

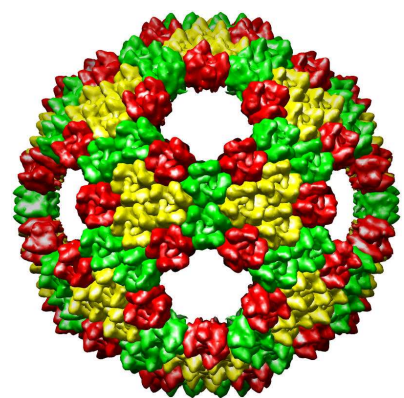
## Uratarina?

Kirjoitin ylioppilaaksi Järvenpään lukiosta keväällä 1997. Koska MAOL-työkirjan käyttöä oli tullut harjoiteltua, tuntuivat Teknillinen korkeakoulu ja Helsingin yliopiston matemaattis-luonnontieteellinen tiedekunta luontevimmilta opiskelupaikoilta. Ajatuksen TKK:sta opiskelupaikkana jouduin tosin hylkäämään kuultuani liian monta kertaa teekkarien ja heidän vanhempiansa käsittelevän Tekniikan ylioppilaiden ylivertaisuutta muuhun ihmiskuntaan verrattuna. Ainoaksi vaihtoehdoksi jäi siis pahamaineinen Helsingin yliopiston matemaattis-luonnontieteellinen tiedekunta.

Sopivimmilta opiskelualoilta (Ludwig Wittgensteinin ohjeen ”Wovon man nicht sprechen kann, darüber muss man schweigen” mukaisesti) vaikuttivat teoreettinen fysiikka, matematiikka ja tietojenkäsittelytiede, vaikka varsinkaan kahdesta viimeisestä minulla ei lukio-opintojen pohjalta ollut vielä juuri minkäänlaista käsitystä. Päädyin lopulta tietojenkäsittelytieteeseen, koska sinne haettaessa ylioppilastutkinnosta hyvitetiin vähiten ja äitini mielenrauhan säilyttämisen vuoksi oli joka tapauksessa syytä mennä pääsykokeisiin. (En tosin ehtinyt vastaamaan kovinkaan moneen pääsykoetehtävään, sillä eväiden nopean nauttimisen vuoksi olin pakotettu palauttamaan vastaukseni hyvin pian kokeen alkamisen jälkeen.)

Lukion jälkeen lähdin kuitenkin täyttämään isänmaallisia velvollisuuksiani Riihimäelle. Hulluuksissani aloin myös tenttimään fysiikkaa ja teoreettista fysiikkaa, sillä en ollut vielä varma oliko tietojenkäsittelytiede sittenkään oikea/hyvä valinta. Tuon vuoden aikana sain integroida ja derivoida vähintäänkin riittävästi varmistuakseni valintani oikeellisuudesta, vaikka teoreettisen fysiikan opiskelu mielestäni onkin toki hyvin yleissivistävää ja hyödyllistä.

Toimittuani kesän ennen opintojen virallista aloittamista kuvankäsittelyohjelmien kehittäjän hommia ilman erityisempää osaamista (ohjelmointikokemukseni oli peräisin lähinnä yläasteen ATK-tunneilla), oli opiskelumotivaatio syksyn alussa korkealla. Ajauduin melko pian myös tutkimusryhmään, joka kehitti laskennallisia menetelmiä biologisten virusten kolmiulotteisten mallien laatimiseksi kryoelektronimikroskooppikuvien pohjalta. Teinkin graduni kyseisestä aiheesta ja valmistuin maisteriksi keväällä 2002. Ohjaajani Esko Ukkonen esitti ennen tutkinnon hakua huolensa siitä, että yliopistosta valmistuvan maisterin olisi osoitettava tiettyä kypsyysää. Päätimme katsoa josko kypsyys saataisiin iskostettua minuun jatko-opintojen aikana. Näin en joutunut myöskään miettimään mitä tekisin isona.



Jatko-opintoni ajautuivat tiedon louhinnan suuntaan. Tiedon louhinnassa kehitetään laskennallisia menetelmiä suurten tietomäärien analysointiin. Viimeisten vuosikymmenten aikana digitaalisessa muodossa tallennetun aineiston määrä on kasvanut räjähdysmäisesti: lukuisat sotilaalliset ja kaupalliset satelliitit kuvaavat maapalloa tarkasti monin eri tavoin, kaupat tallettavat tiedot

ostoksista, Google muistaa kuka on etsinyt mitäkin, verkko-operaattorit ovat velvoitettuja tallentamaan havaitsemaansa verkkoliikennettä, jne.

Väitöskirjani aiheena oli hahmokokoelmien yhteenveto. Tiedon louhinnan tärkeä osa-alue on nimittäin hahmojen etsintä (engl. pattern discovery), jossa tavoitteena on löytää suuresta tietomäärästä kiinnostavia hahmoja. Esimerkiksi usein yhdessä ostettavat tuotteet muodostavat kauppoille mielenkiintoisen hahmon. (Mm. Amazon.com käyttää tällaisia hahmoja suosittelujensa pohjana.) Hahmoja osataan kyllä etsiä nopeasti, mutta usein niitä löytyy niin paljon, ettei niistä tahdo saada mitään tolkkua. Hahmojen etsinnän lisäksi tarvitaan myös hahmojen hyödyntämistä (engl. pattern exploitation; The Exploited näyttelikin merkittävää roolia väitöskirjatyöni etenemisessä), esimerkiksi hahmokokoelmien yhteenvetoa ymmärrettävyyden parantamiseksi.



Tiedon louhinnalla on toki myös pimeä puolensa: Kiinnostus ja kyky kerätä ja analysoida suuria tietomääriä ovat ensi askeleita kohti Orwellilaista valvontayhteiskuntaa. Monissa maissa yritysten edunvalvonnan ja kansallisen turvallisuuden varjolla on otettu merkittäviä edistysaskeleita kyseiseen suuntaan (esim. Yhdysvaltain puolustushallinnon Total Information Awareness Office). Data-analyysimenetelmien kehittyessä myös esimerkiksi tilastojen varomaton julkaiseminen saattaa aiheuttaa vakavia uhkia ihmisten yksityisyydelle: useat tilastot yhdessä voivat paljastaa luottamuksellisia tietoja, vaikka yksittäinen tilasto ei juuri mitään kertoisikaan. Väitöskirjassani tutkin myös tällaisia uhkia.



Väittelin keväällä 2005. Sen jälkeen olin taas vaarassa joutua pohtimaan mitä tekisin isona. Onneksi tarjolla oli vielä yksi viivyttelyn muoto: postdoc-tutkijana toimiminen. (Kypsyys ei tainnut tarttua minuun myöskään jatko-opintojen aikana, joten kenties yliopistolla postdoc-tutkijana jatkaminen oli tohtorin tutkinnon myöntämisen edellytyksenä.) Päätin siis jatkaa tutkimista, koska se oli ollut kivaa, mielenkiintoista ja tajuntaa avartavaa.

En ole vielääkään päättänyt mitä aion tehdä isona. Onneksi, ehkä.

Korsossa 10.4.2006  
Taneli Mielikäinen  
supersankari