speaking of the Gaim project, general information is pretty easy to find, there are docs and FAQs, and friendly users in the forums.”
126, reader, Gaim, interview

Efterfølgende forklarer deltageren dog også, hvorfor han mener, det kan være svært at finde information. Det begrundes med, at det sommetider kan være svært at få pålidelige svar, fordi udviklerne ikke bruger den nødvendige tid. Herudover uddyber deltageren også problemet med de informationskanaler der er til rådighed, forskellige steder. Her beskriver deltageren problemer angående fora og værktøjer på Sourceforge, hvor han oplever navigations- og søgningsproblemer (se citat).

“What seems to be hard to get at times is reliable answers, as the developers don’t seem to take much time to do this.” og “The forums and tools at sourceforge aren’t very good for navigating or searching either, since they are both slow and hard to get good overviews in. That goes for all projects that uses SF’s standard tools, it’s a pain to navigate which makes one reluctant to do so, and more prone to giving up early.” 126, reader, Gaim, interview

De deltagere, der mener, at det er svært at finde den information de leder efter, begrundes det mest med manglende informationskanal eller på grund af dårlig funktionalitet i informationskanalen. Se følgende citater:

“Beacuse MPlayer do not offer a community forum (bulliten board) like most other open source projects I use (like VideoLAN and Xbox-MediaCenter)” 177, bug reporter, MPlayer

“Due to the ‘Mailing list’ archive, the search into it is difficult.” 168, reader, MPlayer

Arketype 1 karakteriserer sig også ved, at deltagerne ikke syntes, at det er blevet meget nemmere, at finde den information de søger efter, men at det dog er blevet lidt nemmere siden de startede på projektet. Dette kan ses ud fra, at deltagerne svarer, ud fra en skal på 1-5 hvor 5 er “easy”, 3 da de startede og at det kun er 3,75 nu (se søjlediagram D.23). Den begrundelse de fleste deltagere giver er, at det er erfaring, der gør, at det bliver nemmere med tiden at finde de informationer de søger. Deltagerne ved, hvor de skal lede efter informationen (se følgende citater).

“I know more now where to look and realize that MPlayer-devs don’t acceptt your questions unless you first thuroly research an subject/issue before you ask them anything, ...or you will get your head chopped off ;-/” 177, bug reporter, MPlayer

“learned the right places to look for things i need most often.” 156, reader, Gaim

At arketype 1 kun synes, at det er blevet lidt lettere, hænger muligvis sammen med, at arketype 1 er den type der har været kortest tid involveret i projekterne (1-2 år, se afsnit 7.1.5), og derved ikke har opbygget den samme erfaring som de andre typer.

Ud over erfaring er der også nogle af deltagerne, der påpeger, at det er fordi de kender og ved hvem andre deltagere i projektet er, og at disse som regel har den information de søger, og derved lettere kan finde den information, de har brug for (se citat).

“Now I know better who is who: I know that core developers’s mails are more informative than others’. I’m better at filtering uninteresting mails.” 219, reader, The Gimp

9.2.4 Hvordan foretrækker arketyperne informationsøgning udføres?

Det er karakteristisk ved arketype 1, at der er en stor overvægt (77%) af deltagerne (se diagram D.24), der helst vil søge information. De fleste af deltagerne begrundede dette med, at der som regel er andre, der har spurgt og fået svar på det samme problem.

“usually someone else has the solution already.” 188, reader, Gaim

“Most questions have been asked and answered. Finding previous answers is faster anyhow, and requires less work by me to gather proper information about my question/problem.” 178, bug reporter, MPlayer

Flere af deltagerne begrundede også, at de først vil søge information, med at de ikke vil genere andre når de skal finde information (se citat).

“When it is written in a clear form, why bother asking it? By asking I’m using someone’s (usually a developer) time which we all know is money (or code line or anything else:)” 185, reader, MPlayer

Nogle af deltagerne beskriver endvidere, at de ikke vil vente på svar og derfor vælger at søge efter informationen i stedet for at spørge (se citat).

“I search first because often, someone else has already asked the same question, and if they got an answer I don’t need to wait for one. Also, if there already is an answer, it is a bad thing too repeat the same question - it make for bad signal to noise ratio, and it wastes someone’s time.” 126, reader, Gaim

“A question means I have to wait/hope for a reply. A search can yield answers immediatly.” 214, reader, The Gimp

De fleste af deltagerne (79%) i arketype 1, vil gerne have mulighed for at følge op på den information de har fundet (se diagram D.25). Hvilket deltagerne stort set alle begrundet med at de kan bidrage med ekstra viden, eksempelvis til andre deltagere der leder efter det samme.

“To provide additional information to others looking for the same stuff, or to request more details on the specific topic.” 125, bug reporter, Gaim

“There might be a chance to add new information to support the existing information.” 216, reader, Gaim

9.2.5 Hvilken information søger arketyperne?

Arketype 1 er karakteriseret ved primært at søge awareness. Det ses af søjlediagram D.20, at 50% af deltagerne søger efter denne type information. Herefter følger forbedringer og manualer med henholdsvis 38% og 31% af deltagerne. Dermed er arketype 1 den type, som søger mest efter manualer. Dette kan hænge sammen med, at det også er den arketype, der har været involveret i projekterne kortest tid (se beskrivelse af arketype 1 i afsnit 7.1.5). De har derfor større behov for at se manualer til, hvorledes systemet fungerer. Derimod besidder deltagere, der har været involveret længere tid, allerede denne viden. Yderligere er arketype 1 karakteriseret ved, at de er dem, der søger mindst efter teknisk dokumentation. Dette kan skyldes, at de er den eneste arketype, som typisk ikke kigger i koden, idet de ikke bidrager med kode.

Følgende citater er eksempler på deltagere, der søger efter manualer:

“It is for practical reasons. How to print? How to get the drawing tablet working for my wife to use? Etc. :)” 149, reader, The Gimp, interview

“I have neither the talent nor intuition to be a great artist, composer or writer, and I beleive How-tos are written by the masters or gods, and if I slavishly follow their lead, I too will become great. (Oh well, you can always hope)” 124, bug reporter, The Gimp, interview

Det første citat understøtter ovenstående beskrivelse, idet deltageren søger manualer for at vide hvordan softwaren virker. Af andet citat fremgår det, at deltageren søger manualer for at lære nyt og blive bedre.

En deltager giver svar på, hvorfor han søger awareness information:

“Most of the time I’m already using a software or I am evaluating its features, so while I do this I come up with things I need or want to do, and thus I look for ways to do this. Sometimes there are plugins that can do this, people provide fixes and so on. Hunting for bugfixes is pretty much the same, I run into a problem of some kind and try to see if anyone knows of a solution.” 126, reader, Gaim, interview

Arketype 1 er den type med det mindste behov for awareness (se søjlediagram D.21). Grunden til dette kan være fordi arketype er den type, som er mindst involveret i projekterne. Derfor er deres behov for at vide hvem der foretager hvad også mindst. De er dog den arketype, der søger mest efter awareness information. Dette skyldes, at de søger efter information om nye releases. Dvs. information om hvad der sker i projektet. For arketype 1 er det derfor ikke vigtigt, at de nemt kan få overblik over, hvad de andre deltagere laver. I stedet er det relevant, at de nemt kan få overblik over nye releases og patches til softwaren.

En deltager svarer i interviewet:

“Mostly out of interesting. Who in the community is working on what projects? Perhaps they can help with other useful projects.” 149, reader, The Gimp, interview

På grund af nysgerrighed følger deltageren med i, hvad andre foretager sig, dvs. person awareness. Det er altså ikke for at kunne koordinere sit eget arbejde, men for at have indsigt i hvad der rører sig i projektet og yderligere, hvis han mangler hjælp.

Fra arketype 1 er der en deltager, der svarer på interview spørgsmål 1 angående, hvilken information deltagerne finder let og svær at finde. Deltagerens svar er:

“Well, speaking of the Gaim project, general information is pretty easy to find, there are docs and FAQs, and friendly users in the forums. What seems to be hard to get at times is reliable answers, as the developers don’t seem to take much time to do this. I’m not putting any special emphasis on if they should or not though, merely an observation. The forums and tools at sourceforge aren’t very good for navigating or searching either, since they are both slow and hard to get good overviews in. That goes for all projects that uses SF’s standard tools, it’s a pain to navigate which makes one reluctant to do so, and more prone to giving up early.” 126, reader, Gaim, interview

Deltager nr. 126 skriver, at han har let ved at finde generel information på grund af dokumentation, FAQs og venlige deltagere. Det han finder svært, er troværdige svar fra udviklerne, fordi de typisk ikke gider bruge tid på at svare. Han siger dog at han ikke ved og de burde bruge tiden på at svare. Endvidere nævner han, at han er utilfreds med de muligheder, Sourceforge stiller til rådighed for projektet. Han mener, at det ikke er godt til navigation, overblik og at søgningen er for langsom.

9.3 Opsummering arketype 1

9.3.1 Informationskanaler:

- Anvender alle informationskanaler undtagen nyhedsgrupper
- Primært mailing lister, fora eller andre kanaler som første sted
- Det de kan lide ved kanalerne er:
 - Chat: real time, føle sig en del af projektet, hjælpe andre
 - Forum: undgå at bruge eksternt program, persistente, søgemulighed, komfortabelt/vane
 - Mailing liste: fokuseret, lidt spam
- 58% bruger både officielle og tredje parts kanaler. En fjerdedel bruger kun officielle og få kun tredje part
- Anvender primært versionsstyringssystem for at vide hvem der har ændret et artefakt
- De har anvendt mailing lister i 1-2 år og de andre officielle kanaler i under 1 år
- Ugentligt anvender de mailing lister, fora månedligt, resten af informationskanalerne mellem månedligt og aldrig

9.3.2 Information:

- 50% søger awareness, dernæst forbedringer og manualer. Få søger teknisk dokumentation.
 - Awareness om releases og hvad der sker i projektet.
 - Manualer for at vide hvordan softwaren virker og lære nyt.
 - Kun få søger teknisk dokumentation da arketypen ikke bidrager med kode.
- Let at finde:
 - generel information pga. dokumentation, FAQs og venlige deltagere
- Svært:
 - om det er troværdige svar fra udviklerne
 - information gennem Sourceforge pga. dårlig navigering, manglende overblik og langsom søgning
- Arketype som har mindst behov for awareness information

9.3.3 Informationssøgning:

- 56% søger flere steder for at finde den information de søger efter
- 62% finder det meget nemt at finde information. Grundene er:
 - de modtager hjælp fra andre deltagere, ved svar eller allerede løste problemer
 - der er meget information i projektet
 - dokumentation og FAQs
- Finder det svært pga. manglende informationskanaler og dårlig funktionalitet i informationskanaler, herunder navigation og søgningsproblemer.
- Det er kun blevet lidt lettere for dem at finde information, de er gået fra 3 til 3,75
 - Grunden til det kun er lidt er, at de ikke har været så længe i projektet og derfor har ikke så meget erfaring.
- 77% af deltagerne foretrækker at søge efter information fremfor at spørge.
 - Vil ikke genere andre.
 - Der er typisk andre der har spurgt og fået svar på samme problemer.
 - Vil ikke vente på svar.
- 80% af deltagerne vil gerne have mulighed for at følge op på den information de har fundet.
 - For at bidrage med mere information

9.4 Arketype 2

9.4.1 Hvor søger arketyperne information?

Omkring halvdelen (48%) af arketype 2 vælger andre informationskanaler, som det første sted de søger, derefter følger versionsstyringsystemer som ca. en fjerde del vælger. Som andet valg vælger de primært chat eller mailing liste (se søjlediagram D.14 samt D.15)

På interviewspørgsmål nr. 2 angående, hvorfor de vælger deres primære informationskanal, er der tre svar fra arketype 2. Der er dog ingen af dem, som svarer angående versionsstyringsystemer, som er den officielle kanal arketype 2 foretrækker. Alligevel bruges et af svarene, idet den ene deltager er blandt de 3 udvalgte som er repræsentative for arketype 2. Deltageren beskriver, hvad det er ved chat, som gør, at han foretrækker at spørge der. De egenskaber deltageren nævner er:

“Chat (e.g. IRC): Is quite easy ask suggestion and send the code you are just developing. Since you don’t have the delay you may expect from email and forum (not to mention bts) you can get feedback faster.” 107, peripheral developer, MPlayer, interview

Deltageren foretrækker altså chat, fordi det er nemt og hurtigt, da der ikke er ventetid med at få svar.

På søjlediagram D.17 ses, at arketype 2 ligner arketype 1 med hensyn til, om de søger officielle eller tredje parts informationskanaler. Ved arketype 2 er fordeling mellem dem som anvender begge og dem, som kun anvender officielle dog mere ligelig. Men også i denne arketype er der færrest (9%) som kun anvender tredje parts kanaler.

Arketype 2 er karakteriseret ved, at størstedelen først kigger i versionsstyrings-system, dernæst mailing lister eller issuetrackers, hvis de vil vide, hvem der har ændret et artefakt (se søjlediagram D.22).

9.4.2 Hvor meget bruger arketypen informationskanalerne?

På søjlediagram D.12 ses en stor forskel på, hvor længe arketype 2 har anvendt de forskellige informationskanaler. Deltagerne har aldrig anvendt fora, mailing lister har de brugt i mellem 1-2 år og de resterende kanaler har de brugt i mindre end 1 år.

Videre ses det af søjlediagram D.13, at den hyppigst anvendte informationskanal er mailing listen, som de anvender næsten dagligt. Derefter følger chat, som de benytter ca. månedligt, versionsstyringsystem og issuetrackers som bruges lidt oftere end månedligt og nyhedsgrupper bruges månedligt. Fora anvender de ikke.

9.4.3 Hvordan oplever arketypen informationssøgning?

Som det er tilfældet med arketype 1, har omkring halvdelen, med en lille overvægt blandt dem der kun har brug for at søge et sted, af arketype 2, brug for at lede flere steder, for at finde den information de leder efter (se søjlediagram D.16).

I arketype 2 har over halvdelen svaret, at det er let at finde information. Dette kan ses ud fra at 65% har svaret 4 eller 5 på en skala fra 1-5, hvor 5 er “easy” (se søjlediagram D.19). De deltagere i arketype 2, der mener, at det er nemt at finde hjælp, begrundet det hovedsageligt med, at det er nemt at få hjælp fra andre der er online og igennem mailing lister. (se følgende citater).

“There are many active people who are often online to help.” 155, peripheral developer, MPlayer

“The developers are very accessible via chat, IM, and e-mail.” 104, peripheral developer, Gaim

“Lots of information is discussed in the mailing list and on irc.” 107, peripheral developer, MPlayer

En af deltagerne uddyber i det skriftlige interview, hvorfor det er nemt at finde information. Han beskriver, at der aldrig har været et spørgsmål, som han ikke kunne få svar på (se citat).

“Because I know the developers and chat with them daily, I don’t think there’s ever been a question I asked and didn’t get an answer on. Be it MPlayer’s internals, why a certain bug appears, how to use MPlayer to do something, MPlayer’s behavior, etc. etc.” 161, peripheral developer, The Gimp

I interviewet er der endvidere en deltager, der uddyber, hvorfor han bruger issuetracker. Grunden er, at han oplever, at søgningen bliver meget nemmere, da han har den nødvendige information til rådighed (se citat).

“Make searching much easier” og “I look at the bug and issue tracker on the site as I have most of the necessary documentation either printed or available from other sites.” 118, peripheral developer, The Gimp, interview

De deltagere, der har svaret, at det ikke er nemt for dem, at finde den information de søger, begrundet det med, at dokumentationen i nogle tilfælde er for lang, hvor andre af deltagerne mener, at den er kan være mangelfuld og kun dækker typiske situationer (se følgende citater).

“The Documentation is very big/long. The way it’s organised is questionable too.” 171, peripheral developer, MPlayer

“..The little documentation that exists is spotty and covers only the utopian “typical” situation.” 142, peripheral developer, The Gimp

Deltagerne i arketype 2 er karakteriseret ved, at de synes, at det er blevet nemmere at finde den information, de har brug for nu (3,96 på en skala fra 1-5, hvor 5 er “easy”), end da de blev tilknyttet til projektet (2,7) (se søjlediagram D.23). Begrundelsen for den forbedring er ifølge deltagerne, at de er blevet klar over hvor, de skal søge efter informationen og at nogle af informationskanalerne er blevet bedre at benytte (se følgende citater).

“the mailing list are more accessible and the irc channel is less crowded” 107, peripheral developer, MPlayer

“I now know of some sources that I didn’t know existed.” 171, peripheral developer, MPlayer

“I’ve grown more familiar with the source code, and now I know where to search for the things I need. I also started joining the project’s IRC channel.” 209, peripheral developer, The Gimp

I det skriftlige interview er der en deltager, der påpeger vigtigheden af at deltagerne får den rigtige information, samt at den er tilgængelig i det rette format, så både nye deltagere og erfarne deltagere kan opnå mere og bedre viden omkring projektet. Derved vil det også blive nemmere for dem at finde den information de leder efter.

“With the growth of The Gimp, it is important to get information in a place and format that makes it easy for new users to learn the basics and for the advanced users to learn how to be more efficient.” og “Many users will go to the www.gimp.org to either see about The Gimp or learn how to use some tools. I have been away from my Gimp usage for the last few months but will be hard into it within weeks. In the past there were broken links to sites that did not exist anymore. This looks better than it did then” 118, peripheral developer, The Gimp, interview

9.4.4 Hvordan foretrækker arketyperne informationsøgningen udføres?

Som det fremgår af søjlediagram D.24, er der 87% af arketype 2 der foretrækker at søge information. En af deltagerne påpeger at grunden, til at han søger information er, at han lærer meget mere, og at han kan finde løsninger på problemer. Herudover er der flere af deltagerne, der svarer, at de søger efter information, fordi de ikke vil besvare andre personer med spørgsmål (se følgende citater).

“You learn much more and you may end up with an already working solution.” 107, peripheral developer, MPlayer

“It’s generally rude to ignore the hard work people have put into writing documentation, by asking questions that are already answered in the documentation.” 172, bug fixer, MPlayer

En deltager nævner også, at der er så meget information, at det ville være en skam ikke at bruge den information. Se efterfølgende citat:

“There is so much searchable information that it would be a shame not to use it. And it is better to discuss whether the found info is still accurate than to ask other people to search for you.” 114, peripheral developer, The Gimp

Deltagerne bliver i spørgsmål 14 bedt om at beskrive, hvorfor de enten søger information i officielle eller tredje parts informationskanaler, eller begge. Hvor der er flest deltagere, der søger begge steder (se analyse af arketype 1 i afsnit 9.2). Begrundelsen deltagerne giver for dette er, at de kan finde informationen, de skal bruge mange forskellige steder (se citat).

“The other developers are on IRC, the mailing list is a kind of knowledge archive, the user and developers docs are on the official web site, and there a lots of contributed sources of information (tutorials, new plug-ins) on third party web sites.” 114, peripheral developer, The Gimp

I spørgeskemaundersøgelsen bliver deltagerne spurgt, om de har lyst til at følge op på den information de finder. Her er det et stor flertal (87%) af deltagerne i arketype 2, der ønsker denne mulighed (se søjlediagram D.25). Deltagerne begrundet dette med, at det sommetider kan være nødvendigt at få afklaret, om de har forstået, den information de har fundet er korrekt. Yderligere mener deltagerne, at der sommetider kan være behov for, at rette den information de har fundet, hvis de ikke synes den er korrekt (se følgende citater).

“Sometimes because it’s very technical, and I want to rephrase with my own words to see if I understood.”171, peripheral developer, MPlayer

“I have made the experience that sometimes something is just not explained correctly. While the facts might be complete, the way it is presented is not the "right one". Of course, other people might have a different opinion, so adding to something is useful.” 114, peripheral developer, The Gimp

9.4.5 Hvilken information søger arketyperne?

Søjlediagram D.20 viser, at arketype 2 primært søger teknisk dokumentation. Dernæst følger forbedringer, awareness og til sidst manualer.

Grunde til at arketype 2 søger teknisk dokumentation fremgår af de følgende citater:

“Work. I need it for doing one of my tasks. As I said, I need to automate some procedures within Gimp.” 118, peripheral developer, The Gimp, interview

“I’m a gentoo developer and one of my task (if not the main task actually) is fix issue in the software we provide.” 107, peripheral developer, MPlayer, interview

Begge deltagere søger altså teknisk dokumentation for at kunne udføre deres arbejde med at bidrage med kode til projektet.

Arketype 2 karakteriserer sig ved at ligge midt imellem arketype 1 og 3 i deres behov for awareness. Dette stemmer overens med, at de er involveret mere end arketype 1 men mindre end arketype 3.

En deltager svarer på, hvorfor han har et middel behov for awareness:

“What is happening on the project in the development phase isn’t much interest to me as a user. I do follow some but not all the issues. It may affect something that I am working on today when there is an upgrade or patch.” 118, peripheral developer, The Gimp, interview

Ud fra deltagerens svar fremgår det, at han ofte ser sig selv i en brugerrolle. I den rolle ser han ikke behov for at kende til, hvad der foregår i udviklingen. I de situationer, hvor han arbejder med noget i projektet, kan det dog påvirke ham, hvis der kommer updates. Det er således ikke person awareness, han er interesseret i, men generel awareness om, hvad der sker i projektet. Ligesom ved arketype 1 er det vigtigt for arketype 2, at de kan få overblik over projektet, updates og releases.

De følgende to citater er to deltageres svar på, hvilken information de finder let/svært. Den ene deltager har svært ved at finde teknisk dokumentation, og den anden har let ved at finde alt:

“Details on programming for a non-programmer. I have had to write some scripts and the necessary calls are not clearly documented for those that need to learn scripting and how to do things in less than 24 hours.

example: gimp_drawable_get_pixel

How do you get all the necessary data and set it up. Or working with layers in a script.

It would be nice if there were clear cut examples for each of the scripts for the NOOBS like me.” 118, peripheral developer, The Gimp, interview

“Because I know the developers and chat with them daily, I don’t think there’s ever been a question I asked and didn’t get an answer on. Be it MPlayer’s internals, why a certain bug appears, how to use MPlayer to do something, MPlayer’s behavior, etc. etc.” 161, peripheral developer, MPlayer, interview

Da arketype 2 er karakteriseret ved at søge teknisk dokumentation, kan vores system ved, at forbedre mulighederne for at søge teknisk dokumentation, forbedre informationssøgningen for arketype 2. Den anden deltagers grund til, at han har let ved at finde information skyldes, at han kender udviklerne af systemet. Han kan derved få hjælp af dem og spørge dem hvis han har problemer. Nogle af arketype 3 skriver, at det irriterer dem når folk hele tiden spørger i stedet for selv at finde svaret. Derved er det ikke umiddelbart en mulighed for andre deltagere.

9.5 Opsummering arketype 2

9.5.1 Informationskanaler:

- 48% vælger andre kanaler som første sted, derefter versionsstyringssystemer
- Som andet valg chat eller mailing liste
- Det de kan lide ved chat:
 - Chat: nemt og hurtigt
- Lige fordeling mellem dem som anvender både officielle og tredje parts kanaler og dem som kun anvender tredje parts kanaler. 9% anvender kun tredje parts kanaler
- Anvender versionsstyringssystemer, herefter mailing lister eller issuetrackers, til at se hvem der har ændret et artefakt.
- Har brugt mailing lister i 1-2 år og resten mindre end 1 år bortset fra fora, som de aldrig har brugt.
- Mailing lister benytter de dagligt og chat månedligt.

9.5.2 Information:

- Primært teknisk dokumentation. Dernæst forbedringer, awareness og til sidst manualer.
 - Teknisk dokumentation for at bidrage med kode.
- Behovet for awareness ligger imellem arketype 1 og 3.
- Behovet er for generel awareness.
 - Overblik over projektet
 - Updates
 - Releases

- Let:
 - Kender udviklerne og har derfor let ved alt information
- Svær:
 - teknisk dokumentation

9.5.3 Informationssøgning:

- Lille overvægt til dem der kun skal søge et sted efter informationen.
- 65% der synes det er let at finde information
 - Hjælp fra andre der er online
 - Nemt at benytte mailing lister
 - Der er den nødvendige information
- Svært at finde information i dokumentation
 - Dokumentation kan være for lang
 - Dokumentation er mangelfuld
 - Dokumentation indeholder kun typiske problemer
- De mener det er blevet en del nemmere at finde information nu, de er gået fra 2,7 til 3,96. Dette skyldes:
 - de er blevet klar over, hvor de skal søge
 - nogle af informationskanalerne er blevet bedre at benytte
- 87% ønsker at søge efter informationen.
 - De søger efter information for ikke at besvare andre
- 87% vil gerne have muligheden for at følge op på den information de finder

9.6 Arketype 3

9.6.1 Hvor søger arketyper information?

Knap halvdelen (43%) af arketype 3 svarer, at den primære informationskanal, de anvender, er mailing liste. Omkring en tredje del vælger versionsstyringsystemer. Som anden prioritet vælger mere end halvdelen chat. Disse resultater fremgår af søjlediagram D.14 samt søjlediagram D.15.

Der er kun et svar angående, hvad deltagerne kan lide ved deres foretrukne informationskanal. Svaret lyder:

“Code repository (e.g. CVS): What I like about it is the fact, I have the source code for the entire application on my hard disk and I can search it, browse it and look for information.” 79, active developer, The Gimp, interview

Denne deltager foretrækker altså at have koden lokalt og kunne søge i den. Dette stemmer også overens med, at deltageren hører til arketype 3 og primært søger efter teknisk dokumentation (beskrevet i afsnit 7.1.5).

Arketype 3 fordeler sig således, at lidt over halvdelen (57%) anvender både officielle og tredje parts informationskanaler mens lidt under halvdelen (43%) kun bruger de officielle. Der er ingen fra arketype 3, som kun anvender tredje parts kanaler. Se søjlediagram D.17 for illustration af dette.

En deltager skriver, at han bruger officielle kanaler, da han oplever, at de dækker hans behov (se citat).

“Because I find them covering what I need.” 79, active developer, The Gimp

Kun få deltagere fra arketype 3 har svaret på, hvor de søger, hvis de vil vide hvem, der har ændret i noget. Af dem som svarer, siger de fleste versionsstyringssystemer, og nogle få siger chat og andre kanaler (se søjlediagram D.22). Arketype 3 er ud fra dette karakteriseret ved kun at fordele sig på tre kanaler, når de søger efter den slags information.

9.6.2 Hvor meget bruger arketyperne informationskanalerne?

Siden arketype 3 i gennemsnit har været involveret længere tid end de andre to typer, er tiden, de har anvendt de forskellige informationskanaler, også længere end i de to andre typer (se søjlediagram D.12). Der er fire kanaler, som arketype 3 har anvendt i 2 år og op efter. Disse er: mailing liste, chat, versionsstyringssystemer og issuetrackers. Nyhedsgrupper samt fora har de anvendt i under et år.

Af søjlediagram D.13 fremgår det, hvor ofte arketype 3 anvender de forskellige kanaler. Arketype 3 anvender chat og mailing lister dagligt, versionsstyringssystemer og issuetrackers ugentlig, fora månedligt og nyhedsgrupper knap en gang om måneden.

Ud fra dette konkluderes, at arketype 3 er karakteriseret som den type, der har anvendt informationskanalerne i længst tid, og ydermere også er den type, som anvender kanalerne oftest.

9.6.3 Hvordan oplever arketyperne informationsøgning?

Hvor det i arketype 2 og arketype 1 er ca. 50%, der har brug for at lede i flere informationskanaler for at finde den information, de søger, er der i arketype 3 en klar overvægt, nemlig 70%, der har brug for at søge i flere informationskanaler (se søjlediagram D.16). Det kan heraf konkluderes, at det er en karakteristik ved arketype 3, at typen har brug for at lede efter information flere steder for at finde den information, de søger.

Arketype 3 er karakteriseret ved, at de finder det let at finde information, de har brug for. Dette kan ses ud fra, at ca. 70% har svaret 4 på en skala fra 1-5, hvor 5 er "easy" (se diagram D.19). Begrundelsen for dette er baseret på de deltagere, der har svaret på spørgsmål 16. En af deltagerne svarer f.eks., at der er masser af information, der er mange personer, der er villige til at hjælpe, og at koden er let at forstå (se citat).

"There's plenty of information, the people are helpful and the code is easy enough to understand." 79, active developer, The Gimp

Dette underbygges af en deltageren, der også svarer, at det er let for ham, fordi han kan forstå koden (se citat).

"Because I understand the code." 121, active developer, MPlayer

Arketype 3 er også karakteriseret ved, at deltagerne mener, at det er blevet nemmere at søge information nu, end da de tilsluttede sig projektet. Dette kan ses ud fra, at deltagerne i starten syntes, det var 3,25 og det nu er 4,5 (se søjlediagram i bilag D.23) ud fra en skala på 1-5, hvor 5 er det letteste. Begrundelsen for, at de mener, det er blevet nemmere ligger i erfaring og større mængde af information. Erfaring kommer fra, at de har fået mere viden omkring værktøjerne de bruger, samt at de bedre kan forstå kode. Se efterfølgende to citater:

"Because I have a better idea of all the tools we use to communicate." 204, active developer, Gaim

"I understand the code better." 121, active developer, MPlayer

En af deltagerne uddyber, at det er blevet lettere at finde information på grund af en stigende mængde af dokumentation, som f.eks. kan hjælpe deltagere med at blive involveret i projektet (se citat).

"We provide more and more documentation, which are describing how the people can get involved with the project." 207, active developer, The Gimp

9.6.4 Hvordan foretrækker arketyperne informationsøgningen udføres?

Største delen af arketype 3 foretrækker at søge efter information. Dette kan ses ud fra, at 70% af deltagerne har svaret "search" i spørgsmål 23 (se søjlediagram D.24). Begrundelsen for at deltagerne ønsker at søge information, i stedet for at spørge ligger i, at de ikke vil besvare andre med deres problemer (se citat).

"Because, most terms are already answered and I don't want to bother others with my (mostly) trivial questions I can answer myself using a search engine." 207, active developer, The Gimp

En af deltagerne siger i samme situation, at når han ikke vil have at andre besvære ham uden at søge selv først, vil han heller ikke besvare andre, før han selv har søgt (se citat)

"I hate when people bother me with trivial questions without searching first, so if I need information, I search first and only ask if I can't find or understand it, which is almost never." 121, active developer, MPlayer

Ud over at undgå at besvare andre er der også en deltager, der begrundet, at han mener, det er hurtigere og mere personligt at søge først (se citat). Vi fortolker, at han med personligt mener, at han ved at søge ikke gør det, han søger, offentligt.

"Because it's quicker and more personal." 204, active developer, Gaim

Omkring en mulighed for at følge op på den information deltagerne har fundet, er der et stort flertal (70%) der har svaret ja. En Deltager vil gerne have en mulighed for, at hjælpe andre der måske har samme problem, og en deltager vil gerne bidrage med sin personlige viden om emnet (se følgende citater).

"In case things are unclear, it might be helpful to me or others to followup on the subject." 121, active developer, MPlayer

"Sometimes it would be nice, to add personal descriptions to it, for other people who are looking for the same information." 207, active developer, The Gimp

9.6.5 Hvilken information søger arketyperne?

Arketype 3 søger primært teknisk dokumentation. Dernæst er det ligeligt fordelt om det er forbedringer eller awareness. Ingen i arketype 3 søger ingen manualer.

De er dermed modsat arketype 1. Dette stemmer også overens med, at de har to forskellige roller i projekterne. Arketype 3 bidrager med kode og søger derfor efter teknisk dokumentation. Arketype 3 har ikke behov for manualer, da de har deltaget i projekterne mere end 2 år (se beskrivelse af arketype 3 i afsnit 7.1.5) og derfor kender til, hvordan systemet fungerer.

En deltager svarer i interviewet på hvorfor han søger teknisk dokumentation:

*“I’m just looking for issues to tackle, and often need to know how to do a certain thing in GIMP or gtk+. So I’m using the API documentation.”
79, active developer, The Gimp, interview*

Idet deltageren er arketype 3, antages det, at “issues” og “certain things” er programmeringsmæssige ting, deltageren vil løse eller se, hvorledes bliver gjort. Antagelsen underbygges endvidere af at gtk+ er grafisk toolkit rettet mod udviklere.

Ud fra søjlediagram D.21 ses, at arketype 3 er den type, som har størst behov for awareness information. Grunden til dette er formodentlig, at de også er den type, som er mest involveret i projekterne og derfor har behov for at vide, hvem der foretager hvad.

Det virker modstridende, at kun 29% af arketype 3 søger awareness information samtidig med, at de svarer, at de er meget interesseret i awareness. Grunden kan være, at de, når de søger øvrig information, eksempelvis teknisk dokumentation eller information omkring forbedringer, opnår awareness som et biprodukt af deres søgning. Eksempelvis anvender arketype 3 mailing lister dagligt, og herved kan de opretholde awareness om, hvad andre laver og hvad der sker i projektet. Vi antager derfor, at deres behov for at søge awareness information er dækket gennem deres øvrige søgninger.

I det følgende citat beskriver en deltager, hvilke typer information han finder let og svært at finde:

“Usually the information that is easy to find is how to use the program. Information that is medium to find is exactly what the code does. Information that I found hard to find is how the program behaved in the past, or how exactly it’s supposed to behave.” 79, active developer, The Gimp, interview

Deltageren synes altså, at manualer er lette at finde, og det som han finder middel eller svært at finde, er teknisk dokumentation dvs. information om koden og systemet. At det er let for ham at finde ud af, hvordan systemet fungerer, hænger formodentlig sammen med, at han tilhører arketype 3 og derfor har været involveret i projektet længe.

9.7 Opsummering arketype 3

9.7.1 Informationskanaler:

- 43% benytter mailing lister som første sted de søger. En tredjedel versionsstyringssystemer.
- Som anden prioritet bruger mere end halvdelen chat.
- Det de kan lide ved versionsstyringssystemer:
 - have koden lokalt og søge i den
- 57% bruger både officielle og tredje parts informationskanaler. 43% benytter kun officielle informationskanaler. Ingen som kun anvender tredje parts kanaler.
- De få der har svaret bruger versionsstyringssystemer til at se hvem der har ændret et artefakt.
- De har benyttet mailing lister, chat, versionsstyringssystemer samt issuetrackers i mere end 2 år.
- Chat og mailing lister anvender de dagligt. Versionsstyringssystemer og issuetracker bruger de ugentligt.

9.7.2 Information:

- Søger primært teknisk dokumentation. Søger ligevægtigt forbedringer og awareness og ingen, der søger manualer.
 - Teknisk dokumentation da de bidrager med kode
 - Ingen manualer da de kender systemet
- Arketype med størst behov for awareness skyldes, at de er mest involveret i projekterne.
- Let:
 - manualer da de har været længe i projektet og ved hvor de befinder sig
- Svært:
 - hvad koden præcis gør
 - teknisk dokumentation

9.7.3 Informationssøgning:

- 70% søger flere steder for at finde den information de søger.
- 70% svarer 4 og dermed at det er let at finde information. Dette skyldes:
 - meget information i projektet
 - mange personer, der er villige til at hjælpe
 - koden er let at forstå
- Det er blevet nemmere at søge information i den tid de har været involveret. De er gået fra 3,25 til 4,5. Grundene er:
 - erfaring da de har fået mere viden om værktøjerne de bruge og de bedre kan forstå koden
 - stigende mængde dokumentation.
- 70% foretrækker at søge efter information fremfor at stille spørgsmål. Dette er pga.:
 - de vil ikke besvare andre med spørgsmål
 - det er hurtigere
 - mere personligt
- Alle søger i de officielle informationskanaler. Nogle søger begge steder.
 - anvender de officielle da det dækker behovet
- 70% Svarer ja til at have muligheden for at følge op på et emne. Grundene er:
 - mulighed for at hjælpe andre
 - mulighed for at bidrage med sin personlige viden om emnet

10 Eksisterende systemer

Følgende kapitel indeholder en analyse af eksisterende systemer. Først analyseres de informationskanaler, der anvendes i de Open Source projekter som benyttes i empiriindsamlingen. Dernæst analyseres eksisterende CSCW systemer. Analysen har til formål at finde aspekter ved systemerne, der kan anvendes til udarbejdelse af designkrav.

Formålet med denne analyse er at identificere aspekter, eksempelvis funktionalitet og design, ved forskellige informationskanaler. Dette inddrages sammen med arketypernes karakteristika 9.1 for at opstille designkrav. Analysen resulterer i en liste af positive og negative aspekter for hver informationskanal. Denne liste inddrages derefter i designet, og sammenholdes med de krav der er udarbejdet på baggrund af arketypernes karakteristika.

I analysen ligges speciel vægt på, hvorledes awareness information formidles og repræsenteres i informationskanalerne, samt hvorledes søgning i group memoryen foretages og repræsentation af resultater fra denne søgning.

De informationskanaler, der inddrages i analysen, er henholdsvis de informationskanaler, der benyttes i de Open Source projekter, der indgår som cases i dette projekt. Hvor beskrivelsen i afsnit 2.2 er generel for informationskanalen, inddrages de specifikke informationskanalers design og funktionalitet i denne analyse.

Udvælgelsen af CSCW systemer beskrives i det følgende afsnit.

10.1 Udvalgelse af eksisterende CSCW systemer

Vi vælger at analysere to CSCW systemer. Formålet med at analysere disse CSCW systemer er at kunne inddrage alternative løsninger i designkravene. De systemer, der inddrages, skal specifikt have til formål at formidle group memory eller awareness information. Der er udvalgt to systemer; Hipikat, et system der foreslår udviklere relevante emner i forhold til, hvad de arbejder med, fra group memoryen, og Jazz, et system der formidler awareness i udviklingsprocessen. Begrundelsen for at vælge ud fra group memory og awareness er, at det er disse begreber projektet fokuserer på.

10.2 Eksisterende systemer i projekterne

I det følgende analyseres de informationskanaler som Open Source projekterne benytter (se tabel 5.2).

10.2.1 Website

Vi anser et Open Source projekts website som det logiske startsted for personer, der ønsker at få mere information omkring projektet. Der er dog ikke data fra spørgeskemaet til at støtte dette (pga. mangler i spørgeskemaet, se afsnit 6.2.8), men enkelte af arketyperne indikerer i deres besvarelser denne antagelse:

“Many users will go to the www.gimp.org to either see about The Gimp or learn how to use some tools.” 118, peripheral developer, The Gimp

Når personerne besøger websitet, er det muligt for dem at finde henvisninger til flere officielle informationskanaler, hvis projektet er i besiddelse af disse. Den følgende deltager påpeger, at de officielle websites indeholder henvisninger til *alt* relevant dokumentation, mens alle andre vigtige informationskanaler dukker op som nogle af de første på en Google søgning:

“The official web sites links to all relevant docs, and all major information sources have excellent Google rankings.” 114, peripheral developer, The Gimp

Vi mener, at det er et positiv aspekt ved et website, at det indeholder henvisninger til andre informationskanaler. Idet det således er muligt for personer, der besøger websitet at opdage disse. Alternativt må personerne aktivt søge efter informationskanaler, f.eks. på Google. Websites er derfor ideelle udgangspunkter til at give

deltagere i Open Source projekter bevidsthed om hvilke informationskanaler, der eksisterer i projektet.

Der er dog også nogle problemstillinger forbundet med websites. Der beskrives eksempelvis i afsnit 5.4, at The Gimp projektet henviser til et andet website, som dog ikke indeholder nogen information. En af The Gimps udviklere påpeger ligeledes, at hvis henvisningerne på websitet ikke virker, er det meget frustrerende:

“Broken links on a WWW site can be very frustrating.” 118, peripheral developer, The Gimp

Hvis en henvisning på websitet ikke længere fungerer, betyder det, at en del af group memoryen ikke længere er tilgængelig. Sker dette, kan hele områder af group memoryen forsvinde. Det er derfor vigtigt at vedligeholde links til de forskellige former for information samt informationskanaler, der stilles til rådighed for deltagerne.

Den samme udvikler påpeger yderligere, at hvis der er henvisninger til for mange forskellige informationskanaler, kan det blive svært at finde den information, man leder efter:

“To many avenues can make it hard to find that one piece of information you are looking for. Looking at the WWW site now, it is much better than the last time I really looked at it. The layout is much better designed.” 118, peripheral developer, The Gimp

Dette illustrerer den problemstilling, dette projekt behandler. Jo flere valg af informationskanaler, der stilles til rådighed, jo mere spredt bliver deltagerne samt information. Vi mener derfor, at det er vigtigt med ét centralt udgangspunkt for informationssøgningen. Dette betyder dog ikke, at de eksisterende informationskanaler skal afskaffes, blot at deltagerne skal have mulighed for at søge i et system i stedet for flere.

Den ovennævnte udvikler peger desuden på, at et bedre layout har gjort det lettere at finde information. Ligeledes henviser en anden deltager fra The Gimp projektet til, at den officielle website for projekter er blevet bedre organiseret:

“because the official gimp site has more info and is better organized now.” 198, bug reporter, The Gimp

Ud fra ovenstående diskussion af websites kan man konkludere, at websites har en række positive og negative aspekter forbundet med sig. Det positive er, at websites kan fungere som udgangspunktet for en deltager til at finde mere information. Et negativt aspekt ved websites er, at få hjemmesider er opbygget på samme måde, og der er derfor normalt nødvendigt at lede på flere websites for at finde det website,

der indeholder den information man søger. Desuden hindrer ustrukturerede og ikke brugervenlige websites deltagerne i at finde den information, de leder efter. Vi er derfor af den opfattelse, at brugervenlighed samt en velstruktureret opstilling af informationen er vigtig, når mange informationer skal formidles. Yderligere mener vi også, at det er vigtigt, at Open Source projekterne tilbyder ét interface, altså ét udgangspunkt til informationssøgningen, så information kan tilgås fra samme sted. I det følgende beskrives et website, der har samlet flere forskellige informationskanaler på et website.

Sourceforge.net

Sourceforge.net er et website, der tilbyder adgang til en række forskellige informationskanaler, som Open Source projekter kan benytte. At Sourceforge samler flere forskellige informationskanaler, er specielt interessant i forhold til vores problemstilling idet, at Sourceforge tilbyder et samlet interface til informationskanalerne. Dog er det stadig muligt at benytte informationskanalerne uden om Sourceforge, eksempelvis kan man tilgå mailing listen via sit normale e-mail program.

Gaim er et af de projekter der benytter faciliteterne stillet til rådighed af Sourceforge. En del af deltagerne i Gaim påpeger nogle problemer ved måden Sourceforge fungerer på.

En af deltagerne beskriver, at der findes meget information på Sourceforge, men søge funktionaliteten er ikke optimal:

“Source forge has a lot of information available, however there search facility is not all it could be.” 216, reader, Gaim

Gaim benytter sig af mange af de informationskanaler, som Sourceforge tilbyder: fora, issue- og bugtracker, mailing lister samt versionsstyringssystemer. Det er således vigtigt for deltagerne i projektet at kunne tilgå de informationer, der eksisterer i de informationskanaler, og hvis Sourceforages søgefunktionalitet ikke fungerer optimalt, besværliggør det adgangen til awareness information og group memory. En anden Gaim deltager påpeger netop dette:

“The forums and tools at sourceforge aren’t very good for navigating or searching either, since they are both slow and hard to get good overviews in. That goes for all projects that uses SF’s standard tools, it’s a pain to navigate which makes one reluctant to do so, and more prone to giving up early.” 126, reader, Gaim

Dette er et meget negativt aspekt ved Sourceforge. Hvis en informationskanal hindrer deltagerne i at opnå deres mål, er der en chance for, at deltagerne opgiver at

benytte denne informationskanal og skifter over til at benytte en anden. Resultatet heraf er, at deltagerne bliver spredt ud over flere informationskanaler, hvilket betyder, at informationen bliver spredt.

Der er dog også positive aspekter ved Sourceforge, nemlig at de tilbyder et konsistent interface til alle de projekter der bliver hostet der. Samtidigt samler de tilgange til mange forskellige informationskanaler, uden at hindre at man kan tilgå dem via andre systemer.

Udover at deltagerne i Gaim påpeger, at Sourceforgeries faciliteter generel ikke er optimale, beskriver en af udviklerne i Gaim projektet også, at der er problemer med Sourceforgeries issue tracker:

“The sourceforge issue tracker is a pain.” 119, peripheral developer, Gaim

At Sourceforge ikke fungerer optimalt, er ifølge vores opfattelse et stort problem, idet der pr. 05-juni-05 hostes 101.021 Open Source projekter på Sourceforge. Hvis bare et en lille del af disse projekter oplever de samme problemer som beskrevet ovenover, betyder det, at kommunikationen i disse projekter lider, og ikke fungerer optimalt. Vi mener, at dette er et meget negativt aspekt ved Sourceforge.

10.2.2 Fora

Gaim er det eneste af de projekter, der undersøges i dette projekt som har et officielt forum (se tabel 5.2), og dette befinder sig på Sourceforge. I beskrivelsen af fora (se afsnit 2.2) nævnes, at fora typisk er emne inddelt, diskussioner er inddelt i tråde, samt at fora også indeholder dedikeret søgning. Gaims forum er ikke inddelt i flere områder efter emne, der er kun, hvad der kaldes “Open Discussion” (se [Gaim Open Discussion, 2005]). Da diskussionen ikke er emne inddelt, omhandler diskussionstrådene mange forskellige emner, eksempelvis forslag til forbedringer (*Suggestions (msn-like)*) og fejl (*BUG? ã é í and others is not allowed...*). Idet diskussionerne ikke er emne inddelt besværliggøres søgning, da det ikke er muligt at begrænse denne til projekt specifikke områder som eksempelvis diskussioner omkring forbedringer.

Søgefunktionen i forumet giver mulighed for at søge efter en tekststreng. Hvis der søges efter “thesis” (ordet er indeholdt i det indlæg omkring vores undersøgelse, der er placeret på Gaims forum, se bilag A) returneres vores indlæg som det første. Ud over kun at returnere resultater fra forumet listes der også en række “Web Search Results” fra Yahoo!. I tilfældet med søgestrengen “thesis” findes forskellige websites der omhandler “thesis”. Umiddelbart virker disse resultater som unødvendig, i det de på ingen måde omhandler informationer, der er at finde i forumet. Man kan dog antage, at disse resultater kan henvise deltagerne til alternative information omkring det, der søges efter.

På trods af at søgefunktionaliteten fungerer tilfredsstillende, er det dog stadig vigtigt med emne inddeling. Ikke alene bliver søgningen, ved hjælp af søgefunktionen, fokuseret, men det antages også at det er lettere for en deltager at browse sig frem til relevant information. Dette ses som et positivt aspekt ved fora, at der er søgefunktionalitet integreret, samt at diskussionerne kan inddeles i emner.

Hvis en deltager er registreret på Sourceforge, kan denne logge sig ind og derved få adgang til en række muligheder i forumet, der ikke er tilgængelig for en ikke registreret deltager. Eksempelvis kan en registreret deltager se forskel på indlæg denne har set tidligere og nye. De nye indlæg i forumet bliver repræsenteret ved, at emnet står med en fed skrifttype. Ligeledes har deltagerne mulighed for at overvåge forumet via linket "Monitor this Forum", hvorved deltageren modtager en e-mail hver gang en ny tråd oprettes eller besvares. Grunden til, at en sådan funktionalitet eksisterer, skal nok findes i forskellen på hvorledes information tilgås i e-mails og websites/webbaseret fora. E-mail er et push media, hvilket betyder, at informationen bliver "skubbet" ud til deltagerne, og det kræver ikke nogen aktivitet fra deres side for at tilgå informationen. Forum adskiller sig fra dette ved, at det kræver, at deltagerne aktivt går ind og puller informationen. Dette betyder, at en deltager skal besøge forumet ofte for at se, om der er lavet nye indlæg. Hvis en deltager har tilmeldt sig "Monitor this forum", bliver deltageren gjort opmærksom når, der er en ændring i forumet via en e-mail.

Vi er af den opfattelse, at kombinationen af pull og push media er et yderst velegnet kombination. Muligheden i at automatisk overvåge eksempelvis et forum, gør at en deltager kan fokusere på andre ting, og først afbryde dette når der er sket en ændring. Det er også et positivt aspekt ved fora, at indholdet kan tilpasses til den individuelle deltager, eksempelvis ved at gøre forskel på de indlæg deltagerne har læst og de indlæg deltageren ikke har læst.

Der er flere positive aspekter ved fora. Idet det er webbaseret, er det let at tilgå, da der ikke kræves ekstra programmer ud over en webbrowser. Dette betyder igen, at fora er tilgængelige for et stort antal deltagere. Den følgende deltager påpeger, at MPlayer ikke er særlig brugervenlig og ikke tager højde for den gennemsnitlige bruger ved netop ikke at tilbyde et forum:

"Beacuse MPlayer is still not very user-friendly and do not accomodate for avarage users on its official means of commuciation, ...they really need a forum." 177, bug reporter, MPlayer

Ud af dette kan der udledes, at deltageren mener, at fora er, hvad gennemsnitlige deltagere ønsker. Det følgende citat belyser en grund til, hvorfor fora virker populære:

“(Forum): A lot of it is probably force of habit - I’m for some reason more comfortable with forums than for instance mailing lists or IRC. I think that forums probably represent a good middle ground there, easy to post and reply without having to fire up an external program (as with mail) at the same time searchable and persistent (not volatile like IRC). So, for any project that provides them, I tend to use the forums first and foremost if it is not apparent that some other channel is much more efficient.” 126, reader, Gaim

Denne deltager beskriver fora som en god mellemting mellem mailing lister og chat. Yderligere påpeger deltageren flere interessante ting; det er let at lave indlæg og besvare indlæg, og der behøves ikke et “eksternt” program til dette som der er brug for til e-mail. Ud fra dette kan man lede den antagelse, at deltageren ikke opfatter webbrowseren som et eksternt program.

10.2.3 Mailing list

Mailing lister er en af de mest brugte informationskanal, hvis ikke den mest brugte i de Open Source projekter der er undersøgt i dette projekt. Ud fra denne observation kan det antages, at mailing lister har nogle karakteristika, der gør, at de er så anvendte i Open Source projekterne. En antagelse er, hvis størstedelen af et projekts kommunikation forgår på mailing listen opsøger nye deltagere også denne informationskanal. Begrundelsen for dette er måske, på grund af at mailing lister indeholder store mængder group memory, altså store mængder viden omkring projektet. Herudover benytter både udviklere og brugere mailing listen til at få awareness information. To deltagere fra MPlayer projekter indikerer yderligere grunde til, hvorfor mailing lister bliver benyttet:

“The documentation is very comprehensive, the experts and developers actively post on the users list, with gold nuggests of insight.” 172, bug fixer, MPlayer

Selvom MPlayer har en dedikeret mailing liste kun til udviklere, tager disse sig tid til også at læse og poste på brugernes mailing liste. Dette betyder, at viden kan formidles meget hurtig, fra dem der typisk ved mest om projektet, nemlig udviklerne, til den deltager der spørger. Endvidere påpeger den følgende udvikler, at der diskuteres meget på mailing listen:

“Lots of information is discussed in the mailing list and on irc.” 107, peripheral developer, MPlayer

At et projekts kommunikation forgår offentligt er meget vigtigt for formidlingen af awareness, idet det er muligt for alle deltagerne i et projekt at “overhøre” hvad

der forgår i projektet. Dette betyder af deltagerne på mailing listerne får en masse indirekte awareness information, bare ved at se hvem der taler omkring hvad.

Udviklerne forventer, at deltagerne har søgt andre steder inden de spørger i chat, hvilket nok er grunden til, at så mange først i anden omgang benytter chat til at finde information. Vi konkluderer derfor, at det er vigtigt at kunne "overhøre" diskussioner, da deltagerne kan opnå indirekte awareness information uden ekstra arbejde. Dette koncept med at opnå awareness information, mener vi også, gælder for flere af de andre informationskanaler, eksempelvis chat og fora.

Som nævnt foregår kommunikationen på mailing listerne offentligt. Yderligere arkiverer de fleste Open Source projekter også alt kommunikation, der foregår i mailing listerne i offentlige arkiver, eksempelvis de tre projekter der benyttes i dette projekt [Gaim mailing liste arkiv, 2005][MPlayer mailing liste arkiv, 2005][The Gimp mailing liste arkiv, 2005]. Dette betyder, at deltagerne har adgang til group memory, der indeholder diskussioner, awareness information, løsninger til problemer samt meget andet. For at kunne tilgå group memoryen effektivt, skal der dog være en velfungerende søgefunktionalitet. En deltager i MPlayer projektet nævner, i denne forbindelse, at søgningen i mailing liste arkivet er svær:

"Due to the 'Mailing list' archive, the search into it is difficult..." 168, reader, MPlayer

MPlayer tilbyder på deres liste over mailing lister (se [MPlayer mailing liste arkiv, 2005]) følgende muligheder: *subscribe*, *archives*, *search*, *search on gmane*. Den første mulighed tillader deltagere at tilmelde sig mailing listen, og de tre sidste muligheder giver adgang til arkiverne. Den første af disse, *archives* indeholder ikke nogen søge funktion, dog har den inddelt alle e-mails sendt til mailing listen, efter hvilken måned og år de er sendt i. Deltagerne har så muligheden for at sortere alle e-mails i en måned efter diskussionstråd, emne, forfatter eller dato. En yderligere mulighed der tilbydes er at download alle e-mails, så man kan søge i dem lokalt. *Search* og *Search on gmane* giver begge mulighed for at søge efter en tekst streng i arkivet. En hurtig søgning på "thesis" (ordet er indeholdt i den mail der er udsendt til Open Source projekterne, i forbindelse med vores undersøgelse, se bilag A) returnerer den e-mail, der er udsendt i forbindelse med spørgeskemaundersøgelsen, som det første hit.

Ud fra den ovenstående diskussion kan det konkluderes at nem og effektiv tilgang til group memoryen, der findes i mailing list arkiverne er vigtig, idet en stor del af kommunikationen i Open Source projekterne foregår der.

Det er derfor et positivt aspekt ved mailing lister, at de er offentlige og arkiveres offentligt. Yderligere er det positivt, at alle deltager har adgang til de andre mailing lister, eksempelvis at alle deltager har adgang til udviklernes mailing liste.

10.2.4 Chat

Både Gaim og MPlayer tilbyder officielle chat kanaler (se tabel 5.2). På samme måde projekterne opdeler deres mailing lister i en for udviklerne, og en for brugerne, har MPlayer to chat kanaler, henholdsvis *#mplayerdev* og *#mplayer*. Gaim har ikke en opdeling på denne måde, men en generel, *#gaim* og en specifik til WindowsTMudgaven af Gaim, *#wingaim*. Begge projekter har valgt at hoste deres chat kanaler på `irc.freenode.net`.

Chat kanalerne tilgås via et program, der understøtter IRC protokollen, eksempelvis XChat. Hver IRC kanal har et emne, der eksempelvis kan indeholde kanalens navn, informationer omkring hvad der diskuteres på kanalen. Udover emnet for kanalen findes også en liste af alle deltagerne på kanalen. Enkelte af deltagerne kan fungere som moderatører og have muligheden for at smide deltagere af kanalen. Flere samtaler kan foregå samtidigt på den enkelte kanal, og det er samtidigt muligt at to deltagere oprette en en-til-en forbindelse, så de kan tale privat.

Listen over deltagere på kanalen giver awareness information om, hvem der er tilstede, og ved at overhøre samtaler på chat er det muligt at få indirekte awareness information.

Chat er realtime kommunikation, og dette adskiller chat fra de andre informationskanaler (se tabel 2.1). Der er derfor muligt hurtigere at få feedback på spørgsmål. Dette underbygges eksempelvis af følgende citat:

“(Chat (e.g. IRC)): Is quite easy ask suggestion and send the code you are just developing. Since you don’t have the delay you may expect from email and forum (not to mention bts) you can get feedback faster.” 107, peripheral developer, MPlayer

I afsnit 2.3.3 omkring awareness i konteksten af Open Source beskrives, hvorledes chat gør det muligt at opnå en vis mængde “just in time awareness”, hvor deltagerne hurtigt og med lethed kan få afgjort nogle ting. Der er dog også nogle problematikker med de områder man kan diskutere i chat. En deltager påpeger at det bliver problematisk at diskutere større stykker kode, da man kun kan skrive en linje ad gangen (se nedenstående citat). Til diskussion af større stykke kode må man antage at e-mail eller lignende informationskanaler derfor er mere velegnede.

“(Chat (e.g. IRC)): I like IRC because it’s realtime chat, and an actual conversation. With almost any other communication, you talk in ‘chunks’. IRC resembles most a real life conversation. It’s disadvantage is it’s hard to talk about large pieces of code or anything of the sort. you can only talk one line at a time.” 161, peripheral developer, MPlayer

I forbindelse med formidling af awareness information giver det følgende citat indsigt i, hvorledes en deltager i The Gimp projektet oplever chat:

“(Chat (e.g. IRC)): I like chat because it’s realtime and also gives me the opportunity to give back to the community by helping others. Even if I’m not participating and just watching a chat, or reading chat messages from earlier in the day, I feel like I’m part of the active community, rather than just a detached end-user.” 149, reader, The Gimp

Ved at deltageren overhører chatten (eller som han kalder det: “watching a chat”), kan deltagerne opnå awareness omkring meget af det, der sker i projektet. Deltageren beskriver, at idet han er bevist omkring, hvad der sker i projektet, føler han sig som en del af et aktivt projekt.

Et positivt aspekt ved chat er, at det er let og hurtigt at få feedback samtidigt med, at det er realtime. Chat er dog ikke nær så velegnet til alle diskussioner, som eksemplet ovenover med større stykker kode. Det er også et negativt aspekt, at chat ikke er persistent. Hvis der er foregået diskussioner der har relevans for andre deltagere, skal en af de deltagere, der har deltager i diskussionen eksempelvis kopiere samtalen over på mailing listen.

Vi er af den opfattelse, at chat giver deltagere mulighed for hurtigt at trække på andre deltagers viden, altså viden som ikke er del af group memoryen, hvilket vi ser som et positivt aspekt. Det følgende citat belyser muligheden for at trække på andre deltagers viden dog kun, hvis man først selv har forsøgt:

“I go to IRC when I can’t find it myself and it doesn’t seem to be obviously answered anywhere. The Gaim developers tend to expect people to have looked around before coming to IRC.” 122, bug reporter, Gaim

Vi ser det som et meget positivt aspekt ved chat, at det er muligt at føre en samtale og trække på andre deltagers viden. Viden der ofte ikke er at finde i group memoryen.

10.2.5 Versionsstyringssystemer

Versionsstyringssystemer benyttes af alle tre projekter, der benyttes som case i vores projekt. Kildekoden til projekterne er offentlig tilgængelig via versionsstyringssystemet, samt alle tre projekter tilbyder webinterfaces til dette (eksempelvis <http://cvs.gnome.org/viewcvs/gimp/> for The Gimp projektet). Igennem webinterface er det muligt at browse, download og se information omkring de enkelte filer i projektet. Via webinterface er det muligt at sortere filerne efter deres navn, hvilken revision filen har, filens alder, forfatter navn på den deltager der sidst har uploadet en ændring, samt sortere efter den kommentar der er knyttet til den sidste ændring. Deltagerne har også muligheden for at se tidligere revisioner af filerne, kommentarerne til hver ændring, samt hvilken deltager der har ændret filen.

Versionsstyringssystemer bliver brugt meget til at se hvem, der har modificeret et artefakt (se diagram D.22), hvilket også er naturligt, da disse systemer er de eneste, af de her beskrevne systemer, der har en direkte kobling mellem, hvad der er ændret, og hvem der har ændret det. Dette er en meget specifik awareness information, og deltagerne får den ikke ved at overhøre diskussioner, som f.eks. det er muligt i chat og mailing lister, men ved at aktivt at søge efter informationen.

Et positivt aspekt ved versionsstyringssystemer er, at det giver direkte adgang den del af group memoryen, der er forbundet med det arbejde, der bliver lavet i projektet, som er defineret som awareness information i dette projekt. Vi mener desuden, at det er en vigtig funktion at have muligheden for at gå tilbage til ældre revisioner af filerne og derved få indblik i, hvorledes projektet har udviklet sig.

10.2.6 Issuetracker

En issuetracker holder basalt set styr på forskellige emner, der har relevans for projektet. Dette kan eksempelvis være forslag til forbedringer og fejl meldinger. Hvert af disse emner kan have en status, eksempelvis kan en fejl være rettet eller ny. I løbet af et emnes levetid kan denne status ændre sig, eksempelvis fra ny til rettet. Emnet kan ligeledes tilskrives en person eller en gruppe, der herefter er ansvarlig for at behandle emnet.

Gaim projektet benytter sig af Sourceforges issuetracker. MPlayer og The Gimp benytter sig af Bugzilla, og ingen af de to projekter benytter andre issuetrackers. Gaims projekt side på Sourceforge indeholder udover en bugtracker også trackere for "Support Requests", "Patches" samt flere (se [Gaim Trackers, 2005]). Disse trackere fungerer på samme måde som en bugtracker, de indeholder bare andre emner end fejl.

Issuetrackere er meget gode til at formidle awareness omkring ting der sker i et projekt. Det er således muligt for en deltager at følge emnet mens det behandles, og hele tiden være bevidst om emnets status. Vi mener derfor at det er et positivt aspekt ved issuetracker.

I det følgende analyseres systemer, der er specialiserede inden for områderne awareness formidling og group memory opbygning. Analysen af disse systemer tager ikke udgangspunkt i spørgeskemaet eller interviewet idet ingen af disse systemer bliver brugt i de Open Source projekter der deltager i undersøgelsen. Disse systemer inddrages på trods af dette, da vi mener, at de kan bidrage med ideer til designkravene til vores system. Systemerne er valgt ud fra, den research som i dette projekt danner grundlag for viden omkring group memory og awareness.

10.3 Andre systemer

De to udvalgte systemer er som nævnt Hipikat og Jazz. Betegnelse “system” er måske ikke velegnet idet både Hipikat og Jazz er plug-ins til udvikleremiljøet Eclipse¹. Vi mener dog, at man alligevel kan betegne disse som systemer, da de tilføjer meget ny funktionalitet til Eclipse, og Eclipse kun fungerer som et fundament til systemerne. Herudover er det to research systemer, som er udviklet ud fra undersøgelse af hvordan man tilbyder group memory (Hipicat) samt hvordan man levere awareness til deltagerne i et team (Jazz).

Idet ingen af de anvendte Open Source projekter benytter nogle af disse systemer, baseres denne gennemgang på information omkring systemerne, og som er tilgængelig i artikler omkring systemerne, henholdsvis Čubranić et al. [2004] og Cheng et al. [2003].

10.3.1 Hipikat

Formålet med Hipikat er at give udviklere relevante informationer fra projektets group memory, omkring det denne arbejder med i Eclipse. Dette fungerer ved, at udviklerne vælger et artefakt, eksempelvis en del af kildekoden eller lignende, og derefter vælger “Query Hipikat” fra menuen [Čubranić et al., 2004]. Hipikat returnerer derefter en liste over forskellige elementer, der har relevans for det valgte artefakt. Listen indeholder informationer om typen af elementet, eksempelvis om det er en fejl fra Bugzilla, et website eller andet. Derudover indeholder listen også navnet på elementet. Grunden til, at navnet vises i listen, er eksempelvis fordi teksten minder om det, der er søgt efter. Til sidst nævnes Hipikats vurdering af, hvor relevant elementet er.

Hipikat udvælger relevante elementer ved at analysere relationen mellem de forskellige artefakter. Dette gøres ud fra eksisterende informationer, der er tilgængelige i projektet, eksempelvis kendes altid navnet på den person, der har ændret noget i versionsstyringssystemet. Det samme gælder for forfattere til et indlæg i nyhedsgruppen [Čubranić et al., 2004, s. 3], hvor Hipikat alternativt udleder denne relation. Dette gøres eksempelvis ved at søge efter om kommentarerne i versionsstyringssystemet indeholder et id, der også findes i issue tracking systemet. For at kunne finde frem til disse relationer, arbejder Hipikat med en model, der indeholder fem elementer. Af disse artefakter findes fire ud af fem i projektets group memory.

- *Change tasks*: Eksempelvis, bug reporter, forbedrings forslag, samt andre emner der er at finde i en issueracer
- *File versions*: Eksempelvis en version af en kildekode fil fra versionsstyringssystemet

¹ Se www.eclipse.org/

- *Messages*: Dette kan være en besked på en mailing liste eller på en nyhedsgruppe.
- *Documents*: Dette kan eksempelvis være dokumenter, der er at finde på projekts website.

Det femte element er *person*, som repræsenterer forfatteren til artefaktet [Čubranić et al., 2004, s. 2-3]. Modellen indeholder forskellige forbindelser mellem disse elementer, eksempelvis hvis en udvikler har valgt en *change task* fra issuetrackeren, er denne muligvis interesseret i andre ændringer, der er blevet foretaget i den del af koden som denne change task omhandler. Hipikat vil derfor returnere elementer, som er *similar-to*, altså minder om den change task udvikleren arbejder med.

Funktionaliteten, som Hipikat tilbyder, er meget relevant for projekter der arbejder distribueret. Grunden til dette er, at Hipikat er specialiseret i de informationskanaler som projektet benytter, og anvender relationerne mellem disse. Čubranić et al. [2004] beskriver dette således:

“This is not possible if the developer is using a general search engine, such as Google, to search for documents in the same set of project archives” [Čubranić et al., 2004, s. 3]

Vi er derfor af den opfattelse, at Hipikats søgning er intelligent, idet den anvender relationerne mellem informationskanalerne, og ikke bare laver en søgning efter hvilke dokumenter, der indeholder en tekststreng. Dette ser vi kan benyttes til at forbedre informationsøgning i Open Source projekter.

10.3.2 Jazz

Jazzes formål er at udvide Eclipse udviklingsmiljøet med nogle af de aspekter, der er at finde i et åbent kontor. I et kontormiljø hvor udviklere er samlet, har disse en generel awareness omkring, hvad der laves, samt hvem der laver det. Kommunikationen i et kontor miljø er simpel og let, da udviklerne blot kan gå over til en af de andre udviklere og snakke. Dette er hvad Jazz prøver at implementere i Eclipse.

Jazz prøver at opnå dette ved at tilføje funktionalitet til Eclipse, som implementerer disse aspekter. Jazz tilbyder de følgende funktioner:

- Jazz band
- Chat
- Skærmdeling
- Concert awareness

Jazz band er et koncept, der tilføjer en bar til Eclipse, som indeholder en oversigt over alle medlemmer i et team. Teams i Jazz er små, uformelle, ad hoc og invitationsbaserede [Cheng et al., 2003, s. 2]. Hvert af medlemmerne i teamet er repræsenteret af et portræt, samtidig med, der vises forskellige informationer omkring medlemmet. Eksempelvis et status ikon der indikerer om medlemmet er *online*, *away* eller *busy*. Hvis musen holdes hen over et af portræterne vises et popup vindue med medlemmets personlige status besked, eksempelvis hvad denne arbejder med pt. [Cheng et al., 2003, s. 4].

Ved at klikke på portrættet gives en række videre muligheder, eksempelvis chat og skærm deling. Chat funktionen fungerer som et normalt chat program, der er dog visse ekstra funktioner. Eksempelvis kan de enkelte ytringer flyttes rundt og grupperes, således relaterede ytringer kan samles. Disse grupperinger er personlige og kan ikke ses af de andre personer, der deltager i chatten [Cheng et al., 2003, s. 4].

Ved hjælp af skærmdelings funktionen er det muligt for en udvikler at dele sin skærm med andre medlemmer i teamet. Dette benyttes eksempelvis til fælles fejlfinding.

Yderligere indeholder Jazz også et whiteboard, hvor flere forskellige informationer kan repræsenteres. Dette kan eksempelvis være dele af en chat, ændringer i koden m.m.

Den sidste funktion Jazz tilbyder, er det Cheng et al. [2003] betegner som *Concert awareness*. Denne funktion tilføjer awareness omkring, hvad andre team medlemmer gør ved artefakterne i projektet. Denne funktion fokuserer på artefakter i projektet, eksempelvis kildekode filerne, og indikerer hvilke filer, der pt. arbejdes med, hvilke filer der er modificerede osv.

Der er flere aspekter ved Jazz der er interessante. Vi mener, at presence indikation (hvilken status medlemmet har pt) kombineret med information om, hvad medlemmet arbejder på, er en værdifuld information i distribueret udvikling. Grunden er at det kan give person awareness. Ligeledes er ideen omkring et whiteboard, hvor mange forskellige informationer kan placeres, er et velegnet værktøj til opbygning af group memory. Konceptet minder om en Wiki (se afsnit 2.2), hvor alle medlemmer i en gruppe kan tilføje, ændre og slette indhold. Vi er af den opfattelse, at Jazz tilbyder mange forskellige måder at formidle awareness på, og at disse kan benyttes i sammenhæng med et andet system og derved fremme awareness blandt dem, som anvender systemet.

11 Designkrav

I dette kapitel opstilles designkrav til systemet. Udgangspunktet er arketypernes karakteristika og analysen af eksisterende systemer.

11.1 Designkrav

Ud fra arketypernes karakteristika identificeres ni designkrav. I de følgende analyseres, hvorledes disse krav fremkommer, både ved at inddrage karakteristika ved arketyperne samt ved at supplere med aspekter fra analysen af eksisterende systemer.

11.1.1 Systemet skal muliggøre informationsøgning i alle officielle informationskanaler

Et designkrav til vores system er, at alle de eksisterende informationskanaler, der er behandlet i projektet, skal kunne tilgås fra systemet. Det vil sige, at det skal være muligt at søge i alle informationskanalerne. Grunden til dette designkrav er, at arketyperne bruger alle informationskanalerne.

En anden grund til at medtage alle informationskanalerne er, at vi har fastslået, at systemet skal være brugbart for alle deltagerne i projekterne.

Ved at samle informationssøgningen til ét sted minimeres antallet af deltagere, der søger flere steder for at finde den information de søger. Vi ser et potentiale for at forbedre alle tre arketyper søgning idet alle tre arketyper karakteriseres ved at søge flere steder for at finde den information de søger.

En yderligere grund til dette designkrav er, at alle typerne benytter tredje parts informationskanaler. Ved at medtage alle kanalerne, samler vi muligheden for at søge information og derved samles også tilgangen til hele group memoryen. Derved

øges chancen for at finde information i de officielle kanaler og dermed minimeres deltagerens behov for brug af tredje parts kanaler.

Udover at systemet kan minimere brugen af tredje parts kanaler, kan vi også gøre det nemmere for arketype 1 at finde information. Et karakteristika ved arketype 1 er, at de har svært ved at finde information fordi de mangler informationskanaler. Dette løses ved at medtage alle kanalerne.

Der, hvor arketype 2 først søger efter information, er i andre kanaler end de officielle. Dette kan tyde på, at de officielle informationskanaler ikke opfylder deres behov for at finde information. Hvis systemet skal samle arketypernes informationssøgning til ét sted, kræver det derfor en forbedring af søgningen i informationskanalerne, således de tilfredsstiller arketype 2's behov. De to andre grupper søger også i andre informationskanaler og deres informationssøgning kan derved også blive samlet mere hvis dette krav implementeres. Som beskrevet i analysen af Sourceforge tilbyder dette sted også søgeresultater fra tredje parts kanaler. Men idet søgeresultaterne ikke altid er brugbare og relevante, mener vi ikke dette skal være en mulighed i vores system.

Idet der kan eksistere informationer, omkring det samme emne, på forskellige informationskanaler vil det gøre det lettere for deltagerne, hvis systemet kan vise disse relaterede informationer sammen. Et eksempel på samme information i forskellige informationskanaler er et stykke kode. Det vil indgå som kode i versionsstyringssystemet, diskussion i mailing listen samt som et emne i issue trackerne der omhandler denne kode. Her kunne modellen med intelligent søgning, som Hipikat benytter inddrages i systemet for at tilbyde at søge på relaterede emner på forskellige informationskanaler.

11.1.2 Systemet skal være webbaseret

Arketype 1 anvender bl.a. fora for at undgå at anvende eksternt program, i dette tilfælde anses en webbrowser ikke for et eksternt program. Dette ser vi, kan overføres som et designkrav til vores system. Ved at vores system er webbaseret, gøres det nemt at tilgå og anvende, idet der typisk er en webbrowser på alle computere der er tilsluttet Internettet, og at mange af de officielle informationskanaler tilgås via en webbrowser, eksempelvis fora, issue trackers og mailing liste arkiver.

Dette er også et positivt aspekt vi har fundet ved websites, at det er et godt udgangspunkt til informationsøgning, hvor man kan

11.1.3 Krav om søgefunktion i mailing lister

Karakteristisk for alle tre arketyperne er, at de alle primært og hyppigst anvender mailing lister. Vi opstiller derfor et designkrav der omhandler, at det skal være muligt i vores system at søge i denne informationskanal, især i arkiverne. Fra analysen af mailing lister fremgår også, at denne kanal indeholder store dele af projekters

group memory. Dette understøtter kravet om, at systemet skal tilbyde søgning i mailing lister, idet der således sikres adgang til en stor mængde group memory. Endvidere skal det være muligt for alle deltagere at tilgå alle mailing lister. Grunden er, som beskrevet i analysen af mailing lister, at forskellige deltagere derved kan hjælpe hinanden på tværs af deres roller.

11.1.4 Der skal være mulighed for chat

Idet alle tre arketyper anvender chat, er det et krav til systemet at det tilbyder adgang til chat. For at undgå at deltagerne skal anvende eksterne programmer skal vores system kunne vise de officielle informationskanaler, i dette tilfælde chat. Det arketyperne foretrækker ved chat er, at det er realtime, samt at det føles som en rigtig samtale. Arketype 1 nævner også, at ved at overvære en chat kan der opnås en følelse af at være en del af projektet, samt at der er mulighed for nemt og hurtigt at kunne hjælpe andre. Chat er nødvendigvis realtime for at en deltager kan give hurtig respons eller føle sig en del af projektet. Ved at gøre relevant snak fra chatten persistent, kunne meget tilføjes til group memoryen. Som der nævnes i eksisterende systemer kan man opnå adgang til viden som ikke er i group memoryen, men som en deltager besidder. En mulighed for at sikre at denne viden bliver en del af group memoryen, er at logge chatten. Dette giver problemer med hvorledes relevante snak fra chatten udvælges. I Jazz løses dette ved at deltagerne i chatten kan placere relevant emner på et whiteboard, hvorefter det indgår i group memoryen. Denne løsning kunne eksempelvis også anvendes i vores system. Det er dog stadig ikke fuldstændig uproblematisk, idet det kræver at deltagerne udfører en handling for at gøre samtalen del af group memoryen.

11.1.5 Systemet skal kunne præsentere data fra versionsstyringssystemer

Dette krav skyldes, at alle arketyperne anvender versionsstyringssystemerne, når de skal se hvem, der har ændret et artefakt. Kravet til systemet er derfor, at det skal kunne præsentere information omkring, hvem der har rettet i de forskellige artefakter, der ligger i versionsstyringssystemer. Endvidere anvender arketype 3 versionsstyringssystemer til at hente koden ned og søge i den. Systemet skal derfor også kunne hente koden og lade deltageren søge lokalt i den. Endvidere beskrives det i analysen af versionsstyringssystemer, at det er positivt at kunne sortere data efter forskellige attributter og yderligere at kunne gå tilbage i versioner. Hvis systemet tilbyder dette vil det understøtte arketype 3's behov for at kunne søge i koden, men også give de tre arketyper mulighed for at sortere data efter hvem der har ændret i artefakterne. Herved kan de nemt finde den information de søger.

11.1.6 Systemet skal være brugervenligt og velstruktureret

I analysen af websites beskrives hvorledes et negativt aspekt ved disse er, at de kan være ustrukturerede og ikke særlig brugervenlige. Det er derfor et krav til vores system, at det ikke er således. Et yderligere aspekt ved dette er, at de tre arketyper siger det er blevet lettere at finde information fordi de har fået mere erfaring med de forskellige informationskanaler. Vi mener at denne proces, med at lære informationskanalerne at kende, kan effektiviseres ved at gøre systemet brugervenligt og struktureret. Dette betyder at nye deltagere lettere kan benytte vores system, og erfarne deltagere kan benytte systemet effektivt.

Blot at samle de forskellige informationskanaler på et website, gør det hverken lettere eller mere effektivt at finde informationer. Dette kan eksempelvis ses ud fra, at en deltager i arketype 1 nævner tre områder, hvor informationssøgningen i Sourceforge ikke opfylder deltagerens behov. Der er dermed tre områder, hvor vores system kan forbedre de eksisterende muligheder for at søge. Vores system skal, sammenlignet med Sourceforge, tilbyde bedre navigering, være bedre til at give overblik samt have en hurtigere søgning.

11.1.7 Systemet skal tydeliggøre awareness information

Alle tre arketyper har behov for awareness, dog forskellige former for awareness. For at tydeliggøre awareness information skal systemet derfor kunne præsentere de forskellige former for awareness.

Arketype 1 søger meget generel awareness. Da denne form for information også søges af de to andre typer, skal systemet præsentere denne type information tydeligt. Dette kan gøres ved, at information om nye releases osv fremgår som noget af det første man ser i systemet. Idet arketype 3 har stort behov for person awareness skal systemet yderligere kunne præsentere informationer som f.eks. hvem der laver hvad.

Ingen af arketyperne anvender issuetrackers i vid udstrækning, men i analysen beskrives, at issuetrackers er gode til at formidle awareness om hvad der sker i projektet. Alle deltagerne er interesseret i awareness, og således kan det fremme deres awareness, hvis de kan tilmelde sig til at modtage e-mail hver gang der sker noget nyt i issuetrackers og andre informationskanaler. Dette beskrives i analysen af fora. For at det ikke påtvinges deltagerne ser vi, at det kan gøres ved at deltagerne selv tilmelder sig hvilke informationskanaler de ønsker at få besked om. Systemet skal således sende e-mails til deltagerne hver gang der er nyt på informationskanalerne deltageren har tilmeldt sig.

11.1.8 Systemet skal give mulighed for feedback

Endnu et designkrav til systemet er, at det giver deltagerne mulighed for at følge op på et emne i en informationskanal. Dette er en karakteristika ved alle arketyperne. For at deltagerne kan gøre det igennem systemet, kræver det, at systemet har to vejs kommunikation med informationskanalerne således, de både kan foretage søgning i dem men også sende deltagerens svar tilbage.

11.1.9 Gøre det let for arketyperne at tilgå den information de søger

Da arketype 1 søger efter manualer for at lære, hvordan softwaren virker eller for at lære nyt, sætter det krav til vores system om, at manualer skal være nemme at finde. Arketype 3 derimod søger primært teknisk dokumentation, hvilket sætter et krav, om at dette er let tilgængeligt. For at disse to krav kan opfyldes er en løsning, at deltagerne, når de foretager en søgning, kan afkrydse hvilke kanaler de ønsker søgningen udført i. Derved undgås at arketype 3 får svar tilbage om manualer, og modsat for arketype 1, at de ikke får teknisk dokumentation. Dette kan eksempelvis være hvis en deltager søger efter navnet på en funktion i koden. Søgningen vil kunne returnere resultater fra manualer men også fra selve koden hvor funktionen er implementeret. Ved at krydse af kan deltagerne begrænse søgning og derved kun få resultater, som er relevante for dem. Dette vil også forbedre arketype 2's informationssøgning, da de synes det er svært at finde teknisk dokumentation. Ved direkte at kunne begrænse søgningen til kun at omhandle teknisk dokumentation antager vi at de vil få brugbare resultater.

Konklusion

I dette kapitel konkluderes på arketypernes karakteristika, hypoteserne samt designkravene og hvorledes disse effektiviserer informationssøgningen. Yderligere konkluderes om resultaterne kan benyttes i andre Open Source projekter.

12.1 Konklusion

Vores problemformulering er som følgende:

“Hvilke designkrav er der til et system, som skal effektivisere informationsøgning og derved samarbejdet blandt deltagere i Open Source projekter?”.

For at besvare denne har vi identificeret en række karakteristika ved deltagere i Open Source projekter. Ud fra disse karakteristika er designkrav til systemet analyseret frem. Endvidere er eksisterende systemer analyseret for at finde positive og negative aspekter, der anvendes i udadbejdelsen designkrav.

Arketyper

Vi kan konkludere, at de forskelle, der er i arketypernes karakteristika ikke udmundede i direkte modstridende krav, men i forskellige krav. Derfor har det været muligt at inkludere alle tre typers karakteristika i designkravene.

Argumentet for at bruge arketyper er en antagelse om, at deltagerne i Open Source projekter har forskellige roller og derved også forskellige karakteristika. Der, hvor vi identificerer, at arketyperne primært har forskellige karakteristika, er i det følgende:

- Forskellig behov for awareness information
- Hvilken type information arketyperne søger

- Hvilke informationskanaler arketyperne benytter
- Hvor længe arketyperne har været involveret i projekterne

Hypoteser

At arketyperne har forskellige karakteristika understøttes også af de opstillede hypoteser. Vi kan konkludere, at alle tre hypoteser er blevet valideret igennem resultaterne opnået igennem empiriindsamlingerne i projektet.

De arketyper, vi har identificeret, karakteriseres blandt andet ud fra, hvor lang tid de har været involveret i projektet. Ud fra de resultater, vi har opnået igennem empiriindsamlingen, kan vi konkludere, at deltageres involvering i Open Source projekterne har indflydelse på, hvilke informationskanaler de benytter og hvilke informationer, de søger efter. Da det derfor er forskellige informationskanaler de benytter og forskellige informationer de søger efter, kan det yderligere konkluderes, at deltagerne har forskellige problemer med informationsøgning. Hvilket yderligere afspejles igennem, at den erfaring, de har i projektet, også har betydning for, hvor svært de synes, det er at finde information. Endvidere kan det konkluderes, at deltagerne har forskellige måder at søge information alt efter deres involvering i projekterne. Ud fra dette er vores hypoteser valideret.

Designkrav

Det kan ikke konkluderes, om det er muligt at implementere de opstillede designkrav til systemet. Besvarelsen på problemstillingen er følgende liste af designkrav:

- Systemet skal muliggøre informationsøgning i alle officielle informationskanaler
- Systemet skal være webbaseret
- Krav om søgefunktion i mailing lister
- Der skal være mulighed for chat
- Systemet skal kunne præsentere data fra versionsstyringssystemer
- Systemet skal være brugervenligt og velstruktureret
- Systemet skal tydeliggøre awareness information
- Systemet skal give mulighed for feedback
- Gøre det let for arketyperne at tilgå den information de søger

Ved at overholde de opstillede designkrav til systemet opnås et system til hele målgruppen og som tilfredsstillende de forskellige deltagers karakteristika i forbindelse med informationssøgning.

Effektivisering

En del af problemformuleringen er: “...som skal effektivisere informationssøgning...”. Det er ikke muligt at sige med sikkerhed om designkravene vil effektivisere processen. For at kunne konkludere på dette er det nødvendigt at implementere systemet, og teste om processen bliver mere effektiv. Vi antager dog, at hvis systemet bliver udviklet ud fra de opstillede designkrav, hvor informationssøgning bliver samlet i et system, at det vil effektivisere søgningsprocessen for de af deltagerne, der skal søge mere end et sted.

Generalisering

Det er ikke muligt at konkludere om vores resultater er repræsentative for Open Source projekter generelt. Dette hænger sammen med, at vi ikke undersøger, om der forskelle mellem deltagerne i de forskellige projekter. Desuden har vi også valgt Open Source projekter, der minder om hinanden og derfor muligvis også har en ens fordeling af deltagere og karakteristika ved disse. Vi antager dog, at de arketyper vi har identificeret også findes i andre projekter. Samtidig er vi også af den overbevisning, at vores hypoteser kan valideres for andre Open Source projekter, hvis de består af flere forskellige typer deltagere.

13 Refleksion

Dette kapitel reflekterer over opdelingen af typer, og det at anvende arketyper som en metode, til beskrivelse af deltager typerne. Yderligere reflekteres der over, hvordan de eksisterende systemer bruges i designkravene og over softwaren, som er brugt til spørgeskemaundersøgelsen.

13.1 Arketyper

Metoden, som er brugt til arketyperne, har været fordelagtig, fordi den ligger op til en iterativ proces. Denne proces gør det muligt at forbedre og præcisere arketyperne med flere iterationer. På den måde kan der ligeledes inddrages flere aspekter i analysen af arketyper, da det er muligt at tage forskellige tilgange/metode ind i de forskellige iterationer.

Det kan diskuteres, hvornår processen har været gennem nok iterationer, og hvornår arketyperne kan klassificeres som værende valide og endelige. Vi mener at de tre iterationer vi har været igennem, har givet en god forståelse af deltagerne, i et Open Source projekt, og dermed målgruppen for systemet.

I interviewet kunne der være spurgt ind til deres type, baseret på deres svar i spørgeskemaet, og hvilke opgaver de varetager i projektet. På den måde kunne der være analyseret på de opgaver, som personen varetager i projektet, og måske revidere typerne udfra dette. Dette ville have givet en mulighed for at tjekke om artiklen om de initiale typer, afspejler det reelle billede af hvilke roller der er i et Open Source projekt og hvilke roller typerne besider.

For at opveje den manglende mulighed for at spørge ind til respondenterne af interviewet, som ved et mundtligt interview, kunne denne iterationen være gentaget flere gange. For på den måde, at kunne få uddybet svarene fra respondenterne yderligere. Dette kunne have ført til en bedre forståelse af, hvorfor de har svaret, som

de har gjort. Det kunne også have den modsatte effekt, at respondenterne ikke ville have svaret tilbage, fordi de har oplevelsen af at vi forlanger for meget af dem.

13.1.1 Opdelingen af typer

I spørgeskemaet er der opstillet 8 typer, som deltagerne skal placere sig under. Nogle af typerne, blandt andet peripheral developers og bug fixers, kan være svære at differentiere i forhold til de opgaver de varetager i et Open Source projekt. Derfor har typerne måske haft for mange inddelinger. I forhold til vores projekt og analyse, har det ikke haft betydning for, hvordan arketyperne er opdelt, da både peripheral developer og bug fixer, høre under arketype 2.

Metoden til opdeling af typerne i undersøgelsen er baseret på, hvor meget deltageren bidrager i form af kode. Det er defineret, at deltageren har indflydelse afhængigt af, hvor meget kode personen bidrager med. En kritik af denne opdeling er, at en deltager også kan have indflydelse, uden nødvendigvis at bidrage med kode. I følge opdelingen i artiklen, som metoden bygger på, er en deltager, der ikke bidrager med kode, en bug reporter eller reader. Dette kan have betydning for hvad respondenterne til spørgeskemaet har svaret. Det er dog svært at vurdere, hvilken betydning der har haft for undersøgelsen. Hvis en person har tænkt, at hvis han skriver dokumentation og dermed ikke koden direkte, er han "kun" reader, og derfor kan konsentrationen af readers være højere end normalt. Da ingen af respondenterne har tilkendegivet at typerne ikke matchede deres beskrivelse og rolle i et Open Source projekt, antager vi ikke at dette har været et problem, der har influeret på resultaterne.

13.2 Den rette løsning

I projektet udarbejdes et designkrav til et CSCW-system, som skal forbedre samarbejdet i Open Source projekter. Det er dog værd at reflektere over, om et system er den rette løsning til at forbedre samarbejdet. Problemstillingen ligger måske i stedet i at informere og lære nye brugere omkring etikken i et Open Source community om, hvordan man søger information. Dette kan være med til at afhjælpe, at der ikke kommer så mange af de samme spørgsmål. Hvis deltagerne i projekterne bliver bedre til at søge, opretter de derfor ikke en dublet af et eksisterende spørgsmål.

En anden løsning kunne være at optimere mulighederne for at søge information på mailing lister. Dette er den informationskanal flertallet af respondenterne i spørgeskemaet anvender, så ved at forbedre søgningen her, ville flere deltagere måske få lettere ved at finde information. Problemet ved denne løsning er dog, at mailing lister ikke bliver brugt til al slags informationsudveksling. F.eks. findes group memory i form af koden, i stedet i versionsstyringsystemer. Det vil derfor ikke være en løsning på alle områder af informationsøgning.

Vores problemstilling, er med udgangspunkt i informationssøgningen, når informationen er genereret og skal findes frem igen. Men måske ligger problemet et andet sted, nemlig i de værktøjer som er med til at generere informationen. Hvis værktøjerne blev udviklet således, at inden man tilføjede et nyt spørgsmål, ville programmet automatisk søge efter problemer med relation, til det spørgsmål som brugeren netop har indtastet.

13.3 Brugen af eksisterende systemer

I projektet er fremgangsmåden først at finde deltagerne krav til systemet og dernæst at analysere eksisterende systemer. Alternativt kunne fremgangsmåden være at inddrage analyse af eksisterende systemer til start. Herved kunne de positive og negative aspekter ved systemerne påvirke resten af processen. Analysen af de eksisterende systemer kunne udmunde i krav til vores system som undersøgelsen af projekterne kunne bygge på. Således ville både spørgeskemaet og interviewet tage udgangspunkt i disse krav. På den måde ville analysen indskrænke den efterfølgende undersøgelse. Måden hvorpå analysen anvendes i projektet bidrager til at løse nogle af arketypernes krav. Det er således svarene fra deltagerne der afgør kravene til systemet, og ikke den viden, som kan hentes om de eksisterende systemer.

13.4 Spørgeskema software

Softwaren brugt til spørgeskemaet havde en række begrænsninger der gjorde at nogle af spørgsmålene, kun kunne opstilles på en bestemt måde. Eksempelvis spørgsmål 17 og 18, hvor formålet med 18 er at belyse relationen mellem, hvilken information deltagerne leder efter, og hvor de leder efter denne. Softwaren har i sin nuværende udgave, ikke mulighed for at benytte svar fra tidligere spørgsmål i senere spørgsmål, således at f.eks. svarene fra spørgsmål 17 kunne bruges i spørgsmål 18, og deltagerne så kunne markere de forskellige informationskanaler hvor de søger efter informationen.

13.5 Empiri

At interviewet er lavet som et skriftligt interview, kan også have betydning for de svar respondenterne har givet. Hvis interviewet var lavet mundtligt, havde det givet mulighed for at stille uddybende spørgsmål, hvis der var tegn på, at spørgsmålet var blevet misforstået, eller hvis respondenterne berørte en interessant vinkel på problemstillingen.

Det skriftlige interview har ikke givet så mange svar som ønsket, trods deltagernes velvillighed til at oplyse deres email adresse, i spørgeskemaet, til samme formål. En grund til de få svar, kan ligge i at respondenterne ikke har ville deltage alligevel, måske synes de der var for mange spørgsmål, for store spørgsmål, eller andre aspekter i den stil. Dette forsøgte vi at gardere os imod, ved at skrive i e-mailen, som blev udsendt sammen med interviewet, at ønskede de ikke at deltage alligevel, ville vi gerne vide besked, det er der kun 2 personer der har gjort.

Dette har betydet at vi er blevet nødt til at inddrage svar fra deltagere som ikke er repræsentative ifølge de parameter vi har opstillet.

14 Videreudvikling af systemet

I dette kapitel perspektiveres over ideer til, hvordan det system, vi har udarbejdet designkrav til, kan videreudvikles. I det følgende diskuteres problemstillinger og overvejelser omkring denne videreudvikling.

14.1 Gør det Open Source

En måde at videreudvikle systemet på, er ved at gøre det til et Open Source projekt. Dette kan selvfølgelig ses som en logisk mulighed, når det er lavet til Open Source projekter. Men yderligere findes en række fordele ved at udvikle systemet som et open source projekt og herved distribueret.

En fordel er, at der er mange potentielle udviklere, der kan tilknytte sig projektet, netop fordi det er Open Source. Herved er kildekoden fri og derved kan deltagerne benytte denne frit. Den store mængde af potentielle udviklere gør, at der er mulighed for at der opstår et solidt fundament for, at systemet videreudvikles. Da der netop er denne store mængde af potentielle udviklere, er der også chancer for at erfarne og dygtige udviklere bidrager med deres viden.

Ved at udvikle systemet som et Open Source projekt, er også muligheden for at udvikle distribueret. Dette giver mulighed for at parallel udvikle systemet, som kan være en faktor der er med til at afgøre projektets mulighed for at overleve. En antagelse er, at hvis folk ikke har mulighed for a bidrage til projektet når de har tid og lyst til det, betyder det, at de vælger helt at opgive projektet.

For at lokke disse potentielle udviklere til at bidrage til projektet, skal der gøres en indsats for at motivere dem. Dette kan f.eks. være igennem selve ideen ved

projektet; at de selv får glæde af den funktionalitet og mulighed systemet giver dem.

En anden mulighed for at motivere udviklerne, er at tilknytte en sponsor til projektet. Ved dette kan tiltrækkes erfarne og dygtige udviklere, der ud over at kunne få glæde af systemet selv også kan opnå en form for betaling for deres arbejde.

Endvidere er en fordel ved at udvikle det som et Open Source projekt, at de udviklere, der bidrager til systemet, selv er potentielle brugere. Derved har udviklerne muligheder for at tilpasse systemet til de karakteristika der er vigtige for dem. Samtidigt kan en udvikler også lave specielle funktioner og muligheder for sig selv ved at påvirke systemet personligt.

Da der er forskel på hvilke informationskanaler Open Source projekter anvender, skal det være muligt at tilpasse systemet, så det er muligt at specificere, hvilke informationskanaler systemet skal søge i.

Ved at opbygge systemet til at kunne søge i specifikke informationskanaler, giver det også mulighed for, at Open Source projekter, der ønsker at anvende systemet, selv kan udvikle systemet til at kunne søge i netop de informationskanaler, de ønsker.

14.1.1 Programmering

Når der skal vælges programmeringssprog til implementeringen af systemet, er der nogle overvejelser, der skal gøres. Eftersom systemet er tiltænkt, at skulle bruges i Open Source projekter, skal programmeringssproget kunne integreres og implementeres op imod de eksisterende værktøjer, der anvendes, som eksempelvis en mailing list eller en nyhedsgruppe.

En overvejelse er også, at bruge et programmeringssprog og guidelines, som allerede anvendes i andre Open Source projekter, til opbygning af kodenstrukturen. På den måde vil det formodentlig være nemmere at få frivillige til at hjælpe med udviklingen af systemet.

14.2 Test af systemet

For at videreudvikle systemet er det vigtigt at teste systemet. Dette gælder både test af de designkrav, der er udarbejdet, men også forskellige test, der skal laves efter systemet er implementeret. Her er det vigtigt at gå ind og vurdere hvordan disse tests udføres. Som beskrevet er det tiltænkt, at systemet skal udvikles som et Open Source projekt. I Open Source projekter er deltagerne typisk spredt ud over hele verden. Derfor er det vigtig at lave overvejelser omkring, hvordan disse brugere inddrages i forskellige test.

14.2.1 Brugervenlighedstest

Et af de krav vi har opstillet for systemet er, at det skal være brugervenligt og velstruktureret. En brugervenlighedstest er med til at sikre, at systemet er anvendeligt, og at det har et godt visuelt design. En sådan test kan have flere formater, forstået således, at det kan være en test af muck-ups eller et beta-release af systemet som testes.

Testen kan være med til at sikre, at informationen, som præsenteres for brugerne efter en søgning, er præsenteret på en fornuftig form. Et yderligere aspekt der skal testes i denne sammenhæng er, hvordan deltagerne foretrækker forskellige group memory og awareness informationer repræsenteres, så det bliver mest effektivt at bruge systemet for dem. Dette kan lede til nye designkrav til systemet eller ændring af de allerede opstillede designkrav.

En mulighed til at udføre brugervenlighedstests er ved at inddrage personer fra de tre arketyper, hvilket giver et repræsentativt udsnit af målgruppen. På den måde kan det være med til at sikre, at systemet dækker alle typer af brugers behov og ønsker.

Et aspekt er også at teste systemet i situationer, hvor systemet skal anvendes. Dette kan gøres ved at lade et projekt anvende systemet og ved hjælp af en bug- og issue tracker til systemet kan deltagerne i projektet kan rapportere tilbage. Dette vil betyde at systemet ud fra feedback og ønsker, hele tiden har mulighed for at udvikle sig og blive bedre og mere effektiv.

14.2.2 Supplerende undersøgelser

For at dette system skal blive en succes, kan der være brug for yderligere undersøgelser. Disse skal ses som supplement til analysen i dette projekt.

Brugsmønstre

Vores projekt analyserer, hvilken information deltagerne søger efter, og i hvilke kanaler de søger, samt hvordan de søger deres information. Det kan yderligere være en mulighed, at undersøge andre af deltagernes brugsmønstre. Dette kan f.eks. være hvorledes deltagerne kommunikerer og information skabes. Ved at finde designkrav til disse brugsmønstre kan systemet udbygges til at forbedre informationssøgningen allerede fra når informationen dannes.

14.2.3 Søgealgoritme og klassificering af information

I designkravene beskrives, at systemet tager Hipikats ideer med at gøre søgningen intelligent. Denne søgning kan forbedres og gøres endnu mere intelligent ved at

finde en søgealgoritme, der kan karakterisere det deltagerne søger efter og returnere resultater, der passer til dette. Endvidere skal det derfor undersøges, hvorledes systemet kan klassificere information, således at de rigtige typer information returnerer på søgningerne.

Adoption af systemet

Det er endvidere nødvendigt at undersøge, hvorledes det sikres at systemet faktisk bliver brugt. Selv hvis systemet lever op til designtkravene kan det stadig ske at systemet ikke bliver brugt. Der kan være flere grunde til dette, eksempelvis nævner Čubranić et al. [2004] i forbindelse med deres design af "Hipikat":

"We believe that this low barrier to adoption is crucial for Hipikat to be accepted in the real world, since it has been repeatedly found that CSCW systems that require significant changes to work practice or that require users to constantly externalize and map their knowledge are doomed to failure." Čubranić et al. [2004, s. 2]

Vi mener at kunne undgå dette ved at vores system fungerer som et supplement til de eksisterende informationskanaler, det skal bare gøre det lettere at finde informationerne i disse. Dette betyder, at deltagerne ikke skal ændre måden de kommunikerer på. Dette giver således en ny problemstilling; hvis deltagerne blot kan forsætte med at benytte de eksisterende informationskanaler, hvorfor skal de så benytte vores system. Deltagerne påpeger, at de mener at det er forholdsvis let at finde information (3,6 på en skala fra 1 til 5, se afsnit 6.3). Dog påpeger de også, at de søger flere steder (55%) for at finde information og det er netop der, hvor vores system kan gøre en forskel. Hvis de 55% af deltagerne er klar over, at de kan benytte et system istedet for at søge flere steder, antager vi at disse må opleve det som endnu lettere at søge information.

Vores system skal således gøre en forskel for deltagerne vil bruge det. I en virksomhed er det muligt at påtvinge medarbejderne at de skal bruge et bestemt system. Et Open Source projekt kan også vælge at benytte et system de kan dog ikke tvinge nogen til noget. Deltagerne skal derfor have lyst til at bruge vores system.

Litteratur

Lucy M. Berlin, Robin Jeffries, Vicki L. O'Day, Andreas Paepcke, og Cathleen Wharton. Where did you put it? issues in the design and use of a group memory. In *CHI '93: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 23–30, New York, NY, USA, 1993. ACM Press. ISBN 0-89791-575-5.

Andrea Capiluppi, Patricia Lago, og Maurizio Morisio. Characteristics of open source projects. In *Proceeding of the Seventh European Conference On Software Maintenance And Reengineering (CSMR'03)*. IEEE, 2003.

Li-Te Cheng, Susanne Hupfer, Steven Ross, og John Patterson. Jazzing up eclipse with collaborative tools. In *eclipse '03: Proceedings of the 2003 OOPSLA workshop on eclipse technology eXchange*, pages 45–49, New York, NY, USA, 2003. ACM Press.

Allan Cooper. *The Inmates Are Running the Asylum*. Sams, Indianapolis, 1999. ISBN 0-672-31649-8.

Davor Čubranić, Gail C. Murphy, Janice Singer, og Kellogg S. Booth. Learning from project history: a case study for software development. In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 82–91. ACM Press, 2004. ISBN 1-58113-810-5.

Eric Steven Raymond. How to ask questions the smart way. <http://www.catb.org/~esr/faqs/smart-questions.html#before>, 2004.

Joseph Feller og Brian Fitzgerald. *Understanding Open Source Software Development*. Pearson Education Limited, 2002. ISBN 0-201-73496-6.

Freshmeat.net. freshmeat.net. freshmeat.net, 2005. Besøgt 19 Apr. 2005.

Gaim mailing liste arkiv. Gaim: Mailing lists. http://sourceforge.net/mail/?group_id=235, 2005. Besøgt 4. juni 2005.

Gaim Open Discussion. Gaim: Discussion forums: Open discussion. http://sourceforge.net/forum/forum.php?forum_id=665, 2005. Besøgt 5. juni 2005.

- Gaim Trackers. Gaim: Trackers. http://sourceforge.net/tracker/?group_id=235, 2005. Besøgt 5. juni 2005.
- Kim Goodwin. Getting from research to personas: Harnessing the power of data. 2002.
- Google. Groups help: 20 year usenet timeline. <http://groups-beta.google.com/support/bin/static.py?page=timeline.html>, 2005. Besøgt 19 Apr. 2005.
- Saul Greenberg, editor. *Computer-supported cooperative work and groupware*. Academic Press Ltd., London, UK, UK, 1991. ISBN 0-12-299220-2.
- Carl Gutwin, Reagan Penner, og Kevin Schneider. Group awareness in distributed software development. In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 72–81, New York, NY, USA, 2004. ACM Press. ISBN 1-58113-810-5.
- JoAnn T. Hackos og Janice C. Redish. *User and Task Analysis for Interface Design*. Wiley computer publisher, 1998. ISBN 0-471-17831-4.
- James D. Herbsleb og Rebecca E. Grinter. Architectures, coordination, and distance: Conway's law and beyond. *IEEE Softw.*, 16(5):63–70, 1999. ISSN 0740-7459.
- Michael Kantor, Beatrix Zimmermann, og David Redmiles. From group memory to project awareness through use of the knowledge depot. *The 1997 California Software Symposium*, pages 19–26, 1997.
- Jesper Kjeldskov og Tom Nyvang. Lessons from being there: Design af et rigere rum for interaktion. Master's thesis, Aalborg University, Denmark, 2000.
- Sari Kujala og Marjo Kauppinen. Identifying and selecting users for user-centered design. In *NordiCHI '04: Proceedings of the third Nordic conference on Human-computer interaction*, pages 297–303, New York, NY, USA, 2004. ACM Press. ISBN 1-58113-857-1.
- Hanny Yulius Limanto, Nguyen Ngoc Giang, Vo Tan Trung, Jun Zhang, Qi He, og Nguyen Quang Huy. An information extraction engine for web discussion forums. In *WWW '05: Special interest tracks and posters of the 14th international conference on World Wide Web*, pages 978–979, New York, NY, USA, 2005. ACM Press. ISBN 1-59593-051-5.
- Audris Mockus, Roy T Fielding, og James D Herbsleb. Two case studies of open source software development: Apache and mozilla. *ACM Trans. Softw. Eng. Methodol.*, 11(3):309–346, 2002. ISSN 1049-331X.

- Audris Mockus og James D. Herbsleb. Open source, design, and collaboration. In *UIUC/NSF Workshop on the Continuous (Re)Design of Open Source Software*, University of Illinois at Urbana-Champaign, 2003. available at <http://www.isr.uci.edu/events/ContinuousDesign/UIUC/data/Herbsleb-Mocku%3.pdf>.
- MPlayer mailing liste arkiv. Mailing lists. http://www.mplayerhq.hu/homepage/design7/info.html#mailing_lists, 2005. Besøgt 4. juni 2005.
- Kumiyo Nakakoji, Yasuhiro Yamamoto, Yoshiyuki Nishinaka, Kouichi Kishida, og Yunwen Ye. Evolution patterns of open-source software systems and communities. In *IWPSE '02: Proceedings of the International Workshop on Principles of Software Evolution*, pages 76–85, New York, NY, USA, 2002. ACM Press. ISBN 1-58113-545-9.
- Jan Noyes og Chris Baber. *User-centred design of systems*. Springer, 1999. ISBN 3-540-76007-5.
- Gary M. Olson og Judith S. Olson. Distence matters. *Human-Computer Interaction, volume 15*, pages 139–178, 2000.
- Open Source Initiative. Open source initiative osi - licensing. <http://www.opensource.org/licenses/>, 2005. Besøgt 16 Maj. 2005.
- Kjeld Schmidt. The problem with 'awareness'. *Computer Supported Cooperative Work 11*, pages 285–298, 2002.
- SourceForge.net. Sourceforge.net: Project info - gaim. <http://sourceforge.net/projects/gaim/>, 2005. Besøgt 20. Apr. 2005.
- Sourceforge.net. Sourceforge.net: Welcome. sourceforge.net, 2005. Besøgt 19 Apr. 2005.
- Telekooperation. Cscw -computer supported cooperative work, 1997. URL <http://www.telekooperation.de/cscw/>.
- Sotirios Terzis og Paddy Nixon. Building the next generation groupware: A survey of groupware and its impact on the virtual enterprise. URL citeseer.ist.psu.edu/terzis99building.html.
- Sotiris Terzis. Cscw & groupware, 1998. URL <http://www.cs.tcd.ie/Sotirios.Terzis/CSCW.html>.
- The Gimp mailing liste arkiv. Mail list links. http://www.gimp.org/mail_lists.html, 2005. Besøgt 4. juni 2005.
- The xine-Project. xine - a free video player. www.xinehq.de, 2005. Besøgt 28 Apr. 2005.

Georg von Krogh, Sebastian Spaeth, og Karim R. Lakhani. Community, joining, and specialization in open source software innovation: a case study. *Research Policy*, 32(7):1217–1241, 2003. available at <http://ideas.repec.org/a/eee/respol/v32y2003i7p1217-1241.html>.

WikiPedia. Weblog - wikipedia, the free encyclopedia. <http://en.wikipedia.org/wiki/Weblog>, 2005. Besøgt 28. Apr. 2005.

Yunwen Ye og Kouichi Kishida. Toward an understanding of the motivation open source software developers. In *ICSE '03: Proceedings of the 25th International Conference on Software Engineering*, pages 419–429, Washington, DC, USA, 2003. IEEE Computer Society. ISBN 0-7695-1877-X.

Figurer

5.1	Initiale deltagertyper	40
5.2	Illustration over hvordan roller i Open Source projekter skifter, når deltagerne bidrager til projektet	41
6.1	Fordeling af deltagere	53
6.2	Spørgsmål 15 og 16 fra spørgeskemaet.	57
7.1	Fordeling af arketyper efter gruppering	62
7.2	Typer af information	64
7.3	Arketyper og information	64
7.4	Relation mellem interesse i awareness og deltager type. (1: Ikke interesseret - 5: Meget interesseret)	65
7.5	Relation mellem interesse i awareness og reviderede arketyper. (1: Ikke interesseret - 5: Meget interesseret)	65
7.6	Hvor længe de initielle typer har været involveret i projekterne	66
7.7	Hvor længe de reviderede typer har været involveret i projekterne	66
7.8	Hvor leder typerne først efter information	66
8.1	Figuren illustrere hvor arketyperne differentiere sig, samt hvor de er ens	70
C.1	Eksempel med interval beskrivelsen placeret udfor afkrydsningsmulighederne	10
C.2	Eksempel på et spørgsmål, hvor der skal besvares ud fra et interval	11
D.1	Diagram over spørgsmål 6	13
D.2	Diagram over spørgsmål 7	13
D.3	Diagram over spørgsmål 8	14
D.4	Diagram over spørgsmål 9	14

D.5	Diagram over spørsmål 10	14
D.6	Diagram over spørsmål 11	15
D.7	Diagram over spørsmål 12	15
D.8	Diagram over spørsmål 13	15
D.9	Diagram over spørsmål 19	16
D.10	Diagram over spørsmål 6, arketyper	16
D.11	Diagram over spørsmål 7, arketyper	16
D.12	Diagram over spørsmål 8, arketyper.	17
D.13	Diagram over spørsmål 9, arketyper.	17
D.14	Diagram over spørsmål 10, arketyper.	17
D.15	Diagram over spørsmål 11, arketyper	18
D.16	Diagram over spørsmål 12, arketyper	18
D.17	Diagram over spørsmål 13, arketyper	18
D.18	Diagram over spørsmål 15, arketyper	19
D.19	Diagram over spørsmål 15, i forhold til arketyper	19
D.20	Diagram over spørsmål 17, arketyper	19
D.21	Diagram over spørsmål 19, arketyper	20
D.22	Diagram over spørsmål 20, arketyper	20
D.23	Diagram over spørsmål 21, arketyper	20
D.24	Diagram over spørsmål 23, arketyper	21
D.25	Diagram over spørsmål 25, arketyper	21

Tabeller

2.1	Informationskanaler	11
5.1	Licenserne som Open Source projekterne er udgivet under.	42
5.2	Tabel over informationskanaler i Open Source projekterne	43
8.1	Personer som udvælges fra de tre arketyper	71
E.1	Justering af spørgsmål 8 og 9	23
F.1	Personer som udvælges fra arketype 1	25
F.2	Personer som udvælges fra arketype 2	26
F.3	Personer som udvælges fra arketype 3	26

Mail udsendt på informationskanaler

E-mailen udsendt til Open Source projekterne omkring spørgeskemaundersøgelsen:

Hi

We are a group of 4 master students at Aalborg University Denmark, who are writing a thesis about communication in open source projects. As an part of this, we are conduction a survey, and we would like your input, it does not matter if you are a user or a developer, we need everybody's input.

If you can spare about 15 minutes filling out this survey: http://www.cs.aau.dk/~tomas/survey/intro/intro_gimp.html we would appreciate it very much.

We thank you for your time.

—

Kind Regards

*Signe, Rasmus, Tomas and Henrik
d624a (at) cs (dot) aau (dot) dk*

E-mailen er udsendt på de følgende informationskanaler, url'en henviser til hvor den kan findes:

Mailing lister:

- GIMP User: <http://www.mail-archive.com/gimp-user@lists.xcf.berkeley.edu/msg07381.html>
- GIMP Developer: <http://www.mail-archive.com/gimp-developer@lists.xcf.berkeley.edu/msg09370.html>
- MPlayer-dev-eng: <http://lists.zerezo.com/mplayer-dev-eng/msg00006.html>
- MPlayer-users: <http://www.newsarch.com/archive/maillinglist/mplayer/user/msg10780.html>
- Gaim-devel: http://sourceforge.net/mailarchive/message.php?msg_id=11266523

Fora:

- Gaim Open Discussion: http://sourceforge.net/forum/forum.php?thread_id=1254465&forum_id=665

Spørgeskema

Introduktion til spørgeskemaet:

“The purpose of this survey is to study how people search and find information in different open source projects.

All answers will be kept confidential, only group members from the university project group will have access to them.

Continue on to the survey

(If you have any questions write us)”

1. Age

2. Gender

- *Male*
- *Female*

3. Where are you from?

- *(liste af lande)*

4. E-mail address (Only supply your e-mail address if we may contact you later, to conduct a e-mail interview)

5. Where did you first hear about this survey?

- *Mailing list*
- *Forum*
- *Chat (e.g. IRC)*
- *Homepage*

- *Other*

6. In which way are you involved in this project?

- *Project Leader (is often the person(s) who has initiated the project)*
- *Core Member (responsible for guiding and coordinating the development of an OSS project and have made significant contributions to the development and evolution of the project)*
- *Active Developer (regularly contribute new features and fix bugs)*
- *Peripheral developer (occasionally contribute new functionality or features to the existing system)*
- *Bug Fixer (fixes bugs that either they discover themselves or are reported by other members)*
- *Bug Reporter (only discover and report bugs)*
- *Reader (people who reads mailing-list or forums, for own purpose and learning)*
- *Passive User (a person which use the software without contributing to the community, but just using the software as if it was proprietary)*

7. How long time have you been using this project?

- *less than 1/2 a year*
- *between 1/2 and 1 year*
- *between 1 and 2 years*
- *between 2 and 3 years*
- *more than 3 years*

8. How long have you been reading/posting on this project's...

(1: never, 2: less than 1 year, 3: between 1 and 2 years, 4: between 2 and 3 years, 5: more than 3 years)

- *Forum*
- *Mailinglist*
- *Chat (e.g. IRC)*
- *Newsgroup*
- *Code repository (e.g. CVS)*
- *Bug and issuetracker (e.g. Bugzilla)*

9. **How often do you read/post on this project's...**

(1: never, 2: monthly, 3: weekly, 4: daily, 5: several times daily)

- *Forum*
- *Mailinglist*
- *Chat (e.g. IRC)*
- *Newsgroup*
- *Code repository (e.g. CVS)*
- *Bug and issuetracker (e.g. Bugzilla)*

10. **When you are searching for information, where do you look first?**

- *Forum*
- *Mailing list*
- *Chat (e.g. IRC)*
- *Newsgroup*
- *Code repository (e.g. CVS)*
- *Bug and issuetracker (e.g. Bugzilla)*
- *Other*

11. **If you did not find the information the first place you looked, where would you then look?**

- *Forum*
- *Mailing list*
- *Chat (e.g. IRC)*
- *Newsgroup*
- *Code repository (e.g. CVS)*
- *Bug and issuetracker (e.g. Bugzilla)*
- *Other*

12. **Do you experience that you have to search several sources (e.g. mailing lists, news groups and so on) to find the information you are looking for?**

- *Yes*
- *No*

13. **When you are searching for information, which sources do you search: "official" sources (e.g. directly related to the project) or 3. party (e.g. google, a general news group, or another person) or both?**
 - *Official*
 - *3. Party*
14. **Why do you search in the sources you have selected above?**
15. **Do you experience it to be difficult/easy to find the information you need in this project? (1: Difficult - 5: Easy)**
 - *Your experience*
16. **Why do you experience it to be difficult/easy to find the information?**
17. **Which two kinds of information are you usually looking for?**
18. **Where are you looking for those two kinds of information?**
19. **How interested are you in knowing what people... (1: Not at all - 5: Very much)**
 - *currently are working on in this project?*
 - *have been working on in this project?*
 - *are going to work on in this project?*
20. **If you want to know who has modified an artifact (e.g. source code, documentation and so on) where do you look?**
 - *Forum*
 - *Mailing list*
 - *Chat (e.g. IRC)*
 - *Newsgroup*
 - *Coderepository (e.g. CVS)*
 - *Bug and issuetrackers (e.g. Bugzilla)*
 - *Other*
21. **Did you experience it difficult/easy to find information... (1: Difficult - 5: Easy)**
 - *when you joined this project?*
 - *now?*

22. **Why do you experience it to be more difficult/easier now?**
23. **If you have the choice, do you prefer to ask a question to get the information you need, or do you prefer to search for it?** (if you prefer different approaches depending on the information you are searching for, please write it in question 24)
- *Ask a question*
 - *Search information*
24. **Why do you prefer the above?**
25. **If you have found the information you searched for, would you like to have the opportunity to post a followup message (e.g. you need more information, or you would like to add some information)?**
- *Yes*
 - *No*
26. **Why would you like to have that opportunity or not?**

Pilotforsøg

For at teste spørgeskemaet udføres et pilotforsøg. Forsøget har til formål at belyse fejl og mangler ved spørgeskemaet. Det kan være problemer angående forståelse af spørgsmål eller fejl i designet. Forsøget er delt op således, at der laves forsøg med to personer, dernæst en evaluering samt retning af fejl og efterfølgende et forsøg med to andre personer. Pilotforsøget er delt op for at kunne evaluere og rette fejl undervejs. Herved undgås, at det er de samme fejl, der påpeges af alle fire forsøgspersoner.

Pilotforsøget udføres som tænke-højt¹ forsøg. Ideen med tænke-højt-forsøget er, at forsøgspersonen skal fortælle, hvad han tænker, mens han udfylder spørgeskemaet. Imens observeres personen og dennes tanker nedfældes.

Som opfølgning på tænke-højt-forsøgene følger dialoger med forsøgspersonerne omkring deres oplevelse af at udfylde spørgeskemaet. Samtalen angår designet, længde af spørgsmål og deres opbygning samt forståelsen heraf. Dette bliver gjort for at mindske risikoen for at sikre, at personerne når ind på de centrale elementer af spørgeskemaet.

Forsøgspersoner

De fire personer, der udvælges som forsøgspersoner er alle bekendt med tænke-højt-tests. Det er en fordel, at personerne har erfaring med at udføre lignende forsøg, idet de herved kan koncentrere sig mere om selve spørgeskemaet og ikke selve testsituationen.

Forsøgspersonerne har endvidere kendskab til informationskanalerne i projekterne. Yderligere har to af forsøgspersonerne også en dybdegående viden omkring Open Source gennem erfaringer og brug af Open Source software. Dette gør, at de forstår spørgsmålene i spørgeskemaet, og at de kan imitere deltagere i projekterne.

¹ Kort beskrivelse af metoden kan se på <http://www.netsteder.dk/raad/brugertest/taenke-test.html>

Resultaterne fra pilotforsøget

I pilotforsøget gør forsøgspersonerne opmærksom på, at det vil være en god ide med en introduktionsside til spørgeskemaet. Dette er også hensigten, men siden var ikke implementeret ved forsøgets start. Dette kan påpeges som en fejl i pilotforsøget, da siden også er en del af selve spørgeskemaet.

Et andet problem, som en af forsøgspersonerne påpeger er, at der i spørgeskemaet ikke er en definition af, hvad der menes med "information". Svaret på denne problemstilling er, at vi ikke vil lægge ordene i munden på spørgeskemaespondenterne og derfor har fravalgt en definition.

Why-spørgsmål kommenteres også af forsøgspersonerne. Kommentarerne lyder på, om vi er bevidste om at ikke alle respondenter vil svare på disse spørgsmål. Formålet med spørgsmålene er, at respondenterne kan uddybe deres svar og det er således ikke påkrævet at de besvarer dem. Svarene på disse spørgsmål vil formodentlig give brugbar information til at besvare hypoteser.

1.	Hvad synes du på en skala fra 1-5?					
	1	2	3	4	5	
1: Godt - 5: Ikke godt		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figur C.1: Eksempel med interval beskrivelsen placeret udfor afkrydsningsmulighederne

En ting, som der ikke har været opmærksomhed omkring under udarbejdelsen af skemaet, er konsistens omkring intervaller. I spørgsmål af typen "*Hvad synes du på en skala fra 1-5*" har intervallerne været placeret forskellige steder. Nogle gange stod intervallet under teksten som i figur C.2, og andre var de placeret udefor afkrydsningsfelterne, se figur C.1. Dette virkede som et forstyrrende element for vores forsøgspersoner. For at undgå denne forstyrrelse rettes det. Så intervallet er placeret som vist på figur C.2.

Der er endvidere et ønske fra forsøgspersonerne om, at intervallerne skrives over afkrydsningsfelterne, så der i stedet for 1 står difficult og i stedet for 5 står easy, se figur C.2. Men da det ikke er muligt, som standard, i det software, phpESP², vi bruger til udarbejdelse af spørgeskemaet, har vi valgt ikke at gøre som ønsket. Da phpESP er udgivet under GPL-Licensen, er der mulighed for selv at udvikle og udvide softwaren med den ønskede funktion, men vi har valgt at bruge vores ressourcer på selve spørgeskemaet frem for udviklingen af denne egenskab.

Spørgsmålet "*When you are searching for information, where do you look first?*" vil flere af forsøgspersonerne gerne krydse af mere end et sted. Ud fra det, kan vi konkludere, at det ikke fremgår tydeligt nok, at der kun tænkes på *look first* og at det således kun kan være ét sted. Problemet løses efter første del af pilotforsøget, ved at formattere teksten "*look first*" som understreget. Som resultat heraf er der ikke blevet knyttet kommentarer til dette spørgsmål i anden del af forsøget.

² <http://phpesp.sourceforge.net/>

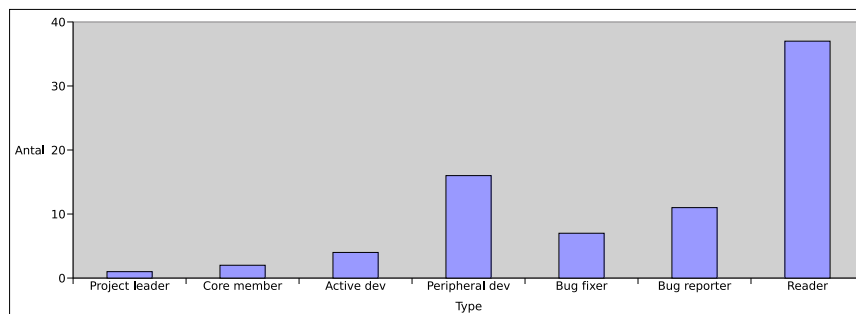
*15.	Do you experience it to be difficult/easy to find the information you need in this project? (1: Difficult - 5: Easy)				
	1	2	3	4	5
Your experience	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figur C.2: Eksempel på et spørgsmål, hvor der skal besvares ud fra et interval

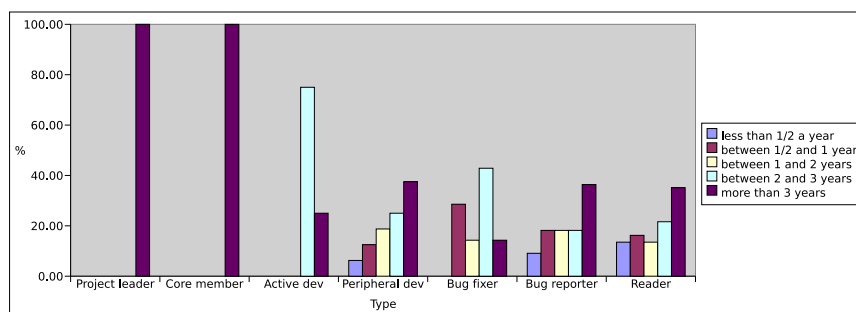
Forsøgspersonerne påpeger, at spørgeskemaet indeholder en del formulerings- og forståelsesfejl. Disse fejl vurderes og rettes.

Data fra spørgeskema

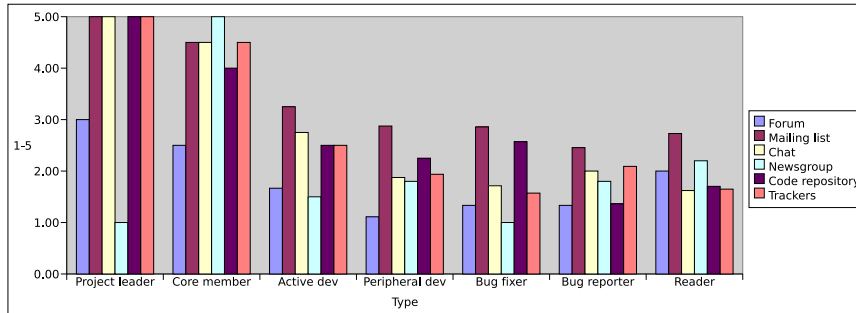
D.1 Diagrammer for 7 initielle typer



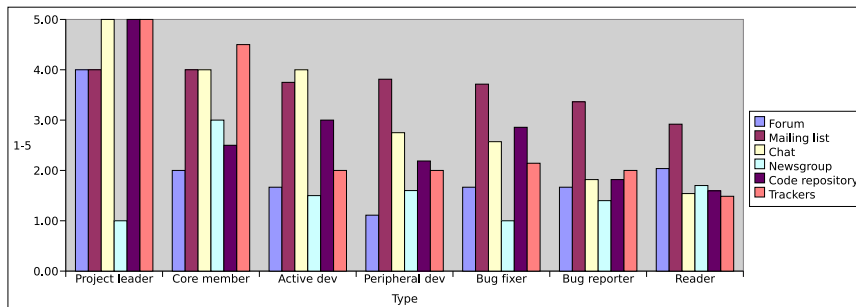
Figur D.1: Diagram over spørgsmål 6: Antal deltagere i de forskellige typer.



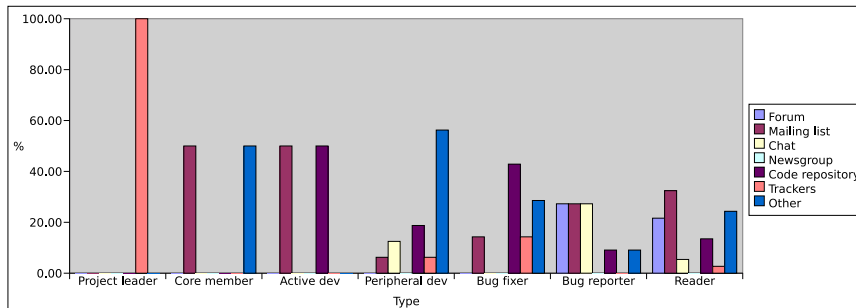
Figur D.2: Diagram over spørgsmål 7: Hvor lang tid deltagerne har været involveret i projekterne.



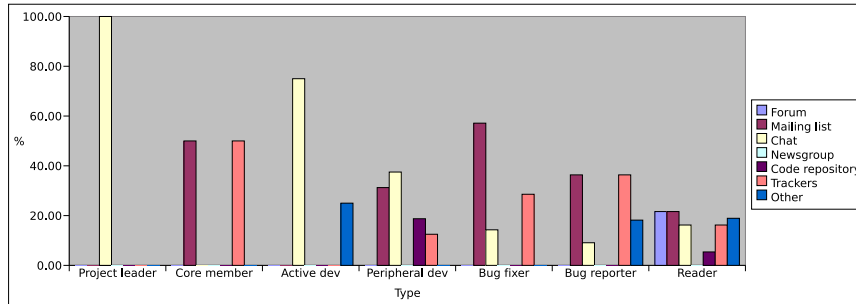
Figur D.3: Diagram over spørgsmål 8: Hvor længe deltagerne har læst/postet på projekts informationskanaler. (1: aldrig, 2: mindre end 1 år, 3: mellem 1 og 2 år, 4: mellem 2 og 3 år, 5: mere end 3 år)



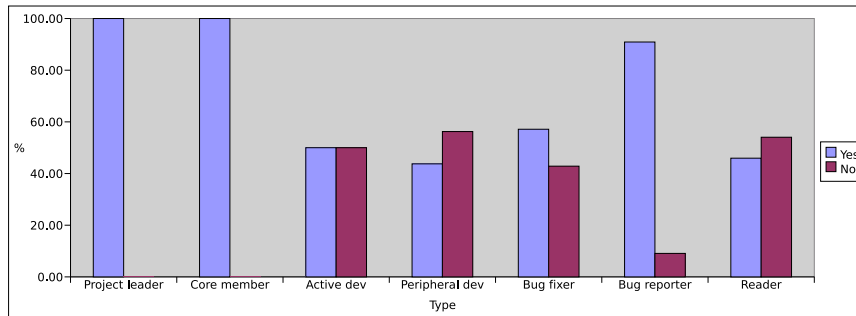
Figur D.4: Diagram over spørgsmål 9: Hvor ofte deltagerne læser/poster på projekts informationskanaler. (1: aldrig, 2: månedligt, 3: ugentligt, 4: dagligt, 5: flere gange dagligt)



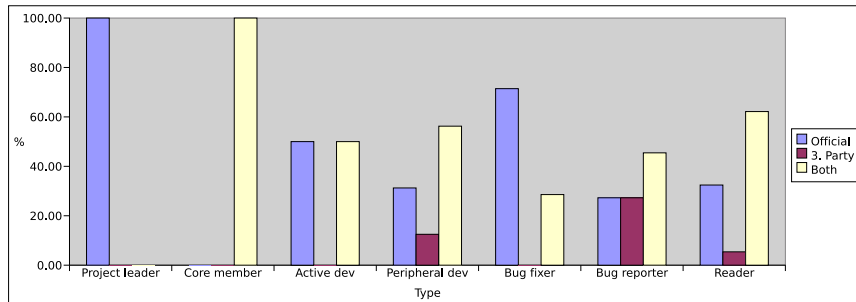
Figur D.5: Diagram over spørgsmål 10: Hvor deltagerne først leder efter information.



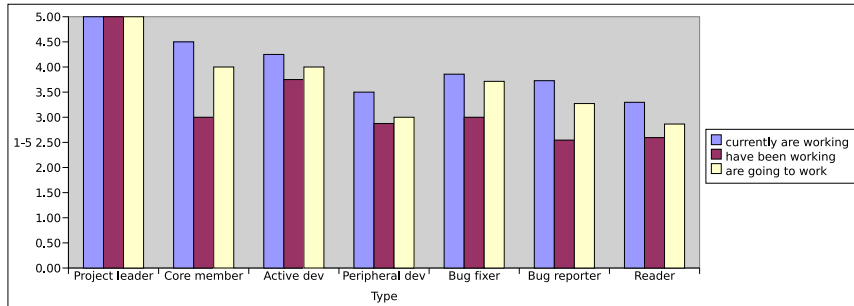
Figur D.6: Diagram over spørgsmål 11: Hvor deltagerne dernæst leder efter information.



Figur D.7: Diagram over spørgsmål 12: Søger deltagerne i flere informationskanaler for at finde det, de leder efter.

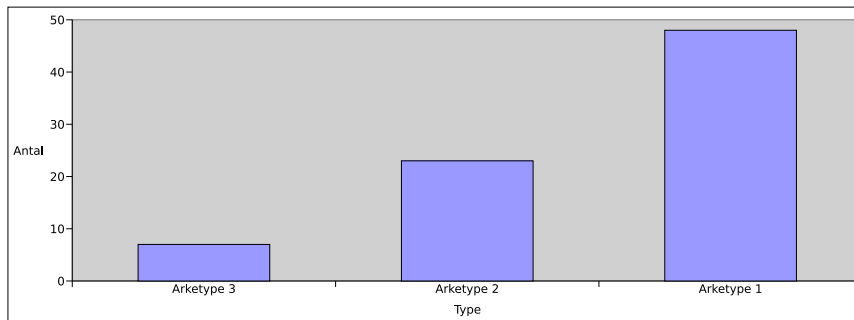


Figur D.8: Diagram over spørgsmål 13: Søger deltagerne efter information i de officielle, 3.parts eller begge typer informationskanaler.

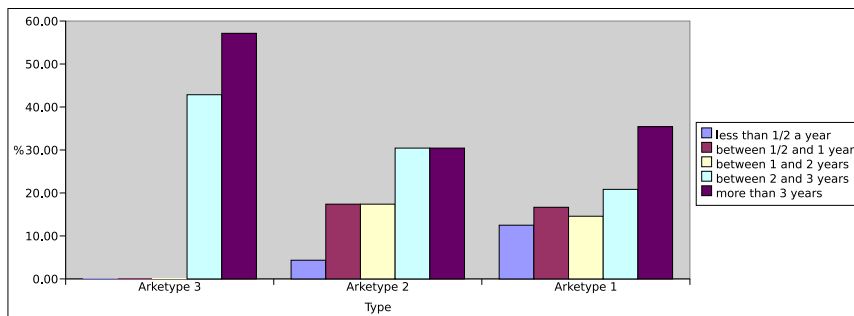


Figur D.9: Diagram over spørgsmål 19: Deltagernes interesse i awareness information. (1: Ikke interesseret - 5: Meget interesseret.)

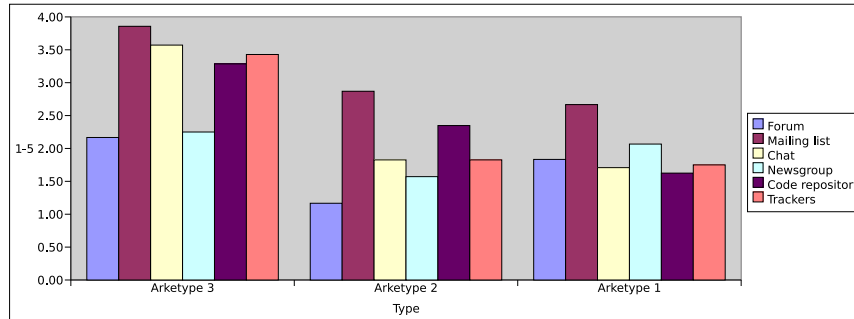
D.2 Diagrammer for reviderede typer



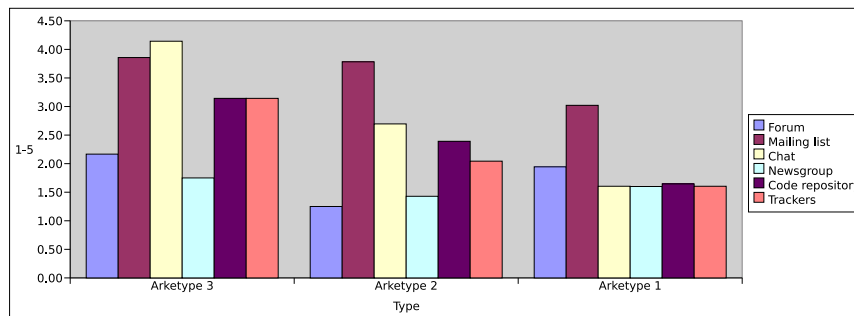
Figur D.10: Diagram over spørgsmål 6, arketyper: Antal deltagere i hver arketype



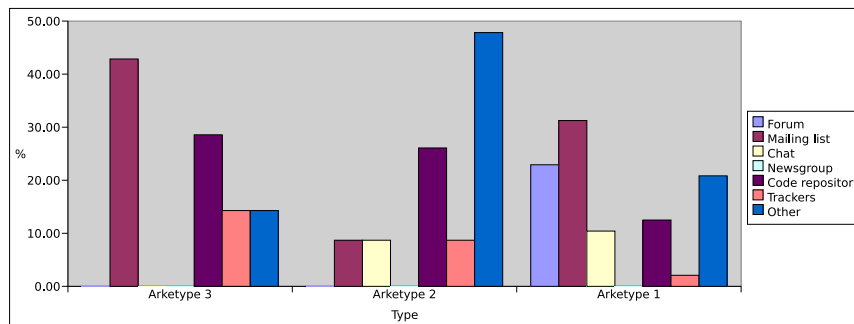
Figur D.11: Diagram over spørgsmål 7, arketyper: Hvor lang tid arketyperne har været involveret i projekterne.



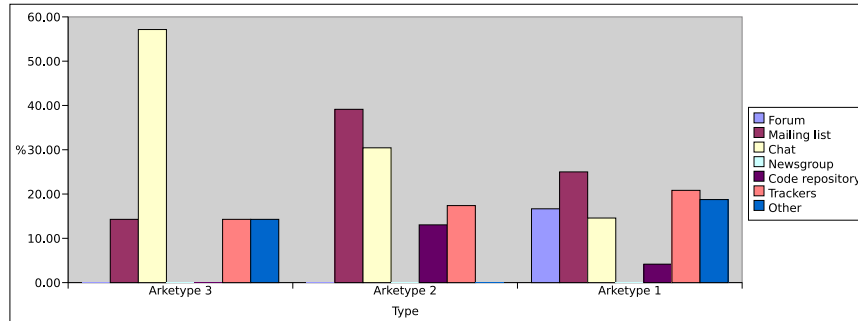
Figur D.12: Diagram over spørgsmål 8, arketyper: Hvor længe arketyperne har læst/postet på projekts informationskanaler (1: aldrig, 2: mindre end 1 år, 3: mellem 1 og 2 år, 4: mellem 2 og 3 år, 5: mere end 3 år)



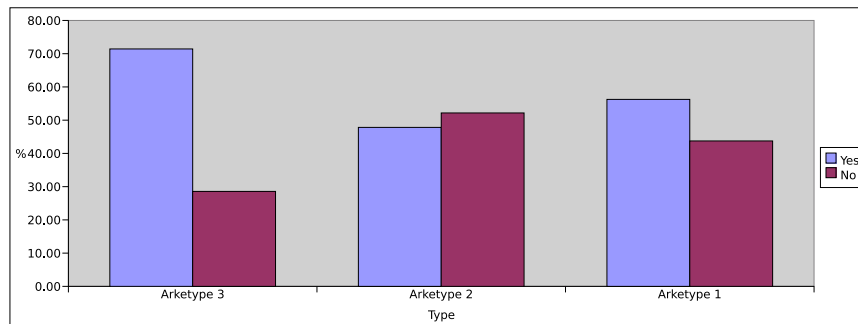
Figur D.13: Diagram over spørgsmål 9, arketyper: Hvor ofte arketyperne læser/poster på projekts informationskanaler. (1: aldrig, 2: månedligt, 3: ugentligt, 4: dagligt, 5: flere gange dagligt)



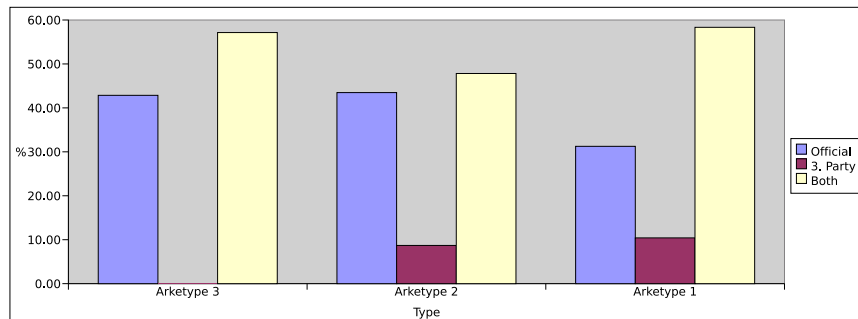
Figur D.14: Diagram over spørgsmål 10, arketyper: Hvor arketyperne først leder efter information.



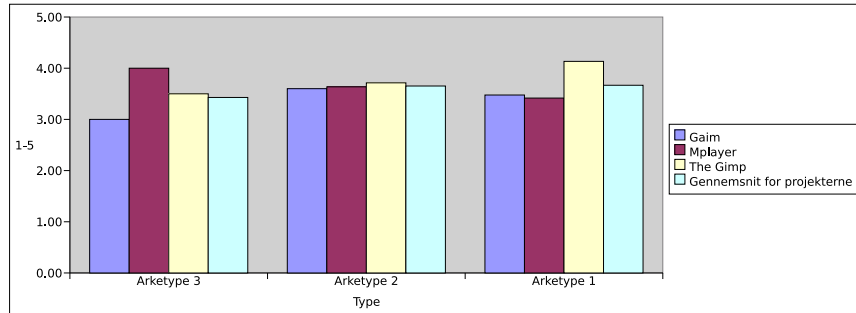
Figur D.15: Diagram over spørgsmål 11, arketyper: Hvor arketyperne dernæst leder efter information.



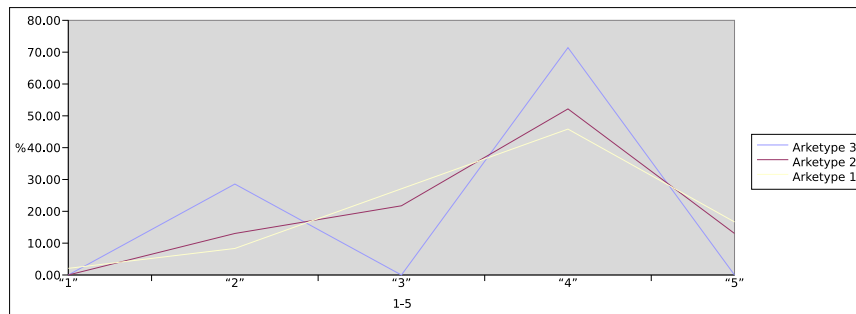
Figur D.16: Diagram over spørgsmål 12, arketyper: Søger arketyperne i flere informationskanaler for at finde det, de leder efter.



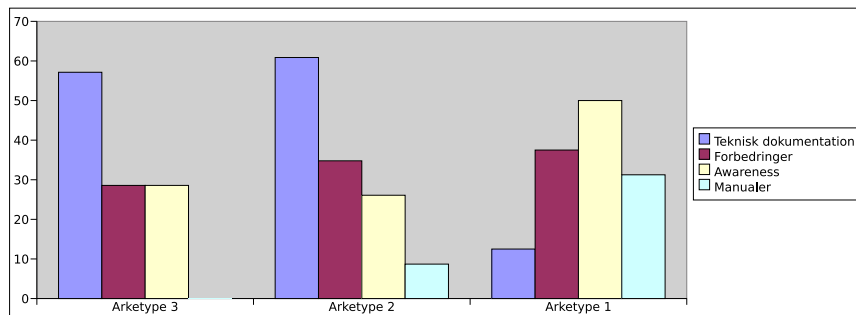
Figur D.17: Diagram over spørgsmål 13, arketyper: Søger arketyperne efter information i de officielle, 3.parts eller begge typer informationskanaler.



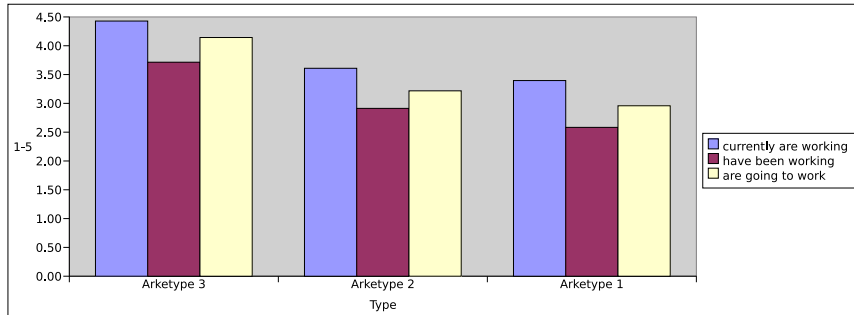
Figur D.18: Diagram over spørgsmål 15, arketyper: Hvor svært/let, synes arketyperne, det er at finde information (1: Svært - 5: Let)



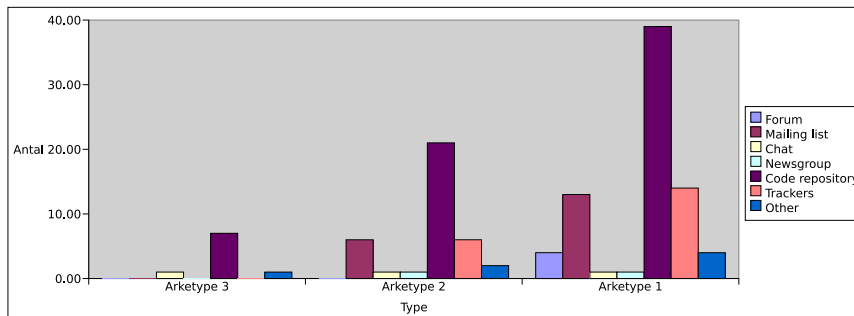
Figur D.19: Diagram over spørgsmål 15, arketyper: Hvor svært/let, synes arketyperne, det er at finde information. Antal procent af hver arketype, som har svaret de forskellige grader af sværhed/lethed. (1: Svært - 5: Let)



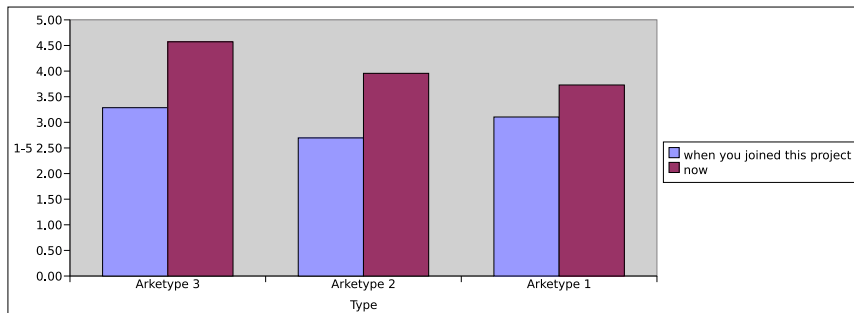
Figur D.20: Diagram over spørgsmål 17, arketyper: Hvilke typer af information arketyperne søger.



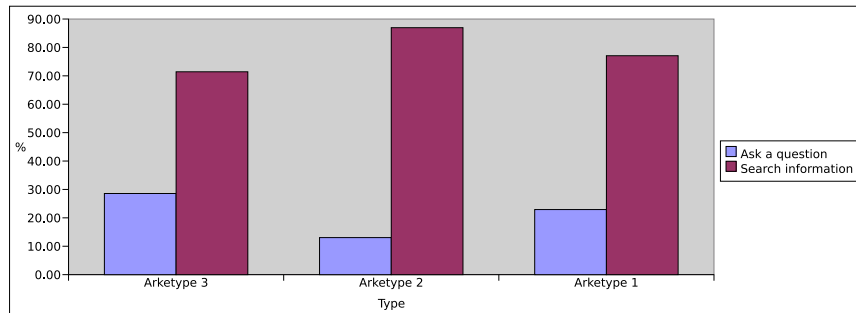
Figur D.21: Diagram over spørgsmål 19, arketyper: Hvor interesseret arketyperne er i hvad andre: arbejder på, har arbejdet på, skal til at arbejde på, (1: slet ikke - 5: meget interesseret)



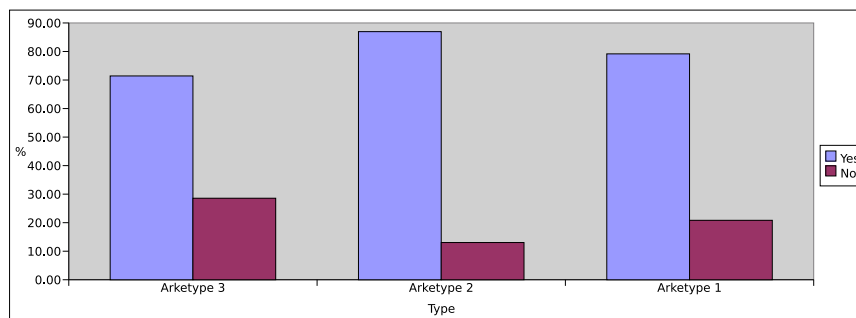
Figur D.22: Diagram over spørgsmål 20, arketyper: Hvor arketyperne søger, hvis de vil vide, hvem der har modificeret en artefakt.



Figur D.23: Diagram over spørgsmål 21, arketyper: Hvor svært/let arketyperne synes det er at finde information både da de startede og nu. (1: Svært - 5: Let)



Figur D.24: Diagram over spørgsmål 23, arketyper: Foretrækker arketyperne at stille spørgsmål eller søge, hvis de skal finde information.



Figur D.25: Diagram over spørgsmål 25, arketyper: Foretrækker arketyperne at kunne give en opfølgende besked til den information, de har fundet.

Justeringer af svar



Det følgende skema viser en oversigt over de deltagere hvis svar er blevet fjernet i beregninger forbundet med spørgsmål 8 og 9 i spørgeskemaet.

Det drejer sig om projekterne MPlayer, ikke i besiddelse af forum og nyhedsgruppe, samt Gaim der ikke har nogen officiel nyhedsgruppe.

Formatet brugt i tabellen er det følgende: XY:Z. X repræsenterer antallet af deltagere, Y er enten M for MPlayer, eller G for Gaim. :Z er summen af de svar der er givet.

	Spg. 8 - Forum	Spg. 8 - Nyhedsgruppe	Spg. 9 - Forum	Spg. 9 - Nyhedsgruppe
Project leader				
Core member		1G:1		1G:1
Active developer	1M:1	1G:1 , 1M:1	1M:1	1G:1 , 1M:1
Peripheral developer	7M:7	7M:7	7M:7	7M:M
Bug fixer	4M:7	1G:1 , 4M:5	4M:5	1G:1 , 4M:7
Bug report	2M:2	4G:6 , 2M:2	2M:2	4G:7, 2M:2
Reader	10M:14	17G:20 , 10M:13	10M:12	17G:17, 10M:11

Table E.1: Justering af spørgsmål 8 og 9

Repræsentative deltagere

ID	Type information	Tid	Primær info-kanal	a current	a have been	are going to
108	Awareness	between 1/2 and 1 year	Mailing list	2	2	2
180	Awareness	less than 1/2 a year	Forum	3	3	3
188	Awareness	between 1 and 2 years	Forum	3	3	2

Tabel F.1: Personer som udvælges fra arketype 1

ID	Type information	Tid	Primær info-kanal	a current	a have been	are going to
107	Teknisk dokumentation	more than 3 years	Other	4	5	4
114	Teknisk dokumentation	between 2 and 3 years	Chat (e.g. IRC)	5	3	4
171	Teknisk dokumentation, awareness	more than 3 years	Other	4	5	3

Tabel F.2: Personer som udvælges fra arketype 2

ID	Type information	Tid	Primær info-kanal	a current	a have been	are going to
207	intet svar	between 2 and 3 years	mailing list	5	5	5
121	Teknisk dokumentation	between 2 and 3 years	code repository	5	4	3
204	Teknisk dokumentation	between 2 and 3 years	mailing list	5	4	5

Tabel F.3: Personer som udvælges fra arketype 3

Interview besvarelses

Det følgende er deltagerens svar på interviewet.

Teksten markeret med **fed** er, det projekt deltageren tilhører samt deltagerens type og ID. Den kursive tekst, som står efter informationerne omkring deltageren, er deltagerens svar til spørgeskemaet, som benyttes i interviewspørgsmålet. I de tilfælde, hvor deltageren ikke har besvaret spørgsmålet, er der ingen tekst efter informationen om deltageren. Til spørgsmål 5 er deltagerens svar markeret med del <1,2,3 >, der indikerer hvilken del af spørgsmål 5 de besvarer.

1. *Which kinds of information do you consider easy to find? Or difficult to find? To answer this question you can describe situations in which you found it easy/difficult to find information. Please also state what you think the reason is?*

The Gimp, reader, 149 (I unfortunately can't think of any answer to this off the top of my head! I guess I do not look for help with Gimp much these days!)

Gaim, reader, 126 Well, speaking of the Gaim project, general information is pretty easy to find, there are docs and FAQs, and friendly users in the forums. What seems to be hard to get at times is reliable answers, as the developers don't seem to take much time to do this. I'm not putting any special emphasis on if they should or not though, merely an observation. The forums and tools at sourceforge aren't very good for navigating or searching either, since they are both slow and hard to get good overviews in. That goes for all projects that uses SF's standard tools, it's a pain to navigate which makes one reluctant to do so, and more prone to giving up early.

The Gimp, bug reporter, 124 Difficult things for me are programs/supporting libraries. Sometimes the READMEs tell you where to find things, sometimes a failed configure will tell you, but most time dependancies can be hard to track down.

Other times, the programs are so constructed that only the gods that made them know how to fix it. I often dream of the next 'must have' invention that is attached to your monitor, reads the puzzlement in your eyes, and correctly feeds Google to provide the answer as number 1 of 50 million hits.

The Gimp, peripheral, 118 Details on programming for a non-programmer. I have had to write some scripts and the necessary calls are not clearly documented for those that need to learn scripting and how to do things in less than 24 hours.

example: `gimp_drawable_get_pixel`

How do you get all the necessary data and set it up. Or working with layers in a script.

It would be nice if there were clear cut examples for each of the scripts for the NOOBS like me.

MPlayer, peripheral, 161 Because I know the developers and chat with them daily, I don't think there's ever been a question I asked and didn't get an answer on. Be it MPlayer's internals, why a certain bug appears, how to use MPlayer to do something, MPlayer's behavior, etc. etc.

MPlayer, peripheral, 107 You may find difficult the usual "paper" information as in brochure or "toilet paper", since it has no use for the people developing mplayer and related project (eg: NUT, Snow) you won't find any Introductory, advertising, commercial like documents. The rest is present and most times some comments are requested in order to commit new features and adding man page /xml documentation is getting compulsory.

Overall is easy to find documentation about how to use mplayer or how to report bugs.

The Gimp, active developer, 79 Usually the information that is easy to find is how to use the program. Information that is medium to find is exactly what the code does. Information that I found hard to find is how the program behaved in the past, or how exactly it's supposed to behave.

2. *In the survey you answered that you primarily use the <informationskanal fra spørsmål 10 > to search for information. What do you like about <informationskanal fra spørsmål 10 >? Is it for example the possibility to ask others, the design or is it because it's widely used or something else?*

The Gimp, reader, 149 (Chat (e.g. IRC)): I like chat because it's realtime and also gives me the opportunity to give back to the community by helping

others. Even if I'm not participating and just watching a chat, or reading chat messages from earlier in the day, I feel like I'm part of the active community, rather than just a detached end-user.

Gaim, reader, 126 (*Forum*): A lot of it is probably force of habit - I'm for some reason more comfortable with forums than for instance mailing lists or IRC. I think that forums probably represent a good middle ground there, easy to post and reply without having to fire up an external program (as with mail) at the same time searchable and persistent (not volatile like IRC). So, for any project that provides them, I tend to use the forums first and foremost if it is not apparent that some other channel is much more efficient.

The Gimp, bug reporter, 124 (*mailing list*): Less spam than newsgroups + more focused group

The Gimp, peripheral, 118 (*Bug and issuetracker (e.g. Bugzilla)*): Make searching much easier.

I look at the bug and issue tracker on the site as I have most of the necessary documentation either printed or available from other sites. The Gimp site is changing and getting more organized.

MPlayer, peripheral, 161 (*Chat (e.g. IRC)*): I like IRC because it's realtime chat, and an actual conversation. With almost any other communication, you talk in "chunks". IRC resembles most a real life conversation. It's disadvantage is it's hard to talk about large pieces of code or anything of the sort. you can only talk one line at a time.

MPlayer, peripheral, 107 (*Chat (e.g. IRC)*): Is quite easy ask suggestion and send the code you are just developing. Since you don't have the delay you may expect from email and forum (not to mention bts) you can get feedback faster.

The Gimp, Active developer, 79 (*Code repository (e.g. CVS)*): Having grown accustomed to Subversion, the CVS successor, I like CVS less and less as time goes by. What I like about it is the fact, I have the source code for the entire application on my hard disk and I can search it, browse it and look for information.

3. *We asked you in the survey which two kinds of information you usually look for are looking for, and you answered: <deres svar til spørsmål 17 >. What is your motivation for searching this information? Is it for example because you: have problems using the software/want to improve your programming skills/want to learn new things/are interested in open source/or something entirely different?*

The Gimp, reader, 149 (*How to use additional peripherals (printer, scanner, drawing tablet)*): It is for practical reasons. How to print? How to get the drawing tablet working for my wife to use? Etc. :)

Gaim, reader, 126 (*Features I want or need, Solutions for problems and/or bugs*): Most of the time I'm already using a software or I am evaluating its features, so while I do this I come up with things I need or want to do, and thus I look for ways to do this. Sometimes there are plugins that can do this, people provide fixes and so on. Hunting for bugfixes is pretty much the same, I run into a problem of some kind and try to see if anyone knows of a solution.

The Gimp, bug reporter, 124 (*howtos*): I have neither the talent nor intuition to be a great artist, composer or writer, and I believe How-tos are written by the masters or gods, and if I slavishly follow their lead, I too will become great. (Oh well, you can always hope)

The Gimp, peripheral, 118 (*Coding - I have to automate some procedures for work. Sample scripts or other examples to learn from.*): Work. I need it for doing one of my tasks. As I said, I need to automate some procedures within Gimp.

MPlayer, peripheral, 161 (*How the program works. How to use the program to do a specific task.*): I like learning how the program works merely because it interests me. I see a program from the outside, and I find myself asking often "how does it do that". How to do a specific task, simply because I'm trying to do something, and I want to do it in the best possible way.

MPlayer, peripheral, 107 (*bug discussion, coding related information (API, coding style, missing features, in development features)*): I'm a gentoo developer and one of my task (if not the main task actually) is fix issue in the software we provide. On the other hand I'm involved in the project itself and if you want to contribute in an effective way you have to send the updates in the right form.

The Gimp, Active developer, 79 (*API documentation*): I'm just looking for issues to tackle, and often need to know how to do a certain thing in GIMP or gtk+. So I'm using the API documentation.

4. *In your answers to the following questions in the survey you estimated your need for information to be <værdier fra deres svar i spørgsmål 19 >.*

- *How interested are you in knowing what people... (1: Not at all - 5: Very much)*
 - *currently are working on in this project?*
 - *have been working on in this project?*

– *are going to work on in this project?*

Please elaborate on why you chose this? You can give examples of when you needed this information, and why you needed it.

The Gimp, reader, 149 (<4; 3 and 2>): Mostly out of interesting. Who in the community is working on what projects? Perhaps they can help with other useful projects.

Gaim, reader, 126 (<5; 4 and 5>): Its always interesting to know who works on a project. If it's an important project to me personally, I usually get to know their track records and online personalities in a way. This can help judging a lot of things, but most of all the health of a project. Not the least that there **are** people working on it. It's hard to elaborate on this one, I just want to know. =)

The Gimp, bug reporter, 124 (<1; 1 and 1>): I once met a great man, Andrew Tridgell of Samba fame. That increased my inferiority complex (IC) 200% ! The other great guy I met was Richard Stallman, but he only raiser my IC 150%. But seriously, names are names and from in front of a monitor, a name does little, can change and conjures little in the mind of the unimaginative.

After reading mailing lists for years you know what name is associated with what type of response. When a stranger intrudes, his (her) comments are instantly assigned a value based on the comment, not on the persons background or claim to fame.

The Gimp, peripheral, 118 (<2; 2 and 2>): What is happening on the project in the development phase isn't much interest to me as a user. I do follow some but not all the issues. It may affect something that I am working on today when there is an upgrade or patch.

MPlayer, peripheral, 161

MPlayer, peripheral, 107 (<4; 5 and 4>): To ask the right question to the right people mostly.

The Gimp, Active developer, 79 (<2; 2 and 3>) : Excuse me? What's <2; 2 and 3>.

5. *Do you think that factors like the amount of users, information, and information channels has an influence on how it is to find information in this project?*

Please describe why you think these factors have or don't have influence?

Are there other factors that influence finding information? Please describe which factors and why?

The Gimp, reader, 149 1. del: Yes. :)

2. del: Simply due to the information and people being more wide-spread, and hence easier to find. :)

Gaim, reader, 126 1. del: Yes, of course. Is there any project where it doesn't in some way or other?

2. del: The right tool for the job goes for information too, so what channels are chosen and monitored of course have impact one way or the other. See my comments on Forums/Lists/IRC above.

Since quite a lot of the information is provided by users, for instance in the forums, they also have an impact.

3. del: Good search and good navigation is imperative - of course, that the info is available is the first one. ;-) With available, I mean either already written down or possible to get by asking the right questions (like from other users in a forum or IRC).

The Gimp, bug reporter, 124 As a general rule, I think that the more users=more questions More questions->better documentation so that people don't have to ask so many questions.

Now the real problem is how to get the users to read the documentation.

And of course there is time. If you have all day, you can generally find what you are looking for.

Mailing lists can provide quick answers but not for those who dont RTFM.

The Gimp, peripheral, 118 With the growth of The Gimp, it is important to get information in a place and format that makes it easy for new users to learn the basics and for the advanced users to learn how to be more efficient.

Many users will go to the www.gimp.org to either see about The Gimp or learn how to use some tools. I have been away from my Gimp usage for the last few months but will be hard into it within weeks. In the past there were broken links to sites that did not exist anymore. This looks better than it did then.

The information must be available and be in a format that is what a user needs. Not all people looking for programming information are programmers. Some are people like me that are looking at making a hard job easier by programming or scripting the routine stuff.

Broken links on a WWW site can be very frustrating. Comparisons or tutorials on how to move from commercial products to The Gimp would also be useful.

To many avenues can make it hard to find that one piece of information you are looking for. Looking at the WWW site now, it is much better than the last time I really looked at it. The layout is much better designed.

Many times in the past I have ended up doing searches on Google to find the correct answer that was not on the site.

Documentation available in pdf format is one very nice feature. I wish all documentation was available like this.

MPlayer, peripheral, 161

MPlayer, peripheral, 107 1. del: The amount of users and even more non educated users is proportional to the amount of noise you get if you have open bts/ml/chat and so on.

On the other hand if you have many educated, users that at least know how to read a manual and how to look for already answered question before asking their questions, that would help tracking issues faster.

2. del: It has influence since if an user can find in an easy way answers to his/her questions,he/she won't ask them again to you.

3. del: I can't think about anything that could be useful to you, sorry

The Gimp, Active developer, 79 The number of users affect the amount of time usage issues can be resolved easily. However, the information about how to develop the package is usually not affected by this directly.

Spørgskema besvarelser