

# **Kuinka ohjelmoijat etsivät koodia**

Jarmo Isotalo

Referaatti  
HELSINGIN YLIOPISTO  
Tietojenkäsittelytieteen laitos

Helsinki, 21. syyskuuta 2015

## **Sisältö**

<b>1</b>	<b>Johdanto</b>	<b>1</b>
<b>2</b>	<b>Tutkimusasetelma</b>	<b>1</b>
<b>3</b>	<b>Tulokset</b>	<b>1</b>
<b>4</b>	<b>Yhteenveto</b>	<b>2</b>
	<b>Lähteet</b>	<b>2</b>

## 1 Johdanto

Ohjelmistoprojektien koon kasvaessa ja kehittyneempien koodinhakumahdollisuuksien myötä ohjelmakoodi etsimisestä on muodostunut yhä oleellisempi osa sekä ohjelmistotuotannossa että ohjelmiston ylläpitovaiheessa. Sadowski, Stolee ja Elbaum esittelevä artikkelissaan ”How Developers Search for Code: A Case Study” [1], kuinka ohjelmistokehittäjät etsivät koodia Googlessa. Tutkimuksensa lähteenä he käyttivät hakutyökalun logeja sekä haun yhteydessä pidettyjä kyselyitä.

## 2 Tutkimusasetelma

Kirjoittajat kertovat artikkelissaan, että Googlessa suurin osa ohjelmakoodista on yhdessä isossa avoimessa koodivarastossa. Googlessa on myös käytössä sisäinen koodinhakutyökalu (code search tool), jossa hakua voi rajata mm. ohjelmointikielen ja tiedostopolun perusteella. Hakutulosten näyttämisen lisäksi koodihaku mahdollistaa koodin läpi navigoinnin perusteella.

Voidakseen tarjota kyselyn koodihakutyökalua avattaessa, he loivat selainlisäosan, joka avaa kyselylomakkeen. Selainlisäosa oli konfiguroitu siten, että jokaisen kyselyn välissä on vähintään kymmenen minuutin tauko. Tämän he tekivät, jottei kysely häiritsisi tutkittavia liikaa, ja siten tutkittavat tuottaisivat laadukaampia vastauksia. Hakutyökalun lokeista he saavat selville käyttäjän, tapahtuman kellonajan, hakutermit ja jokaisen käytetyn selaimen välilehden tunnusteen, sekä mitä tuloksia on klikattu. He yhdistivät kyselytulokset sekä lokimerkinnät kellonaikojen perusteella. Lokianalyysissä lokeista pilkottiin hakusessioita, siten että jokaisen hakusession välissä on vähintään kuuden minuutin tauko. Tutkimuksessa tarkennuttiin vastaamaan, miksi ohjelmoijat etsivät koodia, missä kontekstissa haettiin, millainen tyyppinen hakukysely on, mitä haku tuottaa tulokseksi ja miten haun konteksti vaikuttaa tuloksiin.

## 3 Tulokset

Tapaustutkimuksesta saatujen tulosten perusteella kirjoittajille selvisi, että ohjelmoijat käyttävät koodihakua aktiivisesti ja että tyyppillisessä hakusessiossa hakua tarkennetaan toistuvasti.

Vapaamuotoiset kyselytutkimuksen vastaukset lajittelemalla he tunnistivat, että yleisimmät syyt koodin hakemiseen on selvittää, mitä ohjelmakoodi tekee ja miksi se toimii, kuten se toimii. Hakua käytetään myös aktiivisesti selvittämään missä koodi sijaitsee.

Tapaustutkimuksen perusteella suurin osa hauista tehdään ohjelmakoodimuutoksen parissa työskenneltäessä sekä teknistä ongelmaa ratkaistaessa.

Myös koodikatselmuksen yhteydessä koodihakutyökalun käyttö on tyypillistä.

Lokianalyysistä kävi ilmi, että noin neljännes hauista rajoittaa tuloksia tiedostosijainnin perusteella, kun taas ohjelmointikielen perusteella hakuja rajoitettiin vaan noin viidessä prosentissa hauista. He tunnistivat useita tilanteista, jossa tehtiin kaksi tai useampia peräkkäisiä hakuja siten, että hakujen välissä ei ollut muuta interaktiota koodihaun kanssa. He arvelivat tämän johtuvan siitä, että suurin osa peräkkäisistä hauista on joko kyselyn muokkaamista tai hakualueen rajaamista. Näiden peräkkäisten hakujen välissä on keskimäärin vain kahdeksan sekunnin ero. Tässä ajassa koodihaun tuloksia ehdittiin tarkastelemaan ja sen pohjalta tarkentamaan hakukyselyä.

Tyypillinen hakusessio kestää 3 minuuttia ja 30 sekuntia sekä siinä on avattu kaksi selaimen välilehteä. Tyypillisesti hakuja tehtiin 12 päivässä.

Kyselyiden ja lokianalyysin perusteella ohjelmakoodin ominaisuuksia selvitetessä hauku tuotti harvoin yhtään klikkausta. Tämän he arvioivat johtuvan siitä, että vastaus näkyi jo koodin esikatselussa, tai koska haku tuotti vain yhden tuloksen, jonka koodihakutyökalu näytti suoraan. Koodia heikommin tunteva tyypillisesti käyttää enemmän aikaa koodin selaamiseen.

## 4 Yhteenveto

Tulosten pohjalta he ehdottavat on koodinhakutyökalujen luojille, että haku tarjoaisi tulosten esikatselua, haun rajaamisen ohjelmointikielen ja tiedostosijainnin lisäksi huomioimalla käyttäjän konteksti tarkemmin sekä hakutyökalun integraatiota ohjelmointiympäristöön

Sadowski, Stolee ja Elbaum muistuttavat vielä lopuksi, että Googlen uniikki toimintatapa saattaa vaikuttaa tulosten soveltuvuuteen ja toistettavuuteen.

## 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.