

Kuinka ohjelmoijat etsivät koodia

Jarmo Isotalo

Referaatti
HELSINGIN YLIOPISTO
Tietojenkäsittelytieteen laitos

Helsinki, 9. syyskuuta 2015

Sisältö

1 Johdanto	1
2 Tutkimusasetelma	1
Lähteet	2

1 Johdanto

Projektien koon kasvaessa ja yhä kehittyneempien haku mahdollisuuksien myötä koodin etsimisestä on muodostunut yhä oleellisempi vaihe sekä ohjelmoitaessa että koodia ylläpidettäessä. Sadowski, Stolee ja Elbaum esittelevä artikkelissaan ”How Developers Search for Code: A Case Study” [1], kuinka ohjelmoijat etsivät koodia. He keskittyvät tapaustutkimuksessaan erityisesti tarkastelemaan miksi, miten ja milloin ohjelmoija etsii koodia käyttäen siihen tarkoitettuja työkaluja.

Käyttäen apuna logi analyysiä sekä juuri ennen koodin hakua tehtyjä kyselyitä he selvittivät vastauksia tutkimuskysymyksiinsä.

2 Tutkimusasetelma

Sadowski, Stolee ja Elbaum kertovat artikkelissaan, että Googlella suurin osa ohjelmakoodista on yhdessä isossa repositoriassa, jossa kaikki pääsevät katsomaan ja halutessaan hyödyntämään toistensa koodia. Googlella on käytössä sisäinen koodin haku työkalu, jossa hakua voi rajata mm. kielen ja sijainnin perusteella. Hakutulosten näyttämisen lisäksi koodihaku mahdollistaa koodin läpi navigoinnin kansiorakenteen sekä viittausten perusteella.

Tutkimuksdata kerättiin ennen koodin hakua aukeavalla kyselyllä sekä logi analyysillä. Kyselyä varten he loivat selain pluginin, joka tarjoaa tutkittaville kyselyn, kun he avasivat koodihaun. He kuitenkin rajoittivat sekä kyselyiden määrän enintään kymmeneen kyselyyn päivässä sekä kyselyiden väliajaksi vähintään 10 minuutin tauon. Tällä he pyrkivät vähentämään kyselyn tutkimuksen tuottamaa lisävaikaa tutkittaville, ja siten saamaan laadukkaampia vastauksia.

Toinen tutkimuksen datalähteistä on hakutyökalun logit, minne kaikki interaktio koodihaun kanssa tallentuu. Logeista he saavat selville anonymisoidun käyttäjän, interaktion kellonajan, hakutermiä ja jokaisen käytetyn selaimen välilehden id:n, sekä mitä tuloksia on klikattu. He korreloivat kyselytulokset sekä logimerkinnät kellonaikojen perusteella. Logeista eritettiin haku sessioita, siten että joka session välissä on vähintään kuuden minuutin tauko. Jokainen sessio koostuu siis koodihaun avaamisesta, hakemisesta, tulosten katselmoinnista ja mahdollisista jatkohauista.

Sadowski, Stolee ja Elbaum esittelevät seuraavat tutkimuskysymykset:

1. Miksi ohjelmoijat etsivät koodia
2. Missä kontekstissa haku suoritettiin
3. Millaiset haku kyselyn ominaisuudet
4. Mitä tyypillinen hakusessio tuottaa tuokseksi

5. Miten eri konteksti vaikuttaa hakutuloksiin

Tapaustutkimuksesta he saivat seuraavat tulokset ja näin he hakivat:

1. Miksi ohjelmoijat etsivät koodia

He kartoittivat syitä koodin etsimiseen lajittelemalla kyselyiden vapaamuotoiset vastaukset omiin ryhmiinsä. Tämä tehtiin siten, että vapaamuotoiset kyselylomakkeen vastaukset kirjoitettiin lapuille ja lajiteltiin omiin kategorioihinsa, kunnes kaikki 3 lajittelijaa (artikkelin kirjoittajat) olivat tyytyväisiä lajitteluun. Näin he saivat aikaan korkean luokan kategorioita. Tämä tuotti tuloksiksi seuraavat lajit: miten tehdä jotakin, mitä koodi tekee, miksi koodi toimii kuten toimii, koodin paikantamiseen ja kuka muutti ja milloin.

Lajittelun tuloksista kävi ilmi, että 33.5% hauista oli vastaamaan kysymykseen, miten jokin asia tulisi tehdä. 26% puolestaan vastasivat kysymykseen mitä, eli hakutapahtuma koostui pääosin koodin lukemistesta ja toteutuksen tarkastelusta sekä yleisen koodityylin selvittämisestä. 16% hauista vastaa kysymykseen missä, eli mistä tiettyä koodia käytetään ja mitä koodia se käyttää sekä hauista missä tavoitteena oli selvittää tietyn koodin sijainti repositoriossa. 16% hauista keskittyy vastaamaan kysymykseen miksi tämä koostuu pääosin selvityksistä, kuten miksi jokin toimii eritavalla kuin miten ohjelmoija ajatteli, sekä mahdollisen muutoksen sivuvaikutuksia. 8.5% hauista keskittyy vastaamaan kysymykseen kuka ja milloin; pääosin selvittääkseen kuka muutti koodia ja milloin, sekä koodin omistajuutta, mikä halutaan selvittää mm. jotta oikea henkilö reviewaisi tulevat muutokset.

2. **Missä kontekstissa haku suoritetaan** Koodia haetaan monista eri syistä Googlella; 39% hauista tehdään silloin kun työskentelee muutoksen parissa. Myös sekä koodin katselmuksen vaiheessa että ongelman ratkaisussa koodia tyypillisesti haetaan. Kuitenkin suurin osa hauista tehdään jo osin tuttuun koodiin.

asadsadasd

Lähteet

- [1] Sadowski, Caitlin, Stolee, Kathryn T. ja Elbaum, Sebastian: *How Developers Search for Code: A Case Study*. Teoksessa *Joint Meeting of the European Software Engineering Conference and the Symposium on the Foundations of Software Engineering (ESEC/FSE)*, 1600 Amphitheatre Parkway, 2015.