

*Harri Oinas-Kukkonen  
Vesa Indrén*

# Tekstin muuntaminen hypertekstiksi

Oinas-Kukkonen, Harri & Indrén Vesa, Tekstin muuntaminen hypertekstiksi [Converting Text to Hypertext]. Kirjastotiede ja informatiikka 14 (2): 35–47.

The hypertext approach can provide electronic documents and a user-friendly and flexible way of organizing and accessing them. Via its linking capabilities it supports representation of the associative connections that exist between specific regions of the documents, associations that may exist for either the reader or the author. The question of how to convert traditional linear documents to hyperdocuments is discussed here, the aim being to give the reader an overview of the key conversion issues. The conversion process is discussed in a step-by-step manner, in which the steps recognized are preparing of text, preparing of nodes, indexing, creation of links, organizing of the material and refinement of the links. Some of the existing conversion process models are introduced and discussed based on one of the process models. Both manual and semi-automatic conversion and tools developed for supporting the conversion process models are discussed. Some experiences with an example conversion from a Word document to a HyperCard stack with the Expanded Book toolkit carried out on a Macintosh platform are described.

*Address: University of Oulu, Department of Information Processing Science, P.O.B. 400, 90571 Oulu, Finland. E-mail: hok@rieska.oulu.fi, WWW: http://rieska.oulu.fi/~hok/.*

## 1. Johdanto

Maailmassa julkaistujen dokumenttien määrän alati kasvaessa on lisääntynyt myös tarve käsitellä ja säilyttää dokumentteja elektronisessa muodossa. Huomattava osa tuotetuista dokumenteista kirjoitetaan nykyään tietokoneiden avulla, vaikka lukeminen tapahtuu silti useimmiten paperilta. Tietyissä tilanteissa kaivataan kuitenkin nykyistä parempaa tukea elektronisten dokumenttien selailuun ja lukemiseen. *Hyperteksti* tarjoaa käyttäjystävällisen ja joustavan tavan tarkastella elektronisia dokumentteja asioiden välillä olevien yhteyksien avulla.

Hypertekstiä voidaan kuvata moniulotteisena elektronisena *dokumenttina*, joka koostuu useista suunnittelijan tai suunnittelijoiden assosiaatioiden perusteella järjestetyistä (*linkit*) pienistä tiedon-

palasista (*solmut*). Oleellista on pyrkimys tiedonpalasten joustavaan yhdistelyyn ja käyttöön. (Oinas-Kukkonen 1993, 8)

Osa jo olemassa olevista dokumenteista saataan haluta tai kannattaa siirtää luettavaksi hyperdokumenttina joko työasemalle tai verkkoon. *Konversiolla* tarkoitetaan tässä yhteydessä prosessia, jossa paperi- tai elektronisessa muodossa olevia dokumentteja muunnetaan hyperdokumenteiksi (Berk & Devlin 1991). Laajimmin määriteltynä konversio voi tarkoittaa myös hyperdokumentin muuntamista käytettäväksi myös muissa hypertextijärjestelmissä (Halasz & Schwartz 1990; Leggett & Killough 1991) ja hypertekstin muuntamista tekstiksi (Joyce 1991; Zheng & Rada 1993). Vastaus siihen, mitä dokumentteja kannattaa muuntaa hyperdokumenteiksi, piilee käyttäjien (lukijoiden) tarpeissa. Esimerkiksi tietyissä työtehtävissä täytyy nopeasti pystyä löytämään teknisiä yksityis-

kohtia suuresta määrästä materiaalia tai koulutuksessa halutaan tarjota mahdollisuus tarkastella aineistoa tietyistä näkökulmista. Hyperdokumenttien avulla voidaan tukea toimintaa näiden kaltaisissa tilanteissa.

Useita malleja ja erilaisia tukiohjelmistoja on kehitetty tukemaan konversioprosessia. Tekemämme kirjallisuusanalyysin perusteella tekstin hypertextiksi muuntamisen tutkimus keskittyy tiettyihin ohjelmistoihin sidottujen ratkaisujen esittelemiseen. Tässä artikkelissa rajaudutaan käsittelemään, miten paperimuotoissa tai elektroniset dokumentit voidaan muuntaa hypertextiksi. Käsittelemme konversiota kuvaamalla konversioprosessia varten kehitettyjä malleja (luku 2), sen tukemiseen suunniteltuja ohjelmistoja (luku 3) ja kuvaamalla kokemuksia yhdestä koekonversiosta (luku 4).

## 2. Muuntamisprosessin tarkastelua

*HEFTI*-malli (Hypertext Extraction From Text Incrementally) kattaa muuntamisen aina paperilla olevasta tekstistä hypertexti-välimuotoon, joka voidaan viedä mihin tahansa kyseisen muodon ymmärtävään hypertextijärjestelmään. Malli jakaa konversioprosessin kuudeksi vaiheeksi: tekstin valmistelu, solmujen valmistelu, indeksointi, linkkien luominen, järjestely ja linkkien hienosäätö (Chignell et al. 1991).

*Riner* (1991) on esittänyt käytännönläheisen 12-vaiheisen konversiomallin. Mallin vaiheet ovat sopivan lähtömateriaalin valinta, käyttöliittymän valinta ja kuvaus, loppukäyttäjän näytön huomioinnottaminen, graafisten elementtien käsittely, lähtötekstin konvertointi hypertextijärjestelmään ymmärtämään muotoon, dokumentin kuvauksen luominen, näkymien kuvausten luominen, konversion aloittaminen, lähtödokumentin virheiden käsittely, konversion tarkistaminen ja verifiointi, materiaalin käsittely loppukäyttäjien ympäristöön sopivaksi sekä lopuksi pakkaus ja levitys käyttäjille.

*Dvorak* et al. (1992) ovat kehittäneet menetelmän linkkirakenteiden käyttäjakeskeiseen määrittämiseen. Malli keskittyy selvittämään, kuinka linkkien tulee tukea käyttäjän lukutapoja. Malli ei puutu muihin konversiovaiheisiin kuin linkkien luomiseen ja hienosäätöön, mutta se tarjoaa rikkaan muista malleista poikkeavan näkökulman.

*Rada* (1992) on kuvannut menetelmän oppikirjojen konvertoimiseen välimuotoon ja edelleen tiettyille hypertextijärjestelmille, kuten Guide,

HyperTies, SuperBook. *Gu & Thiel* (1993) ovat esittäneet mallin, jossa ensin muodostetaan semanttinen hypermalli (solmu-linkkimalli dokumentista). Tämä muunnetaan relationaaliseksi hypermalliksi, jonka jälkeen suoritetaan itse konversio. Myös *Furuta et al.* (1989) esittävät tietokanta-orientoituneen lähestymistavan hyperdokumenttien rakentamiseen.

Jatkossa käyttämämme konversioprosessin vaihejako pohjautuu *HEFTI*-malliin (Chignell et al. 1991). Mallien vertailu on koottuna Taulukkoon 1. *Dvorakin* mallia lukuunottamatta kaikki mallit sivuavat ainakin jollain tavalla *HEFTI*-mallin viitettä ensimmäistä vaihetta. *Dvorakin* malli on keskittynyt linkkien luomiseen ja hienosäätöön. *Furutan* ja *Radan* mallit eivät suoranaisesti ota kantaa linkkien hienosäätöön. *Rinerin* ja *Radan* malleille sekä *HEFTI*-mallille on kehitetty myös omat tukiohjelmit.

### 2.1 Tekstin valmistelu

Tekstin valmistelu sisältää tekstin saattamisen elektroniseen muotoon (digitointi) ja tekstin muotoiluun. Tarkastelluista malleista *HEFTI* ja *Riner* ottavat kantaa digitointiin. Digitointi on yleensä hidas toimenpide, jossa apuna voidaan käyttää esimerkiksi skannereita ja optisia tekstintunnistusohjelmia. Laajasti määriteltynä se sisältää nidotun materiaalin irrottamisen ja valokopioimisen, sivujen digitoimisen bittikuvamuotoon, kuvien jakamisen osiin, joissa on joko pelkkää tekstiä tai pelkkää grafiikkaa ja tekstiä sisältävien osien käsittelyn optisella merkintunnistusohjelmalla ASCII-muotoon (*Riner* 1991). Digitoinnissa saattaa syntyä ongelmia erityisesti alkuperäisen dokumentin epäyhtenäisen rakenteen ja tekstin virheiden vuoksi (*Nunn et al.* 1988).

Dokumentin rakenteen ymmärtäminen on ratkaiseva tekijä muunnettaessa tekstiä hypertextiksi. Dokumentilla voidaan aina sanoa olevan sekä fyysinen että looginen rakenne. Dokumentin fyysistä rakennetta tukevat WYSIWYG-tekstinkäsittelyohjelmat (What You See Is What You Get), kuten Microsoft Word, ja dokumentin muotoiluohjelmat, kuten Unixin *troff*. (*Rada & Diaper* 1991.)

Dokumentit voidaan loogisen rakenteen puolesta jakaa karkealla tasolla vahvasti ja heikosti rakenteelliseen tekstiin. Vahvasti rakenteellisen tekstin looginen rakenne on selkeä ja sen konvertoiminen hypertextiksi on usein melko yksinkertainen teh-

tävä. Esimerkki tällaisesta on hakemistotyyppinen teksti, kuten tekniset manuaalit, sanakirjat ja tietosanakirjat (ks. esim. Raymond & Tompa 1987; Rada & Diaper 1991). *Heikosti rakenteellisen tekstin* konvertointi hypertextiksi vaatii paljon työtä dokumentin abstraktin rakenteen ja sen eri osien välisten yhteyksien määrittämiseksi. Esimerkiksi romaanit ja novellit ovat usein laajaa tietoisuuden virtaa, jonka lineaarinen rakenne on selvä mutta jonka loogista rakennetta ei ole lukijalle tarkkaan määriteltä (ks. esim. Harris & Cady 1988; Kahn 1990).

Dokumentin loogista rakennetta tukevat muotoilukielet esittävät dokumentin esimerkiksi puumuodossa tai hierarkkisena esityksenä. Elektroniassa tekstissä olevien muotoilumerkkien käyttöä kutsutaan *merkkaukseksi* tai *merkkaamiseksi* (engl. markup) (Salminen 1992). Tekstin rakenteen merkkaukseen on kehitetty useita standardeja, joiden avulla pyritään välittämään tekstin merkitys oikein käyttäjälle ja lisäämään tekstin käsittelyn automatisointia. Hyperdokumentin tarpeita silmäläpitiäen tarjoavat SGML, HTML (Berners-Lee & Connolly 1993) ja HyTime (Newcomb et al. 1991)

mahdollisuuden dokumentin rakenteen ja/tai esitysmuodon merkkaukseen.

## 2.2. Solmujen valmistelu

Solmujen valmistelu aloitetaan jakamalla teksti solmuksi ja nimeämällä solmut. Kunkin solmun alkuun voidaan haluttaessa lisätä solmun sisällön kuvaava tiivistelmä. Solmujen valmistelussa voidaan usein käyttää tekstin rakennetta hyväksi. Esimerkiksi kukin alaluku voi olla sopiva solmu ja solmun nimeksi voidaan ottaa alaluvun otsikko. (Chignell et al. 1991)

## 2.3. Indeksointi

Solmut indeksoidaan käyttämällä indeksisanoja. Usein alkuperäiset dokumentit tarjoavat oman hakemistonsa, jota voidaan käyttää indeksoinnin lähtökohdana. Kun indeksisanat on määriteltä, ne täytyy etsiä tekstistä (Chignell et al. 1991). Radan (1992) mallissa dokumenttien indeksitermit järjes-

Malli	Furuta	HEFTI	Riner	Rada	Dvorak	Gu
<b>Vaihe</b>						
<b>Tekstin valmistelu</b>						
Paperista elektroniseksi	E	T	T	E	E	E
Muotoilu/merkkauk	T	T	T	T	E	T
<b>Solmujen valmistelu</b>	T	T	T	T	E	T
<b>Indeksointi</b>						
Alkuperäinen indeksi	E	T	T	T	E	T
Vaihtoehtoisia indeksejä	T	E	E	T	E	E
<b>Linkkien luominen</b>						
Viittaavat	T	T	T	T	T	T
Assosiatiiiviset	T	T	T	E	T	T
Semanttiset	E	T	E	T	T	E
<b>Organisointi</b>						
Hierarkiat	T	T	T	T	E	T
Maamerkit	T	T	E	E	E	E
Vaihtoehtoinen rakenne	E	E	T	T	E	E
<b>Linkkien hienosäätö</b>	E	T	T	E	T	T
<b>Tukiohjelmisto</b>	E	T	T	T	E	E

T=tuettu ominaisuus

E=ei tuettu ominaisuus

*Taulukko 1.* Konversiomallien vertailu vaiheittain.

tetään aakkosellisesti indeksisolmuun ja rakennetaan linkit indeksisolmun termien ja tekstissä olevien esiintymien välille. Lisäksi voidaan esimerkiksi tieteellisistä artikkeleista koostuvaan hyperdokumenttiin luoda indeksisolmu, johon kerätään artikkelien ja niiden kirjoittajien nimet sekä muut taustatiedot. Gu & Thiel (1993) ovat raportoineet myös mahdollisuudesta käyttää indeksin rakentamisessa hyväksi tesaurusta.

## 2.4. Linkkien luominen

Linkkien luominen on hypertextin rakentamisen keskeisin tehtävä. Linkkien rakentamisessa voidaan edetä esimerkiksi siten, että ensin rakennetaan hierarkkiset linkit dokumentin kappalejaon perusteella, luodaan linkit erilaisiin viittauksiin, kuten kappaleisiin, kuviin ja avaintermeihin, ja rakennetaan linkit muihin dokumentteihin ristiviittausten avulla (Riner 1991). Oppikirjan ollessa kyseessä kukin kirjallisuusviite linkitetään lähde-tietoihin ja harjoitustehtävät niiden vastauksiin (Rada 1992).

*Linkkien tyypittäminen* tuo lisäarvoa käyttäjälle. HEFTI-mallissa linkit voidaan jakaa viittaaviin, assosiatiiivisiin sekä semanttisiin yhteyksiin perustuviin linkkeihin. *Viittaavat linkit* johdattavat käyttäjän solmun tietystä kohdasta toiseen solmuun tai tiettyyn kohtaan toisessa solmussa. Viittaavat linkit voidaan jakaa edelleen navigoiviin linkkeihin, joilla siirrytään toiseen kohtaan tekstissä, korvaaviin linkkeihin, jolloin teksti-ikkuna korvataan esimerkiksi tekstiin liittyvällä kuvalla, ja annotaatiolinkkeihin, joita esimerkiksi alaviitteet edustavat. *Assosiatiiiviset linkit* kuvaavat kahden solmun aiheiden välistä yhteyttä. Nämä linkit yhdistävät solmut kokonaisina toisiinsa. Ideana on kertoa tietystä asiasta kiinnostuneelle käyttäjälle muista asiaan liittyvistä aiheista. Hypertextiä voidaan pitää eräänlaisena semanttisena verkkona, joten linkkejä voidaan rakentaa *semanttisten yhteyksien* perusteella. Semanttisia yhteyksiä ovat mm. synonyymit ja muut samankaltaisuudet. (Chignell et al. 1991)

Gu & Thiel (1993) käsittelevät kolmenlaisia linkkejä: rakenteen määritteleviä linkkejä, ristiviittausslinkkejä ja tietämyspohjaisia linkkejä. Tietämyspohjaisia linkkejä ovat tilastollisilla 'lähi-naapuri'-laskelmilla (engl. nearest neighbor) saadut solmujen väliset linkit, linkit solmujen ja tesauruksen välillä sekä manuaalisesti määritetyt linkit.

Nykyiset hypertextiksi konvertoidut oppikirjat eivät yleensä ole käyttäjakeskeisiä, vaan lukija joutuu tyytymään paperiversiosta tuttuihin työkaluihin (sisällysluettelo, indeksi, alaviitteet ym.), joihin on ehkä lisätty indeksejä tai tarkempia sisällysluetteloita. Dvorak et al. (1992) ovat kehittäneet menetelmän *linkkirakenteiden käyttäjakeskeiseen määrittämiseen*. Menetelmä keskittyy selvittämään, kuinka linkkien tulee tukea käyttäjän lukutapoja. Esimerkiksi oppikirjoja luetaan *tietyn tavoitteen saavuttamiseksi*. Tavoite jakaantuu tehtäviin, joiden suorittaminen vaatii erilaisia lukutapoja, kuten tekemiseen tähtäävää lukemista (artikkelin kirjoittaminen), opiskelevaa lukemista, etsivää/tutkivaa lukemista, "huvini vuoksi" lukemista ja opettamiseen tähtäävää lukemista. Käyttäjakeskeinen hyperdokumentin rakentaminen vaatii monenlaisia linkkirakenteita auttamaan lukijoita saavuttamaan tavoitteensa (Dvorak et al. 1992, 619, 622–624). Myös Ingwersen (1992, 207–210) käsittelee erityyppisten tehtävien huomioonottoa tiedon haussa.

Lingvistiikan tutkimuksessa on havaittu, että retoriset predikaatit (engl. rhetorical predicates) ovat perusyksiköitä, joiden avulla kirjoittaja välittää tietoa lukijalle. Tekstin sisältämät lukijaa palvelevat tiedonvälitystehtävät täytyy tunnistaa, ryhmitellä ja analysoida sen suhteen, miten lukijat käyttävät niitä tavoitteensa saavuttamiseksi. Yksi tapa tukea lukijan tavoitteita on käyttää kahdentyyppisiä linkkirakenteita, eksplisiittisiä ja implisiittisiä. *Eksplisiittiset linkit* tukevat pääsyä johonkin lukijan helposti tunnistamaan aiheeseen tekstissä sisällysluettelon, indeksin tms. kautta. *Implisiittiset linkit* tukevat tekstin tiedonvälitystä tarjoamalla lukijalle tietoa jostain erityisestä aiheesta. Nämä linkit ovat tehtäväkohtaisia, joten lukija tietää etukäteen, minkälaista tietoa niitä käyttämällä saa esiin. Lukijan tiedonvälitystehtäviä ovat esimerkiksi selitys, luokittelu, kuvaus, vertailu ja vastakkainasettelu, määritelmä ja yhteenveto. Myös muistiinpanojen tekeminen on lukijalle tärkeä apukeino tavoitteensa saavuttamisessa. Jokaisessa solmussa lukijan on pystyttävä tekemään muistiinpanoja ja tarvittaessa tulostamaan niitä paperille. (Dvorak et al. 1992, 625–626.)

## 2.5. Organisointi

Järjestely on tärkeä osa konversiota, koska se vaikuttaa merkittävästi hyperdokumentin käytettävyyteen. HEFTI-malli tarjoaa kaksi keinoa solmujen

organisointiin: *solmujen järjestely hierarkioiksi* sopivien linkkien avulla ja *tärkeiden solmujen painottaminen maamerkeinä*. Linkkien luominen semanttisten yhteyksien perusteella ja solmujen organisointi hierarkioiksi sopivien linkkien avulla ovat osittain päällekkäisiä toimia. Ne on kuitenkin erotettu omiksi vaiheikseen: linkkejä luotaessa tavoitteena ovat paikalliset navigointia helpottavat linkit, kun taas järjestelyssä luodaan laajempi organisaatorakenne, kuten hierarkia. Maamerkkejä voidaan käyttää sisääntuloväylänä tiettyyn aiheeseen ja solmujen välisen navigoinnin apuna. (Chignell et al. 1991) Lisäksi järjestelyssä voidaan hyperdokumenttiin lisätä sitovia solmuja, kuten otsikkosivu-, johdanto- ja yhteenveto-solmuja (Furuta et al. 1989).

## 2.6. Linkkien hienosäätö

Hyperdokumentin laatua voidaan parantaa hienosäätämällä linkkejä. Tällöin pohditaan erityisesti hyperdokumentin käytettävyyttä ja ratkaistaan esimerkiksi, ovatko yksittäisiä solmuja yhdistävät linkit todella järkeviä ja auttavatko solmuryhmien väliset linkit lukijan navigointia ja ymmärtämistä. Yleensäkin kysytään, mitä linkkejä voitaisiin lisätä tai poistaa hyperdokumentin yhtenäisyyden parantamiseksi. (Chignell et al. 1991) Hienosäädettäessä tulisi myös tarkistaa, että jokaiselle linkille löytyy lähtö- ja päätepiste (Riner 1991), ja poistaa epäyhtenäisyydet, mahdolliset duplikaattitietueet yms. (Gu & Thiel 1993). Rada (1992) pohtii hypertextin käytettävyyksymyksiä vaihtoehtoisten organisointien rakentamisen yhteydessä. Hän esittää yksinkertaisen mutta vaikeasti tuettavan käytettävyyssäännön: käyttäjän pitäisi pystyä helposti löytämään linkit, jotka johtavat hänet senhetkisen tehtävän ratkaisuun.

## 3. Automatisoitu konvertointi

Konvertointi tekstistä hypertextiksi voidaan suorittaa joko manuaalisesti tai automaattisesti. *Manuaalisessa konversiossa* dokumentti syötetään haluttuun hypertextijärjestelmään, jonka jälkeen rakennetaan linkit manuaalisesti. Dokumentin syötöä varten useissa hypertextijärjestelmissä on ns. import-toiminto, joka saattaa hyväksyä esimerkiksi ASCII-muotoisten dokumenttien lisäksi monien tekstinkäsittelyohjelmien formaatteja (Riner 1991).

Esimerkiksi ToolBookissa on toiminto, jolla tekstin käsittelydokumentti voidaan tuoda ToolBookiin ja ohjainmerkkien avulla jakaa automaattisesti useaksi hyperdokumentin "sivuksi". Toiminto on vielä suhteellisen kankea, mutta esimerkiksi HyperCardissa ei edes ole vastaavaa ominaisuutta.

*Automaattinen konversio* soveltuu tilanteisiin, joissa samankaltaiset konversiotehtävät toistuvat tai konvertoitavan materiaalin määrä on suuri. Tutkimustyötä varten saattaa kannattaa rakentaa pieniä prototyyppidokumentteja manuaalisesti, mutta jos konversion kohteena on kymmeniä tuhansia sivuja, manuaalisen käsittelyn hitaus ja vaadittu työpanos nousevat esteeksi (Riner 1991). Vaikka toisaalta jotkin materiaalit ovat sellaisia, ettei niiden konvertointia kannata automatisoida, voidaan apuohjelmia usein silti käyttää eri vaiheissa esimerkiksi lähtödokumentin rakenteen muokkaamiseen (Furuta et al. 1989). Nykyiset hypertextijärjestelmät tukevat yleensä hyvin heikosti tekstin automaattista muuntamista hypertextiksi.

Automaattisen konvertoimisen tukemiseksi on esitetty useita tukiohjelmistoja, joilla konversio saadaan yleensä puoliautomaattiseksi. Tällöin solmujen, linkkien ja indeksien rakentaminen on pitkälti automatisoitu. Lisäyksiä ja korjauksia voidaan tehdä manuaalisesti apuohjelmien tuottamiin rakenteisiin ja sisältöön. Ohjelma saattaa esimerkiksi ehdottaa suuren määrän linkkejä, joista käyttäjä karsii pois tarpeettomat (Rearick 1991). Menetelmä on usein nopeampi kuin, että käyttäjä lisäisi itse kaikki tarpeelliset linkit.

Automatisoitu konversio edellyttää aina jonkinlaista tekstin analysointia. Tietokoneavusteiselle linkkien muodostukselle on kehitetty useita tekniikoita, jotka tosin ovat pääasiassa vielä tutkimusasteella. Näitä ovat esimerkiksi leksikaalinen analyysi, tilastollinen analyysi, kieliopin analyysi ja semantiikan analyysi (Rearick 1991). Rajaton luonnollisen kielen ymmärtäminen on nykyisten menetelmien tavoittamattomissa, mutta selvästi rajatun aihealueen käsittely on mahdollista (Castell & Verdejo 1991). Lingvistisestä käsittelystä ja sen ongelmista kirjoittaa Karlsson (1993).

HEFTI-mallia tukemaan on rakennettu Macintoshissa toimiva *HEFTI-0.5* -prototyyppi (Chignell et al. 1991). Rinerin (1991) konversiomallia tuetaan Texas Instrumentsin *HyperTRANS*-ohjelmistolla, joka automatisoi ensisijaisesti linkkien tunnistusta ja rakentamista. *Radan* (1992) mallia tukee joukko Unix-ympäristössä toimivia ohjelmia. Edellä esitettyjen Radan, Rinerin ja

HEFTI-konversiomallien tukiohjelmistojen lisäksi on olemassa myös muita konversion tuki-ohjelmistoja, kuten *SmarText* (Rearick 1991), *Super-Library* (Catenazzi & Argentisi 1991), *Expanded Book* (1992) ja *HyperDOC* (Priha & Kujala 1992). Taulukkoon 2 on koottu yhteenvetona näiden ohjelmistojen antama tuki konversio-prosessille.

### 3.1. Tekstin valmistelu

Mitkään tutkituista konversio-ohjelmista eivät automatisoi tekstin muuntamista paperista elektroniseksi, mihin pitää käyttää esimerkiksi skannereita, tekstintunnistusoohjelmia ja tavutuksen tarkastaja -ohjelmia. HyperDoc-ohjelmiston pääpaino on tekstin muotoilussa ja merkkäamisessä ja se onkin tukiohjelmistoista ainoa, joka tukee tekstin automaattista merkkäamista. Dokumentti merkataan aluksi SGML-muotoon, jolloin määritetään dokumentin fyysinen rakenne. HyperDoc pystyy automaattisesti muuntamaan Xeroxin

Ventura Publisher-ohjelmalla luotuja ja RTF-muotoisia dokumentteja SGML-muotoon. Muunlaisia dokumentteja muunnettaessa joudutaan käyttämään erillistä muunto-ohjelmaa. Esimerkiksi HyperCardiin konvertoitavan dokumentin teksti sijoitetaan automaattisesti korttien tekstikenttiin. Tekstin muotoilu tehdään manuaalisesti. Dokumentin grafiikka voidaan sijoittaa joko kortteihin manuaalisesti tai omiin erillisiin grafiikkaikkunoihin. (Priha & Kujala 1992, 113–115.)

### 3.2. Solmujen valmistelu

Kaikki ohjelmistot tukevat solmujen valmistelua jollain tavalla. Esimerkiksi HEFTI-0.5:ssä kustakin alkuperäisen dokumentin kappaleesta tai alakappaleesta tehdään automaattisesti solmu. Ohjelma nimeää solmun kappaleen otsikon mukaan ja merkkää kunkin solmun sijainnin dokumentin rakenteessa. Tätä merkkiä käytetään hyväksi myöhemmin organisoitaessa hyperdokumenttia. (Chignell et al. 1991)

Tukiohjelmisto Vaihe	HEFTI-0.5	Rada	HyperTRANS	SmarText	Super-Library	Expanded Book	HyperDOC
<b>Tekstin valmistelu</b>							
Paperista elektroniseksi	E	E	E	E	E	E	E
Muotoilukäskyt/merkkäus	E	E	E	E	E	E	A
<b>Solmujen valmistelu</b>	A	A	A	A	A	A	A
<b>Indeksointi</b>							
Alkuperäinen indeksi	P	A	A	E	E	E	?
Vaihtoehtoisia indeksejä	E	A	E	A	A	P	?
<b>Linkkien luominen</b>							
Viittaavat	E	A	A	A	A	P	P
Assosiatiiiviset	A	A	A	A	A	P	?
Semanttiset	E	E	E	E	E	E	?
<b>Organisointi</b>							
Hierarkiat	A	A	A	A	A	A	?
Maamerkit	E	E	E	E	E	E	?
Vaihtoehtoinen rakenne	E	P	A	A	A	E	?
<b>Linkkien hienosäätö</b>	E	E	E	E	E	E	?

A=automatisoitu

P=puoliautomatisoitu

E=ei tuettu

?=lähteestä ei voitu selvästi päätellä kyseistä kohtaa

Taulukko 2. Tukiohjelmistojen vertailu.

### 3.3. Indeksointi

Suurin osa ohjelmistoista tukee ainakin jonkin verran indeksin rakentamista. Esimerkiksi HEFTI-0.5:ssä ohjelmaan syötetään erikseen alkuperäisen dokumentin indeksitermit, jonka jälkeen ohjelma tekee linkit indeksitermien esiintymien ja indeksin välille (Chignell et al. 1991). Vastaavasti Radan ohjelmisto tunnistaa indeksitermit dokumenttiin tehdyn indeksitermien merkkauksen perusteella, kopioi ja järjestää ne aakkosellisesti indeksisolmuun ja luo linkit indeksisolmun termien ja tekstissä olevien termien välille (Rada 1992). SuperLibraryn indeksi sisältää kaikki tekstin sanat. Myös useiden dokumenttien yhteisiä indeksejä voidaan tehdä. (Catenazzi & Argentisi 1991)

Käyttäjä voi rakentaa indeksin myös manuaalisesti. Automaattinen indeksointi on periaatteessa suositeltavampaa, koska manuaalisesti rakennetut indeksit ovat usein puutteellisia. Täysin automaattista indeksointimenelmää ei kuitenkaan ole toistaiseksi kehitetty, joten on löydettävä käytännöllinen kompromissi automaattisen ja manuaalisen indeksoinnin välillä.

### 3.4. Linkkien luominen

Linkkien luominen on konversion ydin. Esimerkiksi HyperTRANSille syötetään dokumentin kuvaustiedot, missä muodossa esimerkiksi dokumentin sisällysluettelo, kuvaluettelo, lähdeluettelo ja indeksi ovat. HyperTRANS jakaa dokumentin solmuiksi, rakentaa hierarkkiset linkit sisällysluettelon mukaan, tai jos dokumentilla ei ole sisällysluettelo, luo sen, rakentaa assosiatiiviset (ei-hierarkkiset) linkit dokumentin solmujen välille perustuen erilaisiin viittauksiin (lukuihin, kappaleisiin, kuviin, taulukoihin, avaintermiin, indeksiin ja sanastoon) ja rakentaa assosiatiiviset linkit muihin dokumentteihin (ulkoiset linkit) perustuen risti-viittauksiin. HyperTRANS huolehtii myös konversiossa ilmenneistä ongelmista. Se pitää yllä poikkeuslokia (engl. exception log), johon se kuvaa kaikki automaattisessa konversiossa kohdatut ongelmat. Ongelmat ovat joko virheitä dokumentin kuvauksessa tai virheitä itse dokumentissa. Tyypillisiä virheitä ovat epäyhtenäiset nimet tai numerointi, puuttuvat kuvat tai taulukot ja indeksitermit, jotka eivät esiinny tekstissä. Dokumentin kuvauksessa olevat virheet on helppo korjata, mutta itse

dokumentissa olevat virheet voivat vaatia neuvoteltua dokumentin kirjoittajan kanssa ja jopa joidenkin dokumentin osien uudelleenkirjoitusta (Riner 1991). HEFTI-0.5 laskee jopa kahden solmun "samankaltaisuudelle" arvon ja luo linkin solmujen välille, jos samankaltaisuusarvo on riittävän korkea. (Chignell et al. 1991)

### 3.5. Organisointi

Hyperdokumentin hierarkkista organisointia tukevat kaikki tutkitut konversion tukiohjelmistot. Lisäksi Rada, HyperTRANS, SmarText ja SuperLibrary tukevat vaihtoehtoisten rakenteiden luomista. Esimerkiksi Super-Libraryssä dokumentit käsitellään aluksi erillisinä ja liitetään myöhemässä vaiheessa yhteiseen kontekstiin dokumenttien välisillä linkeillä, yhteisillä indekseillä ja dokumenttien kokonaisrakenteen esittäväällä kuvauksella. Kukin dokumentti säilyttää silti oman sisällysluettelonsa ja itsenäisen luonteensa. Vaihtoehtoinen rakenne toteutuu siten, että useammasta dokumentista voidaan luoda yhteinen kokonaisrakenne (Catenazzi & Argentisi 1991). Rada (1992) kuvaa tavan, jolla hyperdokumentin eri osissa toistuva samankaltainen rakenne voidaan esittää eksplisiittisenä vaihtoehtoisena rakenteena hierarkkisen rakenteen sijaan.

### 3.6. Linkkien hienosäätö

Linkkien hienosäätöä ei tue mikään tutkituista ohjelmistoista. Kyseistä konversiovaihetta on erittäin vaikea tukea ohjelmistoilla, koska hienosäädössä on usein kyse jo sovellusalueen semanttisesta tuntemisesta. Chignell et al. (1991) ehdottavat, että tarkoitukseen voitaisiin käyttää jotakin erillistä hypertextistä muodostuvaa verkkoa analysoivaa ohjelmaa.

### 3.7. Hyperteksti-selaajat

*SmarText Electronic Document Construction Set* tarjoaa toisenlaisen tavan luoda ja käyttää hyperdokumenteja. SmarText on *hypertekstiselain*, joka ei tallenna hyperdokumenteja, vaan käyttää suoraan tekstinkäsittelyohjelman tiedostoja muuntamalla ne 'lennossa' loogiseen muotoon hypertekstiksi. Se rakentaa dokumentille automaattisesti

indeksin ja sisällysluettelon käyttäen ASCII-tiedostoja. Sisäiset ristiviittaussuhteet SmartText rakentaa olettaen, että jos indeksitermi esiintyy useammassa kohdassa samassa kappaleessa, kappaleen aihe on semanttisesti lähellä itse termiä. Automaattisesti luotuihin ristiviittaussuhteisiin voidaan tehdä lisäyksiä tai poistoja manuaalisesti. SmartText rakentaa automaattisesti näkymän, joka perustuu tekstitiedostoihin. Käyttäjä voi määrittellä lisäksi vaihtoehtoisia näkymiä haluamansa määrän. (Rearick 1991)

Myös *Windows Help* -tiedostojen luomiseen on tehty konversio-ohjelmia, kuten *HyperTrack*, joka muuntaa *Ms Wordin* tiedostoja automaattisesti help-tiedostoiksi. Jotkut konversio-ohjelmat muuntavat tiettyä yleistä lineaarista tekstimuotoa (esimerkiksi RTF:ää) HTML-muotoon, jota World Wide Web-selaimet (Berners-Lee et al. 1994), kuten *Netscape*, *Mosaic*, *Cello*, *Lynx* ja *LineMode Browser* ymmärtävät. Vastaavia SGML-muotoisia dokumentteja ymmärättäviä hypertextiselaimia ovat esimerkiksi *Hlpdk* ja *HyperHelper*.

#### 4. Koekonversio

Koekonversion kohteena oli n. 70-sivuinen dokumentti (16992 sanaa, 18 kuvaa, neljä taulukkoa, viisi alaviitettä ja n. 270 lähdeviitettä (useita viitteitä kuhunkin lähdeluettelon lähteeseen). Dokumentti oli kirjoitettu Microsoft Word versiolla 5.0. Konversio suoritettiin Macintosh SE/30:lla, johon oli asennettu käyttöjärjestelmän versio 7.0.1 ja jonka keskusmuisti oli laajennettu 5 megatavuun. Linkkien hienosäätöä ei tässä koekonversiossa suoritettu. Konversio-ohjelmana käytettiin *Expanded Bookia* (1992). Se on Voyager Companyn kehittämä konversio-ohjelma, jolla voidaan muuntaa tiettyjen tekstinkäsittelyjärjestelmien dokumentteja tai ASCII-tekstiä HyperCard-pinoiksi. Kokeen suoritti artikkelin toinen kirjoittaja ilman aiempaa kokemusta kyseisestä konversio-välineestä.

Expanded Book automatisoi dokumentin jaon solmuihin, sisällysluettelon rakentamisen sekä indeksoinnin. Se jakaa lähtökäsitteiden automaattisesti solmuiksi, joita voidaan edelleen muokata manuaalisesti. Ennen dokumentin viemistä Expanded Bookiin sitä kannattaa muokata ja merkata esimerkiksi hyperdokumenttiin haluamansa sivujaon mukaan. Expanded Book ei kuitenkaan pysty käyttämään hyväksi käyttäjän merkkauksia, vaan merkkaukset on tarkoitettu manuaalista muokka-

usta helpottamaan. Kukin dokumentin luku jaetaan ennen järjestelmään sisääntuomista omaksi tiedostokseen ja tiedostojen nimiksi annetaan kukin luvun otsikko. Expanded Book muodostaa tiedostojen nimistä (luku-otsikoista) sisällysluettelon, jota voidaan muokata manuaalisesti.

Expanded Bookissa annotaatioiden rakentamista tuetaan apuvälineillä. Erilaisia annotaatioita ovat tekstuaaliset annotaatiot, annotaatiot, jotka vievät lukijan toiseen paikkaan kirjassa tai toiseen kirjaan, graafiset annotaatiot, audioannotaatiot, QuickTimen video-annotaatiot, CD-levy- tai videolevyannotaatiot ja erikseen ohjelmoidut annotaatiot. Teksti-annotaatioita ovat esim. alaviitteet ja lähdeviitteet, jotka voidaan toteuttaa teksti-ikkunoiden avulla. Toiseen paikkaan johdattavia annotaatioita eli navigoivia annotaatioita käytetään dokumentin sisältämien ristiviittausten toteuttamiseen. Expanded Book pystyy luomaan sekä yksi- että kaksisuuntaisia linkkejä. Graafisten annotaatioiden avulla kuvat saadaan esille lukijan niin halutessa. Audioannotaatioilla voidaan kirjaan liittää äänihuomautuksia, QuickTimen video-annotaatioilla pieniä "elokuvan" pätkiä ja CD-levy- ja videolevy-annotaatioilla haluttuja osia levystä. Erikseen ohjelmoiduilla annotaatioilla voidaan HyperCardin HyperTalk-ohjelmointikielellä esim. käynnistää muita ohjelmia tai esittää yhtäaikaaisesti eri annotaatiotyyppejä.

Expanded Book auttaa indeksin rakentamista lähinnä indeksisanojen hakutoiminnolla. Käyttäjän on itse määriteltävä indeksisanojen lista ja etsittävä hakutoiminnon avulla sanojen kaikki ilmentymät tekstistä. Indeksisanalista linkitetään tekstin sisältämiin indeksisanoihin automaattisesti. Lisäksi voidaan parantaa luodun hyperdokumentin ulkoasua esimerkiksi upottamalla tekstin sekaan grafiikkaa, parantamalla kannen ja sisällysluettelosivun ulkoasua tai muokkaamalla valikoiden sisältöä. Nämä toimenpiteet ovat pääosin manuaalisia, mutta niissä käytetään apuna Expanded Bookin ja HyperCardin toimintoja.

##### 4.1. Tekstin valmistelu

Tekstin muuntamista paperimuodosta digitaaliseen ei koekonversiossamme tarvinnut tehdä, koska dokumentti on kirjoitettu tekstinkäsittelyohjelmalla. Solmuihin jako suoritettiin osittain manuaalisesti jo Word-dokumentissa. Tekstistä erotettujen kuvien viittaukset tekstissä vahvennettiin ja



alaviitteiden viittaukset merkattiin erikoismerkin '•' sekä alaviitteen numeron yhdistelmällä, jotta myöhempi linkkien rakentaminen helpottuisi. Tekstiin merkattiin vain viittaukset alkuperäisestä dokumentista erotettuihin elementteihin, eli viittaukset kuviin, lähteisiin, alaviitteisiin ja taulukoihin. Expanded Book ei ymmärrä yleisiä merkkauksikieliä. Seuraavaksi tekstiä muokattiin luettavammaksi lisäämällä kursiivilla kirjoitettujen tekstien perään välilyöntejä manuaalisesti. Tätä ei yllättävää kyllä ollut automatisoitu. Muunlainen muokkaus, kuten tekstin sisennykset, suoritettiin hyperdokumentissa.

## 4.2. Solmujen valmistelu

Word-dokumentti oli aluksi yhtenä palana, mutta ennen Expanded Bookiin vientiä kukin pääluku pilkottiin omaksi tiedostokseen ja tiedoston nimek-

si annettiin luvun numero ja nimi. Kaikki kuvat erotettiin omiksi tiedostoikseen ja nimeksi annettiin 'Kuva 1', 'Kuva 2' jne. Alaviitteet erotettiin omaksi Word-muotoiseksi tiedostokseen. Macintoshin tiedostojen nimeämissäntöjen takia näiden tiedostojen nimien pituus voi olla vain 31 merkkiä. Varsinaiseen konversioon kuului myös näytön määrittäminen, syntyvän hyperdokumentin nimeäminen ja muiden asetusten, kuten hakemistopolun, määrittäminen. Seuraavaksi tekstit tuotiin sisään luku kerrallaan. Ennen sisääntuontia määriteltiin muotomäärityksiä, kuten alkuperäisen dokumentin fonttien koon säilyttäminen ja lukujen otsikoiden muotoilu. Import-toiminto (sisääntuonti) lohko dokumentin alkuperäiset sivut näytölle sopiviksi paloiksi (solmuiksi), jolloin tässä 'classic'-koossa yhdestä sivusta tuli noin kolme Expanded Book -sivua. Lukujen nimeämisessä ilmeni sama rajoitus kuin kirjan nimeämisessäkin: maksimi-

**File Edit Go Tools Objects Font Style Kirjat Toolkit** 

### 3.1.3 HyTime-standardi

*HyTime* (Hypermedia/Time-based Document Structuring Language) on SGML:ään perustuva standardoitu muotoilukieli hyperteksti-, hypermedia-, sekä aika- ja tilapohjaisten dokumenttien loogisen rakenteen määrittelyyn. HyTime tarjoaa laajennuksen SGML:n kykyyn ilmaista hypermedian vaatimaa informaatiota kuten linkkejä ja multimediaobjektien aikataulutusta. **Kuva 6** esittää hyperdokumentin DTD:n luomisessa tarvittavia osatekijöitä. (Newcomb et al. 1991)

HyTime koostuu neljästä päämodulista (kts. **kuva 7**):

*perustamoduli* (base module) sisältää perustoiminnot,

*sijaintimäärittelymoduli* (location address module) osoittaa

pituus saa olla 31 merkkiä, joten Expanded Book katkaisi melkein kaikkien lukujen nimet. Lukujen nimet voitiin kuitenkin muuttaa hyperdokumentissa pitemmiksi. Sivut eivät saaneet mitään erityistä nimeä (jos ei sivunumeroa pidetä nimenä). Tekstin kappale- ja sivujaot sekä muut muotoilut muokattiin sopiviksi käymällä dokumentti sivu kerrallaan manuaalisesti läpi ja tekemällä tarpeelliset muutokset lisäämällä/poistamalla rivinvaihtoja sekä leikkaamalla/liimaamalla tekstin osia.

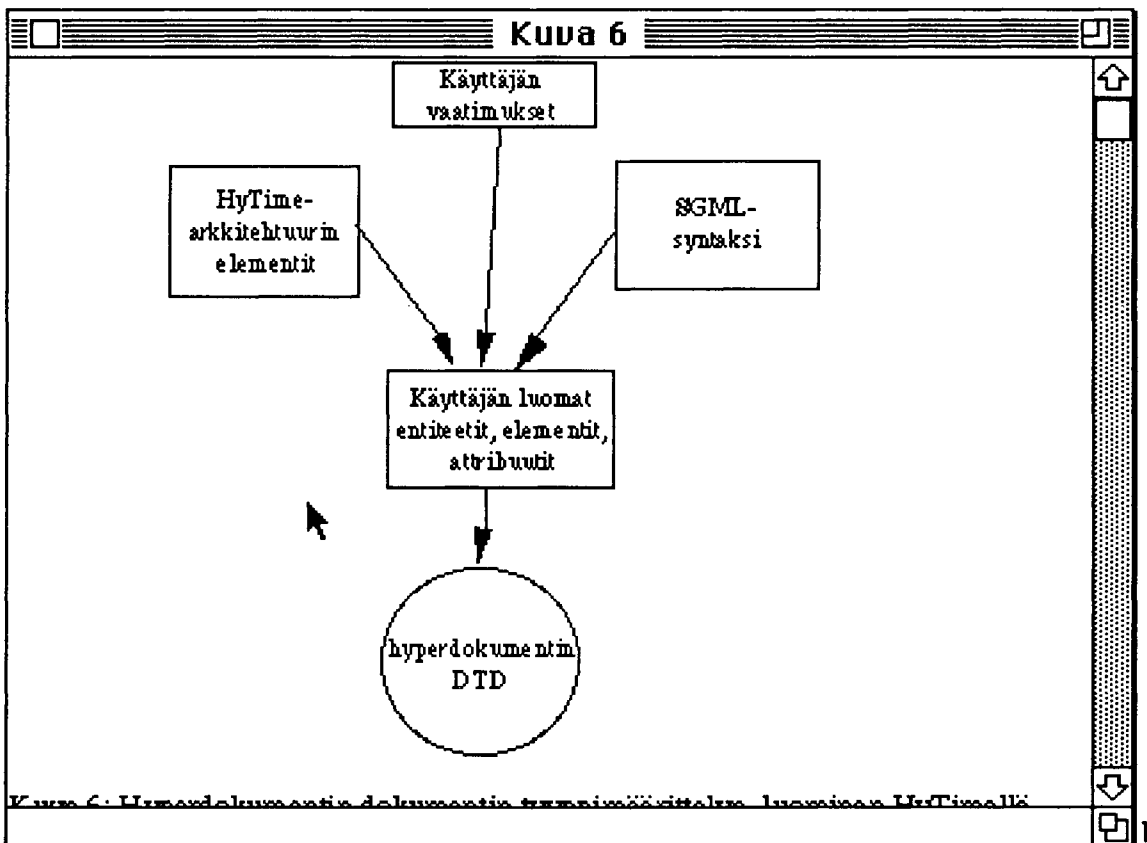
### 4.3. Indeksointi

Dokumentissa ei ollut valmista indeksiä. Indeksitermit oli määriteltävä manuaalisesti ja ne piti hakea tekstistä vapaatekstihaun avulla. Linkkien rakentaminen täytyi tehdä kirjoittamalla sivunumerot termien perään. Tämän jälkeen hyperdokumentti

osasi avata automaattisesti oikean sivun. Jos kuitenkin dokumenttia muokataan siten, että sivunumerointi vaihtuu, myös indeksi menee sekaisin.

### 4.4. Linkkien luominen

Navigoivia linkkejä rakennettiin lähdeviitteiden ja lähdeluettelon välille. Lähteet sijoitettiin omaksi luvukseen kirjan loppuun ja tekstissä olevien tekstiviittausten ja lähdeluettelon välille rakennettiin kaksisuuntaiset linkit. Tämä vaihe kulutti runsaasti aikaa, koska tekstiviitteitä oli paljon ja jokaiselle piti erikseen rakentaa linkki. Tämän toiminnon tulisi olla automatisoitu. Expanded Book rakensi navigoivia linkkejä automaattisesti sivunumerolta toiselle ja luvusta toiseen. Korvaavia linkkejä rakennettiin manuaalisesti taulukoiden, kuvien ja alaviitteiden esittämiseksi lukijalle (ks. kuvat 1 ja



Kuva 2. Kuvaikkuna.

2). Dokumentissa olleet taulukot eivät siirtyneet oikeanmuotoisina hyperdokumentteihin, vaan ne täytyi siirtää erikseen Expanded Bookiin. Ne siirrettiin kopioimalla kukin taulukko Wordissa, sulkemalla Word (muisti ei riittänyt pitämään auki sekä Wordia että Expanded Bookia/HyperCardia), aukaisemalla hyperdokumentti, luomalla tekstiikkunaan linkki tekstin kohdasta, jossa viitattiin ko. taulukkoon, ja sijoittamalla taulukko luotuun ikkunaan. Isot taulukot eivät sovi kerralla Expanded Book -teksti-ikkunan näyttämälle alueelle pienessä näytössä, mutta teksti-ikkunaa voidaan vierittää vierityspalkeilla. Sivuttaisvieritys ei jostain syystä toiminut, mikä aiheutti ongelmia leveiden taulukoiden kohdalla. Wordilla piirrettyjen kuvien siirto piti tehdä MacDraw II -piirto-ohjelman kautta, koska Expanded Book ymmärsi vain PICT-kuvamuotoa. Siirto MacDraw II :een sujui muuten ongelmitta, mutta kuvissa olleiden nuolten varret katosivat. Tämän siirron kierrättämisen olisi tietysti voinut välttää piirtämällä kuvat alunperinkin MacDraw II :lla eikä Wordin piirtotyökalulla. Kuvien siirto MacDraw II :sta hyperdokumenttiin sujui nopeasti. Expanded Bookin avulla rakennettiin linkit tekstissä olevien kuvaviittausten ja kuvien välille. Alaviitteet sijoitettiin teksti-ikkunoihin, jotka tulevat näkyviin lukijan painaessa kyseistä viittausta tekstissä. Poikkeuksena oli kaksi alaviitettä, joihin viitattiin taulukossa. Koska taulukotkin ovat teksti-ikkunassa, eikä kahden teksti-ikkunan välille pystytty rakentamaan linkkiä, alaviitteet sijoitettiin samaan teksti-ikkunaan viitteen sisältävän taulukon kanssa.

Jos linkkien rakentamisen jälkeen tekstiä vielä muokataan Expanded Bookissa, *ankkurit menevät tekstiä muokatessa sekaisin, koska ne eivät seuraa tekstin mukana!* Ilmeisesti ne on toteutettu tekstin päälle läpinäkyvinä sijoitettavilla painikkeilla, jotka eivät ole osa tekstiä. Expanded Book näyttääkin olevan työkalu nimenomaan hyperdokumenttina julkaistaville kirjoille, kuten novelleille, romaaneille, ja teknisille dokumenteille, joita ei julkaisun jälkeen enää muokata (jos tulee uusi "versio", se täytyy konvertoida vastaavasti).

#### 4.5. Organisointi

Expanded Book rakensi automaattisesti hierarkisen rakenteen järjestelmään tuotujen lukujen ja alalukujen järjestyksen perusteella. Hierarkiaa saattoi tarvittaessa vielä manuaalisesti muokata. Hie-

rarkia ilmeni sisällysluettelossa, mutta siihen tulivat alunperin vain pääluvut, eivätkä alaluvut näkyneet siinä ollenkaan. Jos alaluvut haluaa saada näkyviin, täytyy kaikkien hakemistotasojen luvut tallentaa omiksi tiedostoikseen. Toinen vaihtoehto on asettaa alaluvut omiksi luvuikseen vasta hyperdokumentissa. Hakemiston sisennykset täytyy kuitenkin manuaalisesti muokata. Periaatteessa Expanded Book käyttää myös maamerkki-solmuja navigoinnin apuna. Sellaisia ovat lukujen otsikkosivut, joiden välillä pääsee liikkumaan mistä solmusta tahansa.

#### 4.6. Aikataulusta

Konversio kuvailussa laajuudessa vei aikaa noin 15 tuntia (908 min) taulukon 3 esittämällä tavalla. Koska kyseessä oli tekijän ensimmäinen Expanded Bookilla tekemä konversio, oletimme että konversio nopeutuisi toisella kerralla työkalun tultua tutuksi. Näin ei kuitenkaan käynyt, vaan toisella kerralla aikaa kului myös noin 15 tuntia (871 min, minkä lisäksi indeksointiin kului 198 min). On kuitenkin todettava, että toisen hyperdokumentin laatu oli huomattavasti ensimmäistä parempi.

Vaihe	Käytetty aika
Tekstin valmistelu	45 min.
Solmujen valmistelu	155 min.
Indeksointi	-
Linkkien luominen	655 min.
Organisointi	53 min.
<b>Kokonaisaika</b>	<b>908 min.</b>

Taulukko 3. Koekonversion vaiheisiin käytetty aika.

Koekonversiomme perusteella näyttää olevan mahdollista muuntaa tekninen kirja hyperdokumentiksi yhden tai kahden työpäivän aikana. Tätä tukee myös Chignell et al. (1991) tekemä koe, jossa he muunsivat 150-sivuisen teknisen manuaalin HEFTI-0.5-järjestelmän tukemana hyperdokumentiksi. He raportoivat muunnoksessa kuluneen aikaa 649 min, josta yli kaksi kolmasosaa kului tekstin valmisteluun. Tästä puolestaan kului tekstin digitointiin noin 1/3 ja loput digitoidun tekstin virheiden korjailuun. Tekstin valmisteluun kuluva

aika tulee huomattavasti vähentymään tekstin-tunnistusteknologian kehittyessä. Muita vaiheita pystytään nopeuttamaan lähinnä käyttämällä tehokkaampia ohjelmointikieliä konversiotyökaluja kehitettäessä.

## 5. Tulokset

Tekstin konvertoimista hypertekstiksi tukevia *malleja* löytyi kirjallisuudesta muutamia. Yhtä lukuunottamatta kaikki tarkastelluista malleista ottavat kantaa ainakin jollain tavalla konversio-prosessin eri vaiheisiin lukuunottamatta linkkien hienosäätöä. Yksi malleista on keskittynyt kokonaan linkkien luomiseen ja hienosäätöön.

Teksti voidaan konvertoida hypertekstiksi joko manuaalisesti tai automaattisesti. Konvertoitavan materiaalin suuri määrä, samankaltaisten työtehtävien toistuminen ja manuaalisen konvertoinnin hitaus puoltavat automaattista konversiota. Saatavilla on useita *tukiohjelmistoja*, jotka tuottavat yleensä johonkin tiettyyn hypertekstijärjestelmään sopivia hyperdokumenteja. Ohjelmistot keskittyvät tällä hetkellä tukemaan lähinnä solmujen valmistelua, viittaavien ja assosiatiivisten linkkien luomista sekä hierarkkista järjestelyä. Ne eivät yleensä tue tekstin valmistelua ja linkkien hienosäätöä, puhumattakaan semanttisten linkkien automaattisesta generoimisesta.

Yleinen nyrkkisääntö konversiolle on, että on parempi korjata lähtötiedot yhtenäiseen muotoon kuin korjata lopputulosta. Tällä hetkellä hypertekstin konvertoimisessa on havaittavissa selvä trendi kohti hyperteksti-selaajia. Sekä kaupallisten että ei-kaupallisten yritysten kannattaa perehtyä Internetiä ja erityisesti World Wide Webiä varten kehitettyihin selaajiin.

Koekonversiossa suunnitettiin Expanded Bookin avulla Word-dokumentti HyperCard-pinoksi. Konversio vei ilman indeksointia ja linkkien hienosäätöä aikaa noin 15 tuntia. Konversiossa usein ongelmalliset kuvat siirtyivät kohtuullisella työllä, mutta linkityksessä ilmeni vakava ongelma, koska rakennettujen linkkien ankkurit eivät tekstiä muokattaessa liikkuneet tekstin mukana. Lisäksi teksti-ikkunoiden välille ei pystytty rakentamaan linkkejä. Konversion lopputuloksena oli tyydyttävästi toimiva hyperdokumentti, jonka lukeminen on kohtuullisen nopeaa ja jossa navigoiminen on Expanded Bookin ominaisuuksien ansiosta vaivatonta. Expanded Book näyttää sopivan hyvin val-

miin materiaalin muuntamiseen hyperdokumentiksi, mutta ei sen sijaan jatkuvasti päivitetävän hyperdokumentin ylläpitämiseen.

Hyperdokumenttien automaattista konversiota voidaan jo tällä hetkellä hyödyntää tehokkaasti esimerkiksi teknisen dokumentaation yhteydessä. Laajamittainen hyödyntäminen vaatinee kuitenkin yhä kehittyneempiä tukiohjelmistoja, jotta hyperdokumentin voi rakentaa olemassa olevasta elektronisesta dokumentista muutamassa minuutissa tai korkeintaan muutamassa tunnissa, ei muutamassa työpäivässä, sitä voi jakaa verkossa ja tulostaa siitä otteita vapaasti. Toisaalta hypertekstiselaajien kehittyminen voi tarjota uudenlaiset ratkaisut näihin ongelmiin.

### *Kiitokset*

Haluamme kiittää Kalervo Järveliniä ja Marjatta Okkoa artikkelia varten annetuista kommenteista sekä Pentti Kerolaa ja Juha Kämäräistä koekonversiota varten annetuistakommenteista.

Hyväksytty julkaistavaksi 17.1.1995.

## Lähteet

- Berk, E. & Devlin, J. (toim.) 1991. *Hypertext/Hypermedia Handbook*. New York: McGraw-Hill. ISBN 0-07-016622-6.
- Berners-Lee, T. & Connolly, D. 1993. *Hypertext Markup Language (HTML) A Representation of Textual Information and MetaInformation for Retrieval and Interchange*. Internet Draft: IIR Working Group.
- Berners-Lee, T., Cailliau R., Luotonen A., Nielsen H. F. & Secret A. 1994. *The World-Wide Web*. Communications of the ACM. Vol. 37, no. 8, 76–82.
- Castell, N. & Verdejo, M. F. 1991. *Automatic Extraction of Factual Information from Text and its Integration in a Knowledge Base*. In Lichnerowich, A. (toim.) *Proceedings of a Conference on Intelligent Text and Image handling 'RIAO 91'*, Barcelona, Spain 2–5 April. S.l.: Elsevier Science Publishers, 718–737.
- Catenazzi, N. & Argentisi, F. 1991. *Super-library: an Environment for the Conversion of Linear Texts into Hypertexts*. In Lichnerowich, A. (toim.) *Proceedings of a Conference on Intelligent Text and Image handling 'RIAO 91'*, Barcelona, Spain

- 2–5 April. S.I.: Elsevier Science Publishers, 45–66.
- Chignell, M. H., Nordhausen, B., Valdez, J. F. & Waterworth, J. A. 1991. The HEFTI Model of Text to Hypertext Conversion. *Hypermedia*. Vol. 3, no. 3., 187–205. ISSN 0955-8543.
- Dvorak, R., Sommerville, S. & Johnson, P. 1992. A Methodology for User Centred Link Structures for Textbook to Hypertext Conversion. In Shriver, B. D. (toim.) *Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences*. Vol. II: Software Technology, Kauai, Hawaii, January 7–10. Los Alamitos, CA: IEEE Computer Society Press, 619–628. ISBN 0-8186-2425-6.
- The Expanded Book Toolkit User's Guide 1992. Version 1.5. S.I.: The Voyager Company.
- Furuta, R., Plaisant, C. & Shneiderman, B. 1989. A Spectrum of Automatic Hypertext Constructions. *Hypermedia*. Vol. 1, no. 2., 179–195. ISSN 0955-8543.
- Gu, J. & Thiel, U. 1993. Automatically Converting Linear Text to Hypertext: A Case Study. In Frei, H. P. & Schäuble, P. (toim.) *Hypermedia. Proceedings der Internationalen Hypermedia '93 Konferenz* Zürich, 2.–3. März. Berlin: Springer-Verlag. Berlin: Springer-Verlag, 220–232. ISBN 3-540-56477-2.
- Halasz, F. & Schwartz, M. 1990. The Dexter Hypertext Reference Model. In Moline, J., Benigni, D. & Baronas, J. (toim.) *Proceedings of the Hypertext Standardization Workshop* January 16–18, National Institute of Standards and Technology. Washington: U.S. Government Printing Office, 95–134.
- Harris, M. & Cady, M. 1988. The Dynamic Process of Creating Hypertext Literature. *Educational Technology*. Vol. 28, no. 11., 33–40. ISSN 0013-1962.
- Ingwersen, P. 1992. *Information Retrieval Interaction*. Taylor Graham Publishing, Los Angeles, CA, USA. ISBN 0-947568-549.
- Joyce, M. 1991. Storyspace as a Hypertext System for Writers and Readers of Varying Ability. In *Proceedings of Hypertext '91*, San Antonio, Texas, December 15–18. New York: ACM press, 381–388. ISBN 0-89791-461-9.
- Kahn, P. 1990. Linking Together Books: Adapting Published Material into Intermedia Documents. In Delany, P. & Landow, G. (toim.) *Hypermedia and Literary Studies*. Cambridge: MIT Press, 221–256. ISBN 0-262-04119-7.
- Karlsson, F. 1993. Kielitiede. Teoksessa Hyvönen, E., Karanta I., Syrjänen M., *Tekoälyn ensyklopedia*, 47–52. Oy Gaudeamus Ab. ISBN 951-662-559-2.
- Leggett, J. & Killough, R. L. 1991. Issues in Hypertext Interchange. *Hypermedia*. Vol. 3, no. 3., 159–186. ISSN 0955-8543.
- Newcomb, S. R., Kipp, N. A. & Newcomb, V. T. 1991. The Hytime Hypermedia/Time-based Document Structuring Language. *Communications of the ACM*. Vol. 34, no. 11., 67–83. ISSN 0001-0782.
- Nunn, D., Leggett, J., Boyle, C. & Hicks, D. 1988. The REXX Project: A Case Study of Automatic Hypertext Construction. Texas A&M University, Hypertext Research Lab. TAMU 88–021.
- Oinas-Kukkonen, H. 1993. *Hyperkirjallisuus – peruskäsitteet ja historia*. Harri Oinas-Kukkonen (toimittaja): *Hyperkirjallisuus – nyt ja huomenna*. University of Oulu, Department of Information Processing Science, Working Paper Series B27, September 1993, 7–12. ISBN 951-42-3741-2.
- Priha, I. & Kujala, A. 1992. *Elektroninen tuotemanuaali. Kirjassa Tekoälyn uudet suunnat*, Vol. 2, STeP 92, Espoo 9.–11.6., Otaniemi, Finland. Helsinki: Suomen Tekoälyseura ry, 106–115. ISBN 951-96190-7-0.
- Rada, R. 1992. Converting a Textbook to Hypertext. *ACM Transactions on Information Systems*. Vol. 10, no. 3., 294–315. ISSN 1046-8188.
- Rada, R. & Diaper, D. 1991. Converting Text to Hypertext and Vice Versa. In Heather Brown (toim.) *Hypermedia/Hypertext and Object-oriented Databases*. 1st ed. London: Chapman & Hall, 167–200. ISBN 0-412-39970-9.
- Raymond, D. R. & Tompa, F. Wm. 1987. Hypertext and the New Oxford English Dictionary. In *Proceedings of the Hypertext '87 Conference*, Chapel Hill, North Carolina, November, 143–154.
- Rearick, T. C. 1991. Automating the Conversion of Text Into Hypertext. In Berk, E. & Devlin, J. (toim.) *Hypertext/Hypermedia Handbook*. New York: McGraw-Hill, 113–140. ISBN 0-07-016622-6.
- Riner, R. 1991. Automated Conversion. In Berk, E. & Devlin, J. (toim.) *Hypertext/Hypermedia Handbook*. New York: McGraw-Hill, 95–111. ISBN 0-07-016622-6.
- Salminen, A. 1992. Rakenteisen tekstin hallinta. Jyväskylän yliopisto, Tietojenkäsittelyopin laitos. Opetusmonisteita, OM-3. ISBN 951-680-877-8.
- Zheng, M. & Rada, R. 1993. SHyD – a Model for Bridging Text and Hypermedia. In *Proceedings of 1993 ACM Computer Science Conference*, Indianapolis, Indiana, February, 418–424.