



**Norsk
Regnesentral**
NORWEGIAN COMPUTING CENTER

Notat

Resultatrapport for prosjekt 165763 - Reducing Software Entropy

**Notatnr
Forfatter
Dato**

**DART/09/08
Bjarte M. Østvold
30. januar, 2009**

Norsk Regnesentral

Norsk Regnesentral (NR) er en privat, uavhengig stiftelse som utfører oppdragsforskning for bedrifter og det offentlige i det norske og internasjonale markedet. NR ble etablert i 1952 og har kontorer i Informatikkbygningen ved Universitetet i Oslo. NR er et av Europas største miljøer innen anvendt statistikk. Det jobbes med svært mange forskjellige problemstillinger slik som estimering av torskbestand, finansiell risiko, beskrivelse av geologien i petroleumsreservoarer og overvåking av klimaendringer. NR er også ledende i Norge innen utvalgte deler av informasjons- og kommunikasjonsteknologi. Problemstillinger kan være å overvåke inntrengning i datasystemer, e-læring i skole og næringsliv, bruk av datateknologi i markedsanalyser samt anvendelser av multimedia på forskjellige plattformer. NRs visjon er forskningsresultater som brukes og synes.

Tittel	Resultatrapport for prosjekt 165763 - Reducing Software Entropy
Forfatter	Bjarte M. Østvold
Dato	30. januar, 2009
Publikasjonsnummer	DART/09/08

1 Introduksjon

Dette dokumentet er resultatrapporten for prosjekt 165763 - "Reducing Software Entropy: Exploiting software metadata to improve the quality of agile software" finansiert av Norges forskningsråd som et strategisk instituttprosjekt i perioden 2005-2008. Rapporten er en del av sluttrapporteringen for prosjektet.

2 Målsettinger og bakgrunn

'Overall goals' gjengitt fra prosjektsøknaden:

- G1 Advancing the state-of-the-art of software engineering, focusing on techniques for mapping between the implementation level of software development with the conceptual level of human understanding.
- G2 Advancing the state-of-the-art of source code analysis, focusing on techniques that allow the analysis required to improve system quality.
- G3 Advancing the state-of-the-practice in industrial software development, focusing on technologies to automatically extract information that improves software development as it is actually practised.
- G4 Disseminating and exchanging the knowledge gained.

3 Resultater og FoU-aktiviteter

Dette avsnittet gir en oppsummering av prosjektets viktigste resultater ved prosjektslutt med tilhørende forsknings- og utviklingsaktiviteter. De involvert i forskningen har i all hovedsak vært forskere ved Norsk Regnesentral, samt eksterne veiledere for PhD-studentene eller medforfattere av disse ved utenlandsopphold.

Utenlandsopphold: Einar W. Høst besøkte Barbara G. Ryder ved Rutgers University og Ingvar Tjøstheim besøkte Daniel R. Fesenmaier ved Temple University, begge USA.

3.1 Resultat: Et verktøy for programvareutviklere i mellomstore og store prosjekter

Beskrivelse: Verktøyet, kalt Bark, hjelper utviklere med å holde oversikt over programkoden under systemutvikling og å unngå utilsiktede koblinger mellom deler av koden. Sentralt står visualisering av koden som en grafrepresentasjon. Bark har et avansert brukergrensesnitt basert på systemet Eclipse. Mer informasjon om Bark på dets hjemmeside: <http://bark.nr.no>

FoU-aktiviteter: Vi kartla krav i samarbeid med bedriftspartnere, designet og implementerte en prototyp av verktøyet, prøvde ut verktøyet på partnerprogramkode og på programkode fra velkjente applikasjoner fra åpen kildekode-verdenen.

3.2 Resultat: Metode for å avdekke strukturelle sammehenger i programkode

Beskrivelse: Store programkodebaser er vanskelige å forstå pga. størrelsen og komplekse relasjoner mellom delene. Vi brukte en grafrepresentasjon av koden, hentet fra Bark (se forrige avsnitt), og laget en metode for å avdekke grafmønstre som gjentar seg i koden. Hovedresultatet er at forekomsten av en del mønstre henger sammen med størrelsen av programkoden, noe som igjen indikerer at gjenbruk av kode øker med kodestørrelse.

FoU-aktiviteter: Vi kartla krav til, designet og implementerte en algoritme for finne mønstre i programkode. Denne algoritmen ble så brukt til data-analyse av velkjente applikasjoner fra åpen kildekode-verdenen. Å lage en effektiv algoritme var en særlig utfordring: å analysere store grafer er tidkrevende.

3.3 Resultat: Metode for å forstå navn i programkode

Beskrivelse: Med navn i programkode menes tekstlige betegnelser som programmer setter på biter av programkode, f.eks. "Utskrift-Av-Rapport-Data". Gode

navn, dvs. intuitive og letteste navn, er essensielle for sunn programvareutvikling og godt vedlikehold. Vi har laget en metode som avdekker sær og uforståelig navnebruk, samt gir råd om hvordan bruke de navn som forekommer hyppigst i programkode. Metoden omfatter både enkeltord og fraser satt sammen av flere ord. Vi brukte metoden for å avdekke uheldig navnebruk i velkjente applikasjoner fra åpen kildekode-verdenen.

FoU-aktiviteter: Vi utviklet begreper og algoritmer for å analysere navn på ulike måter. Innsamling og analyse av et korpus av eksisterende programkode for å få data om gjennomsnittlig (normativ) bruk av navn var viktige delaktiviteter. Matematisk modellering av meningen til navn, og implementasjon av analysen, var andre aktiviteter.

PhD-student: Einar W. Høst

3.4 Resultat: Kunnskap om brukeroppførelse i avanserte brukergrensesnitt

Beskrivelse: Avanserte grafiske brukergrensesnitt for virtuelle miljøer kan brukes til å simulere deler av vår materielle verden. Blant resultatene er at oppførelse i de to miljøtypene er korrelert.

FoU-aktiviteter: Vi studerte forskjellen på virtuelle miljøer med to og tre dimensjoner, dvs. 2D- versus 3D-brukergrensesnitt. Vi så også på hvordan oppførelsen til brukerne i et virtuelt miljø kan sammenlignes med oppførelse i et materielt miljø.

PhD-student: Ingvar Tjøstheim

4 Sammeholding av resultater og målsettinger

Målene (se avsnitt 2) og hvordan de adresseres av resultatene i det foregående avsnittet.

- G1 Adresseres av resultatet i avsnitt 3.1 siden verktøyets visualisering av kode et mer høynivå enn koden selv. Adresseres også delvis av aktiviteten i avsnitt 3.4 siden kunnskapen derfra er til hjelp ved utvikling av nye brukergrensesnitt.
- G2 Adresseres av resultatene i avsnitt 3.1 – siden visualisering gir økt forståelse og dermed økt kvalitet – og avsnitt 3.3 – siden gode navn er en forutsetning

for god kvalitet av koden. Begge avsnittenes resultater innebærer kodeanalyse, og navne-analysen er ny mht. litteraturen. Resultatet i avsnitt 3.2 er relevant for programkodekvalitet siden det gir kunnskap om gjenbruk, og dette er også kodeanalyse.

G3 Adresseres av resultatet i avsnitt 3.1 siden verktøyet Bark kan benyttes av utviklere uten krav til en spesiell utviklingsprosess. Adresseres også delvis av resultatene i avsnitt 3.2 og 3.3 siden teorien og implementasjonene her kan omarbeides til generelle verktøy for utviklere.

G4 Disseminating and exchanging the knowledge gained.

Adresseres av prosjektets publiserings- og presentasjonsvirksomhet.

5 Prosjektgjennomføring og ressursbruk

Prosjektgjennomføring:

- Prosjektet har hatt en relativt stabil gruppe ved NR. Gruppen har arbeidet med ulike problemstillinger, men disse utfyller hverandre til et hele, og er i godt samsvar med de opprinnelige målsettingene. Dog har det vært redusert fokus på mål G1 siden Jason Baragry sluttet på NR før prosjektstart; Baragry var tenkt å arbeide i hovedsak med dette målet og veilede en PhD-student innenfor temaet.
- Veiledningen av Høst, som blir utført av en NR-forsker (Østvold), oppleves å ha fungert godt av begge parter.
- Bedriftspartnerne fungerte godt som samtalepartnere vedrørende krav til og nytteverdi av verktøyet Bark. Å overføre slik teknologi til partnerne, dvs. få dem til å bruke den i egen organisasjon, viste seg imidlertid vanskeligere enn først antatt, bl.a. pga. krav til dokumentasjon, lettbrukbarhet og partnernes prioriteringer.

Ressursbruk:

- I sum har prosjektet brukt noe mer ressurser enn budsjettet og dette har blitt dekket av interne midler.
- Den forespeilede gjesteforskerstillingen ble ikke besatt, og midlene ble om-disponert.
- PhD-studentenes stipender løper videre inn i 2009 pga. senere enn forventet oppstart og også enkelte korte planlagte avbrudd.

6 Resultatenes betydning og nytteverdi

Resultatenes relevans i forskningsfeltet vises av deres publisering. En av de to PhD-studentene (Tjøstheim) forblir på NR etter prosjektslutt mens den andre (Høst) begynner i industrien og vil ta med seg kunnskapen til en programutviklerrolle der.

7 Oppfølging av resultater

Resultatene vil videreføres i søknader til EU og Forskningsrådet, og også utnyttes i løpende prosjekter. De sentrale resultatene fra prosjektet er publisert eller vil bli publisert etter prosjektslutt (se neste avsnitt).

8 Resultater etter prosjektslutt

Følgende påbegynte arbeider forventes publisert etter prosjektslutt.

- To forsvarte doktorgrader, én for hver PhD-student.
- Høst, E., Østvold, B. M. (2009). The Vocabulary of Programmers. In preparation for submission to a conference.
- Høst, E., Østvold, B. M. (2008). Debugging Method Names. Submitted to ECOOP 2009: European Conference on Object-Oriented Programming in Genova, Italy.¹
- Tjøstheim, I., Go, F. (2009) Travelers in VE and Telepresence. Virtual Destinations in Computer Games. Submitted paper to The 6th CPTHL Symposium in Vienna: Psychological, Marketing, Socio-Economic, and Sociological Drivers of Cultural Experiences in Leisure and Tourism.

Disse arbeidene kommer i tillegg til allerede publiserte arbeider som er innrapportert i den elektroniske rapporten på Forskningsrådets "Mitt nettsted".

1. The paper was accepted at the conference and shared a best paper prize with another paper.