

Turun yliopisto  
Informaatioteknologian laitos  
Tietojenkäsittelytiede

## Lausunto Jyri Lehtosen pro gradu –tutkielmasta ”Characterizing the deep Web”

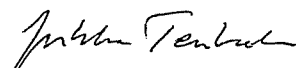
Verkkohaut ja niitä suorittavat hakukoneet ovat äärimmäisen keskeinen toiminto nykyaikaisessa, vahvasti web-informaatioon ja –palveluihin tukeutuvassa tietojenkäsittelyssä. Verkkosivut olivat alunperin staattisia, mutta nykyinen suuntaus on voimakkaasti kohti dynaamisia, pyydettäessä generoitavia sivuja. Tämä muodostaa ongelman hakukoneille, jotka nykyisellään pystyvät indeksoimaan lähinnä vain staattisia sivuja. Dynaaminen, hakukoneilta piilossa oleva web-sisältö on kuitenkin moninverroin laajempi kuin staattinen osa. Lehtosen tutkielma tarkastelee mahdollisuuksia päästä tähän ns. *syväverkkoon* (deep web) käsiksi automaattisilla tekniikoilla. Todettakoon, että kyse ei ole salaisista tiedoista, vaan dynaamisten sivujen pohjana olevista, periaatteessa julkisista tiedoista.

Lehtonen tarkastelee työssään ensin hakukoneiden yleisiä toimintaperiaatteita ja rajoitteita. Sen jälkeen hän luonnehtii syväverkon ominaisuuksia ja arvioi sen volyymiä, sekä erityisesti kasvun nopeutta. Seuraavaksi pureudutaan web-navigointiin yksityiskohtaisemmin, tarkastelemalla tyypillisiä linkitysrakenteita, ja niissä kulkemisen automatisointia. Keskeisimmän osan työstä muodostavat tekniikat, joilla pystytään automaattisesti täyttämään web-lomakkeita, jollaisten takana valtaosa syväverkosta sijaitsee. Tässä yhteydessä esitellään pari olemassaolevaa, lähinnä kokeellista syväverkon hakukonetta. Viimeisessä luvussa Lehtonen esittelee itse laatimansa syväverkkorobotin (deep web crawler) periaatteita ja raportoi sen onnistuneesta toiminnasta.

Jyri Lehtosen tutkielma pureutuu ajankohtaiseen ja kiinnostavaan aiheeseen. Työ on omaperäinen ja osoittaa aiheen perusteellista hallintaa, sekä myös innovatiivisuutta. Tekijän laatima syväverkkorobotin prototyyppi on ansiokas. Varovaisuussyistä sen testauksessa pyrittiin välttämään web-palvelinten ylikuormittamista, joten tavoitteena oli lähinnä teknisen toimivuuden osoittaminen, ei niinkään tulosaaineiston laadullinen analyysi.

Tutkielma on kirjoitettu englanniksi. Kieli on sujuvaa ja vivahteikasta, sekä aiheen käsittely perusteellista. Työn ydinajatuksukset tulevat selkeästi esille. Katsomme, että Jyri Lehtosen tutkielma täyttää erinomaisesti pro gradu –tutkielmalle asetettavat vaatimuksen ja ehdotamme sen hyväksymistä arvolauseella *eximia cum laude approbatur*.

Turussa 22.12.2011

  
Jukka Teuhola

  
Pentti Riikonen

Nähty: 