

hyväksymispäivä

arvosana

arvostelija

Koneoppimisen sovellukset merkintunnistuksessa

Tuomo Salmenkivi

Helsinki 26.2.2018

HELSINGIN YLIOPISTO

Tietojenkäsittelytieteen osasto

Tiedekunta — Fakultet — Faculty		Laitos — Institution — Department	
Matemaattis-luonnontieteellinen tiedekunta		Tietojenkäsittelytieteen osasto	
Tekijä — Författare — Author			
Tuomo Salmenkivi			
Työn nimi — Arbetets titel — Title			
Koneoppimisen sovellukset merkintunnistuksessa			
Oppiaine — Läroämne — Subject			
Tietojenkäsittelytiede			
Työn laji — Arbetets art — Level		Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
		26.2.2018	9 sivua + 1 liitesivua
Tiivistelmä — Referat — Abstract			
Todo abstract			
ACM Computing Classification System (CCS): General and reference → Document types → Surveys and overviews Applied computing → Document management and text processing → Document management → Text editing			
Avainsanat — Nyckelord — Keywords			
ulkoasu, tiivistelmä, lähdeluettelo			
Säilytyspaikka — Förvaringsställe — Where deposited			
Muita tietoja — Övriga uppgifter — Additional information			

Sisältö

1	Johdanto	1
2	Kirjoituksen rakenne	1
2.1	Tiivistelmä	1
2.2	Johdanto	2
2.3	Käsittelyluvut	2
2.4	Lähdeviittausten käyttö	3
2.5	Yhteenveto	3
2.6	Lähdeluettelon laatiminen	4
3	Ulkoasulliset seikat	6
3.1	Työn osien järjestys	6
3.2	Tekstin yleinen sijoittelu	7
3.3	Kuvat ja taulukot	7
3.4	Otsikot	9
3.5	Mallin käyttö	9
4	Yhteenveto	9

Liitteet

1 Malli ABC

1 Johdanto

Optista merkintunnistusta (OCR) käytetään muuttamaan skannattujen ja tyypillisesti konekirjoitettujen dokumenttien sisältö tietokoneen luettavaan ASCII merkitöön. Tämä mahdollistaa tekstin muokkaamisen, lukemisen ja säilömistä huomattavasti tehokkaammalla tavalla.

Hyvin suuri osa arkipäiväisesti informaatiosta ei kuitenkaan ilmene dokumenteissa vaan esimerkiksi liikkeiden julkisivuissa tai tiekulteissa. OCR ei toimi tällaisissa luonnollisissa tilanteissa yhtä hyvin, sillä tyypillisesti OCR-ratkaisut vaativat toimiakseen mustavalkoisen, tasaisen, linjoihin perustuvan ympäristön.

Keskityn tässä tutkielmassa käsittelemään aluksi optisen tekstintunnistuksen käyttöön ja ongelmiin yleisellä tasolla. Seuraavissa kappaleissa käsittelen optisen merkintunnistuksen sovelluksia koneoppimisessa.

2 Kirjoituksen rakenne

Tarkastellaan aluksi tieteelliseltä tekstiltä odotettuja kirjoituksen osia. Samoihin asioihin on luonnollisesti syytä kiinnittää huomiota myös muussa teknisessä kirjoittamisessa. Huomattakoon, että tämä teksti ei ole tieteellinen teksti, eikä siten itse sisällä kaikkia niitä elementtejä, jotka tieteellisen tekstin sisällölliseen antiin kuuluvat. Tällaisia puutteita ovat esimerkiksi johdannon tutkimuskysymyksen asettelun puuttuminen sekä arvoivan materiaalin puute tekstin lopussa, sekä yhteenvedon laiteus. Teksti rajoittuu siten otsikkonsa mukaisesti vain tekniseen sisällön asetteluun.

2.1 Tiivistelmä

Tiivistelmäsiivu sisältää seuraavat osat: työn bibliografiset tiedot, tiivistelmäteksti, aihehuokat ja avainsanat. Bibliografiset tiedot koostuvat työn otsikosta, tekijän nimestä, julkaisupaikan tiedoista, julkaisuaikajankohdasta ja sivumäärästä.

Tiivistelmäteksti on lyhyt, yleensä yhden kappaleen mittainen (maksimissaan noin 100 sanaa) selvitys kirjoituksen tärkeimmästä sisällöstä: mitä on tutkittu, miten on tutkittu ja mitä tuloksia on saatu.

Aiheluokat kuvataan ACM Computing Classification System -luokituksen (CCS) luokituksen mukaisesti. Luokittelussa käytetään täysia polkuja juurisolmun CCS osoittamista lähtöposteista lehtisolmuihin. Polkuja voi antaa 1-3 aihepiirien soveltuvuuden mukaan, mitä alempi opinnäyte, sen vähemmän polkuja se tarvitsee. Poluissa tasot erotetaan toisistaan nuolella eteenpäin. Kun polun nimisanoja arvioidaan suhteessa työn sisältöön, merkitään boldface-fontilla tärkein termi, italics-fontilla toiseksi tärkein. Näin menetellään, mikäli jotkin termeistä ovat olennaisesti paremmin kuvaavia kuin muut polun termit. Nimettyjen polkujen lisäksi lukija voi siten tarkastella lisäulottuvuutena myös tärkeiksi merkittyjen termien joukkoa sinänsä.

Avainsanoiksi valitaan kirjoituksen sisältöä keskeisesti kuvaavia käsitteitä.

2.2 Johdanto

Johdannon tarkoituksena on kertoa yleiskielisesti työn tavoite. Kerrotaan (kuten tiivistelmässäkin, mutta laveammin), mitä on tutkittu, miten on tutkittu ja mitä tuloksia on saatu. Jotta kysymyksenasettelu ja tulokset on lukijan helppo oikein tulkita on syytä aloittaa johdanto asettelemalla tutkimus asiayhteyteensä, esimerkiksi kertomalla aluksi, minkälaisessa yhteydessä tarkasteluun otettavat haasteet esiintyvät ja keiden on ratkaisusta tarkoitus hyötyä.

Johdannon pituus määräytyy suhteessa koko kirjoitelman pituuteen. Parisivuinen kirjoitus ei erikseen otsikoitua johdantoa kaipaa, sillä se itsessään on laajennettu tiivistelmä. Kymmensivuisen kirjoituksen johdanto voi olla vaikkapa sivun tai puolentoista mittainen. Pro gradu -tutkielman 50-70-sivuisen kokonaisuuteen tuntuu 2-4-sivuinen johdanto kohtuulliselta.

Johdanto kertoo siis lyhyessä, yleistajuisessa muodossa koko kirjoitelman kysymyksenasettelun, juonen sekä tulokset ja johtopäätelmät. Tämän luettuaan lukija voi päätellä, haluaako syventyä asiaan tarkemmin lukemalla koko kirjoituksen.

2.3 Käsittelyluvut

Käsittelylukujen työnjako määräytyy käsiteltävän asian luonteen mukaisesti. Lukijan ohjailemiseksi kukin pääluku kannattaa aloittaa lyhyellä kappaleella, joka paljastaa mikä kyseisen luvun keskeisin sisältö on ja kuinka aliluvuissa asiaa kehitellään eteenpäin. Erityisesti kannattaa kiinnittää huomiota siihen, että lukijalle ilmaistaan selkeästi miksi kutakin asiaa käsitellään ja miten käsiteltävät asiat suhtautuvat toisiinsa.

Jäsentelyongelmista kielivät tilanteet, joissa alilukuja on vain yksi, tai joissa käytetään useampaa kuin kahta tasoa (pääluku ja sen aliluvut). Kolmitasoisia otsikointeja saatetaan tarvita joissakin teknisissä dokumenteissa perustellusti, mutta nämä muodostavat poikkeuksen.

Perusohjeena on käyttää tekstin rakenteellisesti painokkaita paikkoja, kuten lukujen avauksia ja teksikappaleiden aloitusvirkkeitä juonenkuljetukseen ja informaatioaskeleiden sitomiseen toisiinsa. Tekstikappaleiden keskiosat, samoin kuin lukujen keskiosat selostavat asiaa vähemmän tuntevalle yksityiskohtia, kun taas aihepiirissä jo sisällä olevat lukijat voivat alkuvirkkeitä silmäilemällä edetä tekstissä tehokkaasti eksymättä tarinan juonesta.

Kullakin kirjoittajalla on oma temponsa, joka välittyy lukijalle tekstikappaleiden pituudessa ja niihin sisällytettyjen ajatuskulkujen mutkikkuudessa. Kussakin tekstikappaleessa pitäisi pitäytyä vain yhdessä informaatioaskeleessa tai olennaisessa päätelyaskeleessa, muuten juonen seuraaminen käy raskaaksi olennaisten lauseiden etsiskelyksi. Yksivirkkeisiä tekstikappaleita on syytä varoa.

2.4 Lähdeviittausten käyttö

Olenaisia opittavia asioita viittaustekniikoissa ovat viitteen paikka tekstissä, oikea lähdeluettelojärjestys valitun viitetyylin parina sekä taito ja tahto noudattaa annettua tyylimääräystä. Väitöskirjoissa ja lehti- tai konferenssiartikkeleissa tekstin hyväksyminen riippuu myös näiden yksityiskohtien asianmukaisesta käsittelystä. Tästä syystä laitoksella nähdään tarpeelliseksi opiskelijoiden tutustua edes pinnallisesti myös muihin tyyllilajeihin ja oppia käyttämään automatisoituja muotoilutyökaluja tehokkaasti, jolloin tyylinuutokset ovat tehokkaita.

Lähdeviitteet sijoitetaan aina virkkeen sisäpuolelle. Siten esimerkiksi tekstikappaleen lopussa irrallaan oleva viite ei ole asiallinen. Tilanne ei muutu, vaikka viite sujautettaisiin tekstikappaleen viimeisen virkkeen sisään. Lähdeviittauksen yhteyteen merkitään mukaan tarkentavat sivunumerot, mikäli lukijan olisi työlästä löytää asianomainen kohta viitatusta lähteestä.

Tehokkaita viitteensijoittelupaikkoja ovat esimerkiksi uuden käsitteen nimeämiskohdasta ja virkkeen loppu kun kyseessä on lähteestä lainattu väite. On myös muistettava lainausmerkkien käyttö silloin kun tehdään suoria lainauksia.

Tekstin jäsentelyn on tuotava selkeästi esiin, mihin asiaan viite liittyy. Samalla tulee ymmärrettäväksi se, kuinka pitkään tekstikatkelmaan ko. viite liitetään. Ei ole siten asiallista aloittaa lukua nimeämällä yhtä tai useampaa lähdettä luvun taustaksi, vaan viitteitä on kiinnitettävä täsmällisemmin väitteisiin ja käsitteisiin. Luvun avaus viitetiedolla voi olla oire myös suuremmasta ongelmasta: lähderiippuvuudesta. Aloitteleva kirjoittaja helposti toistaa lähteestä oppimaansa ilman että tarpeellinen analysointi ja prosessointi suhteessa muuhun opittuun olisi vielä tapahtunut.

Viitteillä ja sanamuodolla on myös tuotava selkeästi esiin se, mikä teksissä on lainattua ja mikä oman pohdinnan ja valikoinnin tulosta.

Lähdeviittauksiin käytetään Tietojenkäsittelytieteen osastolla numeroitua tyyliä ja APA-tyyliä, valinnan näiden välillä tekevät kunkin ryhmän valvoja ja ohjaaja yhdessä. Numeroitu tyyli on esimerkiksi IEEE- ja ACM-julkaisuissa yleisesti käytetty ja puolustaa siten paikkaansa. APA-tyyli on poikkeuksellinen ns. kovissa tieteissä, mutta monet valvojista pitävät siitä sen luettavuuden vuoksi. Numeroita joutuu nimmiä useammin tarkistamaan lähdeluettelosta, sillä tarkastus- ja arvointiprosessiin kuuluu arvioida myös lähteiden valitua ja niiden käyttötapaa.

2.5 Yhteenvedo

Yhteenvedo vaatimattomimmillaan on vain lyhyt kertaus kirjoituksen keskeisistä asioista. Arvokkaamman yhteenvedon saa aikaan kommentoimalla työn tulosten arvoa, työn liittymistä ympäristöön ja tulevaisuudennäkymiä. Tällaiset arviot huolellisesti perusteltava.

2.6 Lähdeluettelon laatiminen

Tieteellisen kirjoittamisen kurssin töiden lähdeluetteloiden laatimisessa noudatetaan seuraavia ohjeita.

Niiden taustalla on kaksi keskeistä pyrkimystä: tehdä viitatus lähteen hankkiminen luettavaksi mahdollisimman helpoksi ja ilmaista, millaisen arviointiprosessin läpi käyneeseen kirjoitukseen vedotaan. Näistä syistä

- lähdeviitteen tulee aina olla niin tarkka, että lähde on sen perusteella tunnistettavissa ja löydettävissä luetteloista ja kirjastoista,
- erityyppisten lähteiden (kirjat, konferenssit, lehdet) on erotuttava toisistaan ja
- luettelon eri osien tulee olla mahdollisimman yhdenmukaisia, erityisesti lähde-tyypin sisällä.

Riippumatta käytettävästä viitetyylistä, lähteet ovat Tietojenkäsittelytieteen osaston opinnäytteiden lähdeluetteloissa tekijän nimen mukaisessa aakkosjärjestyksessä, saman tekijän (tekijäryhmän) työt julkaisuaajan mukaisessa järjestyksessä. Jos jollakin lähteellä ei ole henkilötekijää, se aakkostetaan julkaisun nimen mukaisesti.

Kustakin lähteestä annetaan seuraavat tiedot, edelleen viitetyylistä riippumatta:

- (tarvittaessa lähdeviitelyhenne).
- tekijän tai tekijöiden nimet (sukunimi, etunimien alkukirjaimet) alkuperäisessä järjestyksessään; jos tekijöitä on enemmän kuin kolme, voidaan toimia siten, että vain ensimmäinen tekijä nimetään ja muiden tilalle kirjoitetaan *et al.*
- julkaisun tai artikkelin nimi alkuperäisessä muodossaan
- julkaisupaikan tiedot:
 - kirjasta: kustantaja, julkaisupaikka (voidaan jättää pois, jos kyseessä on tunnettu kustantaja), vuosi ja
 - lehtiartikkelista: lehden nimi, volyymi, numero, vuosiluku ja kuukausi (suluissa) ,
 - artikkelikokoelmassa (esim. konferenssijulkaisussa) ilmestyneestä artikkelista:
 - * kokoelman nimi, toimittaja, kustantaja, julkaisupaikka ja vuosi *tai*
 - * konferenssin nimi, järjestäjä, paikka ja aika,
 - raportista: julkaisusarja, raportin numero, julkaisupaikka, julkaisija ja vuosi

- www-lähteestä: verkko-osoite, voimassaoloajankohta, mahdollisesti viit-
tausajankohta hakasuluissa
- sivunumerot, mikäli lähteenä käytetty julkaisu on artikkeli tai kokoomateoksen
itsenäinen luku.

Normaaliin suomalaiseen tapaan artikkelin nimessä ainoastaan ensimmäinen sana kirjoitetaan isolla alkukirjaimella, sen sijaan konferenssien ja kokoelmajulkaisujen nimissä käytetään isoa alkukirjainta jokaisen sanan alussa (artikkelisanoja ja prepositioita lukuunottamatta). Katso mallia oheisista esimerkeistä. Kokoelman nimen edessä on syytä selvyyden vuoksi käyttää sanaa *Teoksessa*, paitsi kun on kysymys konferenssijulkaisusta, jonka nimi alkaa lyhenteellä *Proc.* (sanasta Proceedings). Tällöin ei tarvita mitään täydennystä. Tämän eron näkee esimerkiksi vertaamalla lähdeviitteiden "[?]" ja "[?]" ulkoasuja.

WWW-lähteiden käytössä on syytä muistaa, että verkossa julkaisukynnys on ole-
maton. Kannattaa siten keskittyä tunnettujen tieteellisten kustantajien julkaisuihin
ja niihin teknisiin standardeihin, joille WWW on ainoa julkaisukanava. Mikäli sama
julkaisu on saatavissa myös perinteisessä muodossa, viitataan ensisijaisesti siihen ja
käytetään verkko-osoitetta lisätietona. Lähdeluettelossa on annettu esimerkit useita
kanavia julkaistusta kirjoituksesta [?, ?] sekä pelkästään WWW-julkaisuna leviäväs-
tä standardista [?].

Erityisesti varoitetaan Wikipedian käytöstä tieteellisessä tekstissä. Vaikka sen avulla
on helppo alustavasti tutustua joihin aihepiireihin ja asiantutva lukija voisi teksin
kelvolliseksi tiettyä hetkenä hyväksyäkin, ei se foorumina millään lailla täytä tie-
teellisesti vertaisarvoidun tutkimusfoorumin kriteerejä. Jos Wikipedia-artikkeliä ei
mitenkään malta ajankuvana olla mainitsematta, käytettäköön jotain muuta kuin
lähdeviitetekniikkaa tähän taiteelliseen otteeseen, vaikkapa alaviitteitä. Olennaista
silloinkin on, että tieteellinen sisältö ei tule tällä korvatuksi vaan sen puute koros-
tetuksi.

WWW-lähteeseen viittaamisessa pätevät samat periaatteet kuin perinteisiin läh-
teisiin viitattaessa: lähdeviitteessä ilmaistaan otsakkeet, kirjoittajat, toimittajat ja
muut seikat. Eroa on ainoastaan verkko-osoitteen ja sen voimassaoloajankohdan il-
maisemisessa. Mikäli lähde on julkaistu ainoastaan verkossa, voidaan web-osoitetta
(URL) käyttää vastaavasti kuin perinteisen julkaisun paikannusinformaatiota (leh-
den ja se numeron julkaisutiedot). Lähdeluettelossa on WWW-viittausten yhteydes-
sä aina syytä ilmaista päivämäärä, jolloin linkin voimassaolo ja lähteen sisältö on
tarkastettu. Esimerkkeinä verkkoviitteistä soveltuvat seuraavat:

- Abiteboul, S., Quass, D., McHugh, J., Widom, J., Wiener, J.L., The Lorel que-
ry language for semistructured data. International Journal on Digital Libraries
1 (1997), 68-88. Myös <http://link.springer.de/link/service/journals/00799/bibs/7001001/70010068.htm> (18.1.2000).
- Dietinger, T. et al, Dynamic Background Libraries - New Developments in
Distance Education Using HIKS (Hierarchical Interactive Knowledge System).

Journal of Universal Computer Science, 5, 1 (1999), http://www.iicm.edu/jucs_5_1/dynamic_background_libraries_new (18.1.2000)

Kirjoituksen lähdeluettelossa luetellaan täsmälleen ne lähteet, joihin viitataan kirjoituksen tekstiosassa. Tämän kirjoituksen lähdeluettelo on tarkoitettu lähinnä esitustavan esimerkiksi, mistä syystä siinä on ”ylimääräisiä” lähteitä.

Pääsääntöisesti julkaisun tai artikkelin nimen perään tulee piste, samoin kunkin lähteen bibliografisten tietojen perään. Muut tiedot erotetaan toisistaan pilkulla. Useimmissa tapauksissa voidaan noudattaa teknisten välineiden antamaa mallia, sillä edellytyksellä, että ylläolevat vaatimukset muuten täyttyvät.

3 Ulkoasulliset seikat

Tässä luvussa käsitellään yleisimpiä tekstin tekniseen esittämiseen liittyviä seikkoja.

3.1 Työn osien järjestys

Kirjoituksen alussa on aina erillinen, mallin mukainen kansilehti. Toisena sivuna on tiivistelmäsiivu, sen jälkeen sisällysluettelo (yksi tai useampia sivuja) ja sitten varsinainen teksti. Sivunumerointi aloitetaan vasta ensimmäiseltä tekstisivulta (arabialaisella ykkösellä). (Tarkat jättävät ykkössivun numeromerkittä.) Sisällysluetteloon merkitään kaikki (numeroidut) otsikot ja vastaavat sivunumerot. Monet tekstinkäsittelyjärjestelmät muodostavat itse sisällysluettelon, jolloin kirjoittajan ei tarvitse huolehtia luettelon sivunumeroiden päivittämisestä tekstin kehittyessä. Sisällysluettelosiivu ja sitä edeltävät sivut voidaan haluttaessa numeroida erikseen (roomalaisin numeroin) esimerkiksi tämän mallin mukaisesti.

Varsinaisen tekstin jäljessä, mutta itse työhön kuuluvana on ensimmäisenä lähdeluettelo, jonka otsikkoa ei numeroida. Lähdeluettelon jälkeen sijoitetaan mahdolliset liitteet, jotka otsikoidaan ja varustetaan sisäisillä sivunumeroilla.

Mikäli kuvista, algoritmeista ja taulukoista halutaan tehdä yhtenäinen luettelo, sijoitetaan luettelot sisällysluettelon jälkeen. Luetteloiden käyttöarvosta on eriäviä mielipiteitä, joten niiden laatimiseen ei varsinkaan ilman tekstinkäsittelyjärjestelmän tukea kannata ryhtyä ilman tarkastajan erityistä toivetta.

Mikäli kirjoitukseen erityisistä halutaan liittää aakkosellinen hakemisto, sijoitetaan se lähdeluettelon jälkeen ennen liitteitä. Indeksimerkitään sisällysluetteloon samoin kuin lähdeluettelo (numeroimaton luku). Mikäli indeksin tekemiseen ryhdytään, on syytä käyttää tekstinkäsittelyjärjestelmän tarjoamaa automatiikkaa.

Teksin luonnollisen juonenkuljetuksen mukana esiin tulevien käsitteiden määrittelyjen sijasta ei pidä yrittää sen enempää pakata kaikkia määritelmiä johdantoon kuin laatia johdantoa ennen käsitelistää tai lyhenteiden selityslistaa. Kumpikaan ei sovi tavanomaiseen argumentoivaan tieteelliseen tekstityyliin, vaikka teknisessä yhteydessä niillä liitteinä voi olla lisäarvoa.

3.2 Tekstin yleinen sijoittelu

Lopullinen tutkielmaversio voi olla yksi- tai kaksipuoleiseksi aseteltua ja riviväliltään yksi. Erityyppisissä teksteissä haasteet ja asetteluvaatimukset voivat olla erilaiset. Erotta kappaleet toisistaan yhdellä ja luvut muutamalla tyhjällä rivillä tai vaihtoehtoisesti käytä tekstinkäsittelytyökalujen ominaisuuksia hyödyksesi ja määrittele tekstikappaleiden väliin jäävä tila hieman normaalia riviväliä suuremmaksi.

Luonnosvaiheessa voi olla avuksi käyttää yksipuolista tulostusta ja suurempaa riviväliä, jotta palautteen antaja voi kirjoittaa kommentteja tekstirivien väliin. Luonnoksissa myös molempien marginaalien on hyvä olla melko leveät (noin 3 cm): vasen (hieman leveämpi) marginaali tarvitaan kirjoitelman nitomista varten ja oikealle puolelle tarkastaja voi kirjoittaa merkintöjään. Jätä myös ylä- ja alareunaan riittävästi tilaa (2–3 cm).

Kirjoituksen lukujen, kuvien ja taulukoiden erottumisen kannalta tärkein keino on riittävän tilan käyttö niiden ympärillä. Kuvan ja nimekkeen tulee olla selkeästi yksi kokonaisuus, joka eroaa muusta tyhjän tilan rajaamana. Kuvan tai taulukon on aina numerointinsa ja nimekkeensä kanssa mahdollista yhdelle sivulle tai varmasti kaksipuolisena paperidokumenttina tarkasteltavassa tekstissä aukeamalle. Kuvissa fonttikoko ei saa alittaa 8 pistettä.

Jos uusi luku tulisi alkamaan aivan sivun alareunasta (vain yksi tai kaksi riviä varsinaista tekstiä), aloita mieluummin uusi sivu. Jokaista uutta lukua ei kuitenkaan ole tarpeen — etenkin lyhyessä kirjoituksessa — aloittaa uudelta sivulta: jos kirjoituksessa on paljon melkein tyhjiä sivuja, lukija voi epäillä, että kirjoittaja on yrittänyt saada kirjoituksensa näyttämään pitemmältä kuin se onkaan.

Tyhjää tilaa kannattaa käyttää hyödyksi myös kuvien ja taulukoiden yhteydessä. Erityisesti jos kirjoituksessa käytetään kauttaaltaan samaa tekstityyppiä, tyhjät rivit ovat välttämättömiä erottamaan esimerkiksi tekstiä ja taulukkoa toisistaan. Tyhjä tila on halpaa, mutta se lisää selkeyttä ja luettavuutta.

3.3 Kuvat ja taulukot

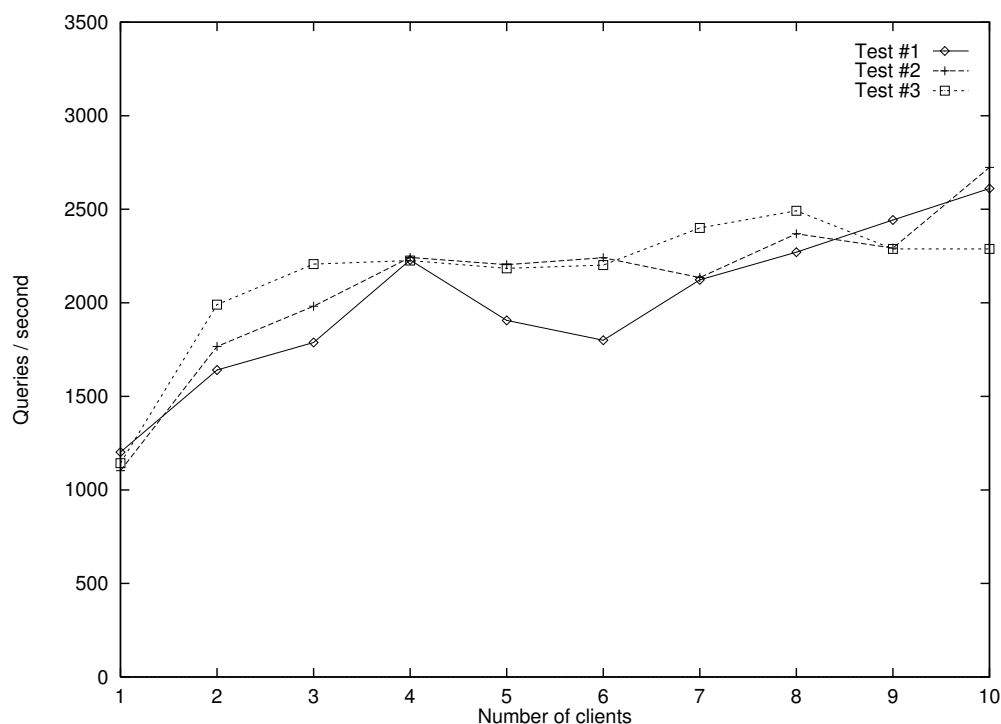
Kuva tai taulukko sijoitetaan mahdollisimman lähelle (ensimmäistä) tekstikohtaa, jossa siihen viitataan, ei kuitenkaan kyseistä viittausta aikaisemmaksi. Tekstissä on syytä myös kertoa, mitä kuvalla halutaan havainnollistaa. Kuvan voi lukea monella eri tavalla, joten lukijaa on ohjattava.

Kuvaa ei pidä sijoittaa välittömästi luvun otsikon alle, vaan on aloitettava tekstillä. Kuvaa ei pidä sijoittaa keskelle tekstikappaletta (saati virkettä), paitsi jos kuva tulee sivun alkuun tai loppuun eikä kappaleen jatkumisesta tule epäselvyyttä.

Kuvan ei aina tarvitse olla välittömästi viittaavan kappaleen perässä. Esimerkiksi viittauskohdan ja vasta seuraavalle sivulle mahtuvan kuvan väliin jäävää sivun loppuosaa ei jätetä tyhjäksi. Kuvaa ei kuitenkaan pidä viedä seuraavaa sivua kauemmas viittauskohdasta.

Varsinaista kuvan esittämistä havainnollistaa kuva 1. Huomiota on kiinnitettävä

kuvan osien ja tekstimerkintöjen näkyvyyteen, kuvan numerointiin ja otsikointiin.



Kuva 1: Kuvan elementit.

Kuvien kokoon on kiinnitettävä huomiota. Käytettyjen merkintöjen on oltava helposti luettavissa ja selkeät. Esimerkiksi suorituskäyriä esitettäessä akselit on nimettävä, asteikot merkittävä ja käytetyt yksiköt tuotava selkeästi esiin. Samankaltaisia asioita esitettäessä useammalla kuvalla on syytä käyttää samaa mittakaavaa vertailun helpottamiseksi.

Kuvan otsikko kirjoitetaan kuvan alle ja sen tulee olla mieluummin lyhyt ja ytimekäs kuin liian selittelevä. Samoin toimitaan taulukoiden otsikoinnissa.

Kuvat ja taulukot numeroidaan juoksevasti. Pitkissä teksteissä käytetään kaksitasoista numerointia (esimerkiksi Kuva 3.1) pääluvuittain, lyhyissä riittää yksitasoinen numerointi.

Kuva- ja taulukko-otsikoiden yhdenmukaiseen esitystyyliin on syytä kiinnittää huomiota, samoin mm. välimerkkeihin. Luontevaa on käyttää kuvatekstin lopussa pistettä, ovathan useimmat kuvateksteistä virkkeitä.

(Kuvien ja taulukoiden otsikointityyli vaihtelee kustantajittain ja julkaisuittain. Samoin tuntuu suositeltava käytäntö Tietojenkäsittelytieteen laitoksen sisällä vaihtelevan taulukon otsikon sijainnin suhteen.)

3.4 Otsikot

Otsikoissa voi käyttää muusta tekstistä poikkeavaa kirjasintyyppiä, alleviivausta, suurempaa kirjasinkokoa tms. erotuskeinoa, yleensä kuitenkin vain yhtä näistä, koska kovin monta erilaista kirjasintyyppiä ja -kokoa tekee ulkoasusta helposti sekavan. Otsikoiden esitystavan on oltava johdonmukainen läpi koko kirjoituksen. Numeroimattomia ”ylimääräisiä” otsikoita ei tule yleensä käyttää.

3.5 Mallin käyttö

Voit käyttää tätä kirjoitusta mallina oman opinnäytteesi ulkoasua varten. Eri tekstinkäsittelyjärjestelmissä käytössä olevat yksityiskohdat kuten kirjasintyypit ja -koot ja rivivälit poikkeavat toisistaan, joten pienet poikkeamat ovat toki hyväksyttäviä.

Tieteellisen kirjoittamisen kurssin luennoilla ja liitteenä olevassa ohjeessa annetut töiden ohjeelliset sivumäärät koskevat työtä, joka vastaa ulkoasultaan tätä ohjetta (kirjasinkoko 12 pistettä). Tässä tekstissä keskimääräinen rivin pituus lienee noin 80 merkkiä ja sivun pituus 35 riviä. Sivumääriin lasketaan varsinaisen tekstiosuuden pituus ja lähdeluettelo (arabialaisin numeroin numeroitu osuus), ei kansilehteä, tiivistelmää eikä sisällysluettelo.

4 Yhteenveto

Tämän kirjoituksen tarkoituksena on toimia muistilistana eräistä esitystavallisista säännöistä, joihin harjoitusten ja tutkielmien kohdalla on syytä kiinnittää huomiota.

Annetut ohjeet on laitoksen henkilökunta muotoillut yhdessä keskustellen ja noudattaen oman tieteenalansa perinteitä. Eri erikoistumisaloilla ja erilaisilla määräävässä asemassa olevissa julkaisufooruilla käytänteet vaihtelevat ja nuorten tutkijoiden onkin tiedostettava ero yleisten sisältöohjeiden ja teknisten muotoilusääntöjen välillä. Aina tekstin valmistuessa on tarkastettava erikseen, täyttääkö se annetut pituusrajoitteet ja vastaako se annettuja muotoiluohjeita, olivatpa ne kuinka pikkutarkkoja tahansa. Tarkasta sääntöjen noudattamisesta syntyy yhtenäisyyttä kokoovan julkaisun tasolla, mikä helpottaa lukijoiden työskentelyä.

Tämä ohje vastaa vain asettelullisiin kysymyksiin ja sen rinnalla on syytä tutustua materiaaliin ja luentoihin, joissa keskitytään tekstin varsinaiseen sisältöön. Olennaisin väline on kuitenkin akateemisesti pidemmälle ehtineen, jo julkaisuja rakentaneen ohjaajan palaute ja mentorointi.

Liite 1. Malli ABC

Liitteet ovat tässä vain sisällysluettelon ja esitystavan mallina. Jokainen liite aloitetaan yleensä uudelta sivulta, jonka alkuun tulee liitteen numero ja nimi. Kunkin liitteen sivut numeroidaan erikseen.

Liite on paitsi dokumenttia täydentävä osuus myös itsenäinen kokonaisuus. Liite ei siten voi olla pelkästään kuva tai ohjelmanpätkä, vaan liitteessä on ilmaistava sen sisällön laatu ja tarkoitus.