

Webcrawling til analyse af hjemmesider

Afgangsprojekt udført hos Morning Train Technologies ApS

Macintosh HD:Users:rubatharisan:Downloads:morning-train-logo-left-black-big.png

**Udarbejdet af:**

Rubatharisan Thirumathyam

[ruthi12@student.sdu.dk](mailto:Ruthi12@student.sdu.dk)

[rt@morningtrain.dk](mailto:rt@morningtrain.dk)

Eksamens nr.: 324539

**20. maj 2017**

**Vejledt af:**

Ole Dolriis **(SDU)** [od@mmmi.sdu.dk](mailto:od@mmmi.sdu.dk)

Bjarne Bonde **(Morning Train)** [bb@morningtrain.dk](mailto:bb@morningtrain.dk)

**Synopsis**

**Forord**

Dette projekt er udarbejdet som et afslutningsprojekt, på uddannelsen som diplomingeniør indenfor Information- og kommunikationsteknologi.

Projektet er lavet i samarbejde med Morning Train Technologies ApS (ref. Morning Train), hvor jeg har i fællesskab med Bjarne Bonde (CTO v/Morning Train) defineret projektets vision. Dette projekt er blevet vejledt af Ole Dolriis (lektor v/SDU) og Bjarne Bonde. Under projektforløbet har jeg fået råd fra alle konsulenter fra webbureaet Morning Train.

Som led i mit uddannelsesforløb har jeg haft et praktikophold på 6 måneder hos Morning Train. Morning Train ledelsen har herefter valgt at ansatte mig som serveransvarlig, hvor mit primære ansvarsområde har været serveropsætning, servervedligeholdelse og serversikkerhed. Mit sekundær ansvarsområde har været fullstack udvikling af systemer. Da jeg startede på dette projekt, havde jeg allerede været ansat i virksomheden i 1 år og 7 måneder.

Dette projekts målgruppe er webbureauer, hjemmesideejere og SEO konsulenter.

**Indhold**

1. Indledning 5

1.1 Problemområde 5

1.2 Læsevejledning 6

2. Planlægning 6

2.1 Udviklingsmetode 6

2.2 Agil softwareudvikling 6

2.3 Scrum 6

2.2 Projektplan 7

3. Research 9

3.1 … 9

3.2 … 9

4. Design 10

4.1 … 10

4.2 … 10

5. Implementation og test 11

5.1 … 11

5.2 … 11

6. Evaluering 12

6.1 … 12

6.2 … 12

7. Konklusion 13

7.1 … 13

7.2 … 13

8. Referencer 14

8.1 … 14

8.2 … 14

9. Appendiks 15

9.1 … 15

9.2 … 15

# 1. Indledning

## 1.1 Problemområde

Morning Train er et webbureau lokaliseret i Odense, drevet af kvalitetsbevidste webnørder. Morning Train specialisere sig primært indenfor tre kerne områder: digital design, intelligente webløsninger og online markedsføring.

Indenfor disse tre kerneområder er der mange problematiske udfordringer, som hver kunne være et projekt i sig selv. Men jeg søgte et projekt som strækkede sig over alle tre kerne områder, et projekt der ville kunne være til gavn for alle tre kerne områder.

Men for at kunne gøre dette, blev jeg nød til at finde ud af præcis hvordan de forskellige afdelinger fungere. Kort beskrevet så når Morning Train udvikler en ny hjemmeside til en kunde, så er der tre afdelinger som tager hånd om opgaven. Morning Train har front-end udviklere som varetager det digitale design og layout opsætning, de har back-end udviklere der får programmering og funktionaliteten på plads. De har online markedsførings konsulenter som analysere og vurdere hjemmesiden for at efterfølgende at optimere på det de føler der vil gøre en forskel for hjemmesidens online tilværelse.

For at forstå hvad mit projekt skulle handle om, besluttede jeg at snakke med medarbejdere fra de forskellige afdelinger, for at høre dem om hvordan et afgangsprojekt ville kunne bidrage til deres afdeling. Frontend afdelingen kunne ikke rigtigt se at jeg kunne lave noget brugbart for dem, deres problemer er primært grafisk orienteret. Online markedsførings afdelingen ville have mig til at lave et system som skulle systematisk søge forskellige søgeord på Google og finde placeringer af en bestemt hjemmeside, dette projekt kunne ikke lade sig gøre, da Google ikke tillader systemer at crawle deres søgemaskine – ret banalt. Backend afdelingen havde nogle spændende projekter som alle kunne have været det perfekte projekt at bruge som afgangsprojekt. F.eks. et projekt hvor man samlede data om brugeradfærden på en hjemmeside, for at derefter bruge dette data til at forbedre hjemmesiden – hvor brugeren bliver centrum.

Men det projekt jeg syntes der vil være til gavn for alle tre afdelinger blev ”webcrawling for analyse af hjemmesider”. I en mere forstående term: en bot som systematisk besøger hver eneste underside på en hjemmeside, for at vurdere om der nogle punkter som der kan forbedres. Vurderinger af siden, kan herefter udvides, til at kunne tage hensyn til flere ting. En vurdering kunne være f.eks. en undersøgelse om alle grafiske billeder ikke overgår en bestemt filstørrelse, da større en billede er, jo længere tid skal en besøgende vente på at billedet bliver fremvist.

## 1.2 Læsevejledning

# 2. Planlægning

## 2.1 Udviklingsmetode

Udviklingsmetoden som jeg har brugt i dette projekt er Scrum, årsagen til at jeg har valgt Scrum er på baggrund af dets agile tilgang til software udvikling. For at kunne forklare hvordan jeg har implementeret Scrum, bliver jeg nød til at forklare hvordan min agile tilgang har været. Min agile tilgang har været baseret på manifesten for agil softwareudvikling.

## 2.2 Agil softwareudvikling

I den agile manifest er der nogle værdier og principper der gør udvikling af software systemer nemmere og.

<http://agilemanifesto.org/>

<https://www.youtube.com/watch?v=Z9QbYZh1YXY>

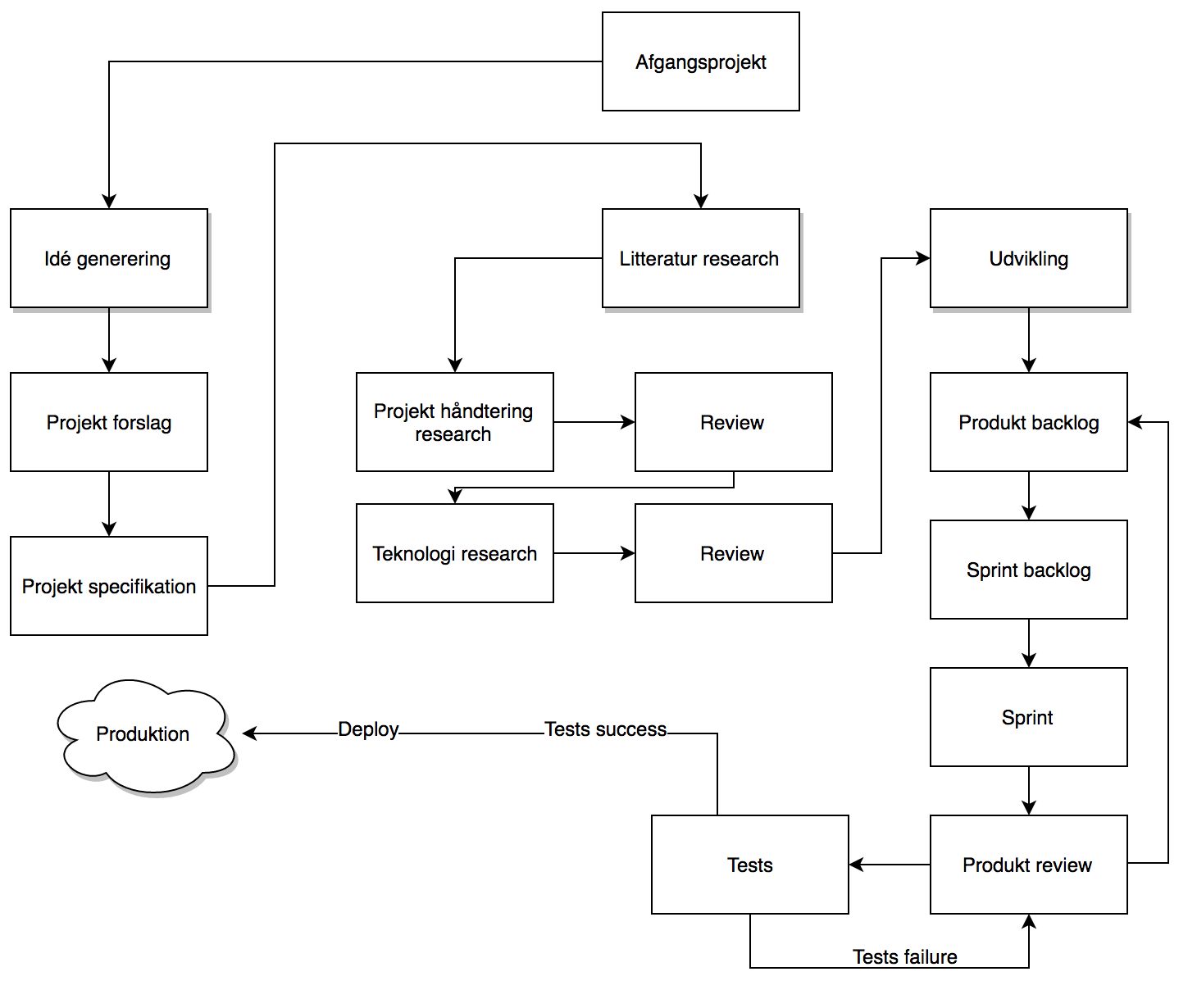
<https://en.wikipedia.org/wiki/Agile_software_development#The_Manifesto_for_Agile_Software_Development>

## 2.3 Scrum

<https://www.youtube.com/watch?v=9TycLR0TqFA>

Hvad, hvorfor.

## 2.2 Projektplan



|  |  |
| --- | --- |
| Aktivitet | Estimeret varighed |
| Idé generering | 1 uge |
| Projekt forslag | 1 uge |
| Projekt specifikation | 2 uger |
| Litteratur research | 1 uge |
| Projekt håndterings research  + Review | 2 uger |
| Teknologi research  + Review | 2 uger |
| Udvikling (opstart) | 1 uge |
| Produkt backlog (iterativt forløb) + Sprint backlog + Sprint + Produkt review + Tests + Deployment | 6 uger |

# 3. Research

## 3.1 …

## 3.2 …

# 4. Design

## 4.1 …

## 4.2 …

# 5. Implementation og test

## 5.1 …

## 5.2 …

# 6. Evaluering

## 6.1 …

## 6.2 …

# 7. Konklusion

## 7.1 …

## 7.2 …

# 8. Referencer

## 8.1 …

## 8.2 …

# 9. Appendiks

## 9.1 …

## 9.2 …