

Hakukoneiden suorittaman kaupallisen tekstin- ja tiedonlouhinnan oikeudelliset haasteet Euroopan unionissa

Sofia Kangas

Alustatalous ja EU:n digitaaliset sisämarkkinat

Turun yliopiston oikeustieteellinen tiedekunta

Pro gradu -tutkielma

Maaliskuu 2020

TIIVISTELMÄ

TURUN YLIOPISTO

Oikeustieteellinen tiedekunta

SOFIA KANGAS: Hakukoneiden suorittaman kaupallisen tekstin- ja tiedonlouhinnan oikeudelliset haasteet Euroopan unionissa

Pro gradu -tutkielma, XX + 89 s.

Kauppaoikeus

Maaliskuu 2020

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin Originality Check -järjestelmällä.

Euroopan unionissa kappaleen valmistamista koskeva yksinoikeus on rakentunut analogisen maailman tarpeille. Tämä yksinoikeus on vahvasti sidottu oikeudenhaltijalle teknologianeutraalilla määritelmällä. Uudet teknologiset innovaatiot ovat kuitenkin digitalisoineet kappaleen valmistamista koskevan määritelmän. Näkymättömiä, tilapäisiä ja digitaalisia kopioita voidaan nimittäin valmistaa osana teknologista prosessia. Esimerkiksi hakukonetoiminta perustuu tekstin- ja tiedonlouhinta-tekniikkaan, joka analysoi valtavia tietomassoja uuden informaation tuottamiseksi. Osana tätä analysointitekniikkaa valmistetaan kopioita. Näiden kopioiden ja tekijänoikeuden välille on muodostunut jännite, jota on pyritty selventämään uudella tekijänoikeusdirektiivillä.

Tutkielman tarkoituksena on tarkastella, mikä on tekijänoikeusdirektiivin myötä tämän hetkinen oikeustila Euroopan unionissa hakukoneiden suorittaman kaupallisen tekstin- ja tiedonlouhinnan suhteen. Tutkimusongelmaan vastaaminen edellyttää ymmärrystä käytännön tasolla siitä, miten hakukoneet toimivat. Lisäksi on tunnistettava tekijänoikeuden ja tekstin- ja tiedonlouhinnan yhteensovittamisen haasteet, sekä mihin tekijänoikeusdirektiivin säätämistarve on perustunut.

Keskeisenä tutkimusmetodina tutkielmassa käytetään oikeusdogmaattista tutkimusotetta. Tutkimusongelman vahva liityntä teknologiaan edellyttää oikeudelliseen analyysiin teknologisen ulottuvuuden tuomista. *Law and Technology* -tutkimusmetodi mahdollistaa teknologisen ulottuvuuden hahmottamisen niin käytännöllisellä kuin yhteiskunnallisella tasolla. Tutkielman lähdemateriaali koostuu erityisesti Euroopan unionin lainsäädännöstä, oikeuskäytännöstä sekä asiantuntijalausunnoista, mutta myös merkittävä rooli on ollut tietotekniikka käsittelevällä kirjallisuudella.

Tutkielman loppupäätelminä voidaan esittää, että teknologianeutraali kappaleen valmistamista koskeva määritelmä on digitaalisessa ympäristössä vanhentunut, sekä oikeustila on epäselkeä etenkin lisensoinnin ja TPM-rajoitusten suhteen. Uusi tekijänoikeusdirektiivi sisältää tulkinnalle alttiita avoimia käsitteitä eikä lisää oikeusvarmuutta kaupalliseen tekstin- ja tiedonlouhintaan. Edellä esitetystä johtuen tutkielmassa esitetään kannanotto Euroopan unionin tekijänoikeuslainsäädännön muutostarpeesta. Osana kannanottoa tutkielmassa ehdotetaan tekijänoikeuslainsäädäntöön punnintaa perustuvaa joustavaa rajoitusnormia, joka ottaisi huomioon markkinavaikutusanalyysin ja palauttaisi tekijänoikeuden takaisin sen peruseriaatteille.

Avainsanat: digitalisaatio – EU-oikeus – koneoppiminen – kopiot – tekijänoikeus – teknologia – tekoäly – tietoyhteiskunta

SISÄLLYSLUETTELO

SISÄLLYSLUETTELO	II
LÄHTEET	IV
LYHENTEET	XVIII
1 JOHDANTO	1
1.1 Johdatus aiheeseen.....	1
1.2 Tutkimusongelma ja aiheen rajaus	4
1.3 Tutkimusmenetelmät ja käytettävät oikeuslähteet	7
1.4 Kirjallisuuskatsaus ja tutkielman tarpeellisuus.....	9
1.5 Tutkielman rakenne	11
2 TEKSTIN- JA TIEDONLOUHINTA.....	12
2.1 Määritelmä.....	12
2.2 Google-hakukoneen toiminnan logiikka.....	17
2.2.1 Yleisesti	17
2.2.2 Datan järjestäminen.....	19
2.2.2.1 Valintalajittelu	19
2.2.2.2 Pikalajittelu	21
2.2.3 Binäärihaku	24
2.2.4 Tietokannan muistirakenne ja indeksointi.....	25
2.2.5 Hajautusfunktiot	27
2.2.6 Googlen algoritmit	31
3 TEKIJÄNOIKEUS DIGITALISOITUVASSA EUROOPASSA	36
3.1 Tekijänoikeuden ja tekstin- ja tiedonlouhinnan välinen jännite.....	36
3.1.1 Tekijänoikeuden sisältö.....	36
3.1.2 Tarkempi oikeudellinen tarkastelu	41
3.2 Tekijänoikeuden digikriisi.....	42
3.2.1 Analogisesta maailmasta digitaaliseen ympäristöön.....	42
3.2.2 Kappaleen valmistaminen digitaalisessa ympäristössä	45
4 TEKIJÄNOIKEUDEN RAJOITTAMINEN	50
4.1 Yleisesti	50
4.2 Tietoyhteiskuntadirektiivin 5 artiklan 1 kohta.....	53
4.2.1 Kappaleen valmistaminen on väliaikaista tai satunnaista	53
4.2.2 Erottomaton ja välttämätön osa teknistä prosessia.....	56
4.2.3 Kappaleen valmistamisen tarkoitus.....	57

4.2.4	Ei taloudellista itsenäistä merkitystä	58
5	DIGITALISAATIOKEHITYS JA TEKIJÄNOIKEUSDIREKTIIVI	62
5.1	Digitalisoituva Eurooppa ja tekijänoikeusdirektiivi	62
5.1.1	Taustaa.....	62
5.1.2	Tekstin- ja tiedonlouhinta tieteellistä tutkimusta varten	63
5.1.2.1	Tekijänoikeusdirektiivin 3 artikla.....	63
5.1.2.2	Aineellinen ja organisatorinen soveltamisala	64
5.1.2.3	Laillinen pääsy.....	68
5.1.2.4	Turvallisuustoimenpiteet	70
5.1.3	Tekstin- ja tiedonlouhinta muuta tarkoitusta varten.....	73
5.1.3.1	Tekijänoikeusdirektiivin 4 artikla.....	73
5.1.3.2	Aineellinen ja organisatorinen soveltamisala	74
5.1.3.3	Oikeuksien pidättäminen asianmukaisella tavalla	75
6	KANNANOTTO MUUTOSTARPEESTA	77
6.1	Kannanoton rakenteesta.....	77
6.2	Kannanotolle asetettavat lähtökohdat	77
6.3	Paluu tekijänoikeuden peruseriaatteille	79
6.4	Todellinen vaikutus markkinoihin.....	82
7	YHTEENVETO	85

LÄHTEET

Artikkelit ja kirjat

Aarnio, Aulis: Laintulkinnan teoria. Yleisen oikeustieteen oppikirja. WSOY, 1989.

Cockfield, Arthur: Towards a Law and Technology Theory. *Manitoba Law Journal*, Vol. 30, s. 383–415, 2004.

Cocoru, Diana – Boehm, Mirko: An analytical review of text and data mining practices and approaches in Europe – Policy recommendations in view of the upcoming copyright legislative proposal. Openforum Europe, 2016.

Cormen, Thomas – Leiserson, Charles – Rivest, Ronald – Stein, Clifford: Introduction to Algorithms. Cambridge, Mass.: MIT Press, 3rd edition, 2009.

Depreeuw, Sari: The Variable Scope of Economic Rights in Copyright. Information law series. Wolters Kluwer Law & Business. Bryssel, Vol. 31, 2014.

Drassinower, Abraham: What is wrong with copying? Harvard University Press, 2015.

Ducato, Rossana – Strowel, Alain: Limitations to Text and Data Mining and Consumer Empowerment Making the Case for a Right to "Machine Legibility". CRIDES Working Paper Series, 2018.

Ezrachi, Ariel – Stucke, Maurice: Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy. Harvard University Press, 2016.

Filippov, Sergey – Hofheinz, Paul: Text and Data Mining for Research and Innovation. The Lisbon Council – think tank for the 21st century. Issue 20/2016.

Geiger, Christophe: Copyright's Fundamental Rights Dimension at EU Level. Estelle Derclayen editoimassa teoksessa: Research Handbook on the Future of EU Copyrights, Cheltenham, 2009.

Geiger, Christophe – Griffiths, Jonathan – Hilty, Reto: Towards a balanced interpretation of the "three-step test" in copyright law. Julkaisussa: EIPR: European Intellectual Property Review Dec, Vol.30(12), s. 489–496, 2008.

Geiger, Christophe – Frosio, Giancarlo – Bulayenko, Oleksandr: The exception for text and data mining (TDM) in the proposed directive on copyright in the digital single market - legal aspects. Centre for International Intellectual Property Studies, Research Paper No. 2018–02.

Han, Jiawei – Kamber, Micheline – Pei, Jian: Data mining: Concepts and Techniques. The Morgan Kaufmann Series in Data Management Systems. Ed. 3, 2012.

Hirvonen, Ari: Mitkä metodit? Opas oikeustieteen metodologiaan. Helsingin yliopisto, yleisen oikeustieteen julkaisuja, 2011.

Hugenholtz, Bernt: Adapting copyright to the information superhighway. The Hague: Kluwer Law International, 1996.

Hugenholtz, Bernt – Okedijii, Ruth: Conceiving an International Instrument on Limitations and Exceptions to Copyright. Amsterdam Law School Legal Studies Research Paper number 2012–43, 2012.

Hugenholtz, Bernt: Flexible Copyright: Can EU Author's Right Accommodate Fair Use? Ruth Okedijiin editoimassa teoksessa: Copyright Law in an Age of Limitations and Exceptions. Cambridge University Press, s. 275–291, 2017.

Hughes, Justin: The Philosophy of Intellectual Property. Georgetown University Law Center and Georgetown Law Journal, 77 Geo. L.J. 287, 1998.

Husa, Jaakko: Oikeusvertailu. Helsingin Kamari Oy/Helsingin seudun kauppakamari ja tekijä, 2013.

Husovec, Martin: Injunctions against Intermediaries in the European Union – Accountable but Not Liable? Cambridge University Press, 2017.

Klein, Susanne: Search Engines and Copyright – An Analysis of the Belgian Copiepresse Decision in Consideration of British and German Copyright Law. International Review of Intellectual Property and Competition Law 4/2008, s. 451–483, 2008.

Leaffer, Marshall A.: Understanding Copyright Law. Lexis Nexis. Edition 6, 2014.

Levendowski, Amanda: How copyright law can fix artificialIntelligence’s implicit bias problem. Washington law review Vol. 93:579, 2018.

Locke, John: Second Treatise of Government. Arlington Heights (Ill.): Harlan Davidson, cop. 1982.

Ohisalo, Juho: Tietokannan sisältämän tiedon suojaaminen sui generis -oikeudella ja sopimusmenettelyllä, s. 74–101. Teoksessa: Immateriaalioikeudet ja yleinen etu. Alma Talent, painos 1, 2017.

Ohly, Ansgar: Economic rights, s. 212–243. Teoksessa: Research Handbook on the Future of EU Copyright, Edward Elgar, Editor Estelle Derclaye, 2009.

Ohly, Ansgar: A Fairness-Based Approach to Economic Rights, s. 83–119. Teoksessa: Information law series: Copyright Reconstructed – Rethinking Copyright’s Economic Rights in a Time of Highly Dynamic Technological and Economic Change. Wolters Kluwer, Editor P.Bernt Hugenholtz. Volume 41, 2018.

Patton, Desmond Upton – Rambow, Owen – Auerbach, Jonathan – Li, Kevin – Frey, William: Expressions of loss predict aggressive comments on Twitter among gang-involved youth in Chicago. npj Digital Medicine volume 1, Article number: 11, 2018.

Peguera, Miquel: Copyright Issues Regarding Google Images and Google Cache, s. 169–202. Teoksessa: Google and the Law Empirical Approaches to Legal Aspects of Knowledge-Economy Business Models, TMC ASSER PRESS Springer, Editor Aurelio Lopez-Tarruella, 2012.

Pila, Justine – Torremans, Paul: European intellectual property law. Oxford University Press, toinen painos, 2019.

Poort, Joost: Reconstructing the Reproduction and Communication to the Public Rights: How to Align Copyright with Its Fundamentals, s. 283–332 Teoksessa: Information law series: Copyright Reconstructed – Rethinking Copyright’s Economic Rights in a Time of Highly Dynamic Technological and Economic Change. Wolters Kluwer, Editor P.Bernt Hugenholtz. Volume 41, 2018.

Price, Monroe E.: Reexamining Intellectual Property Concepts: A Glimpse Into the Future Trough the Prism of Cahakarbarty. Cardozo Arts & Entertainment Law Journal. Volume 6, number 2, 1998.

Quintais, João Pedro: Rethinking Normal Exploitation: Enabling Online Limitations in EU Copyright Law. AMI: Tijdschrift voor Auteurs-, Media- & Informatierecht, 41(6), s. 197–205, 2017.

Rosati, Eleonora: Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and Its Role in the Development of AI Creativity. *Asia Pacific Law Review*, 2019.

Sag, Matthew: The New Legal Landscape for Text Mining and Machine Learning. *Journal of the Copyright Society of the USA*, Vol. 66 p.291, 2019.

Salokannel, Marjut: Tekijänoikeus ja tutkimuksen raaka-aineet – Uuden teknologian haasteet tekijänoikeudellisesti suojatun materiaalin tutkimuskäytölle. *Koneen säätiö. Mynäprint, Mynämäki*, 2013.

Samuelson, Pamela: Does Copyright Law Need to Be Reformed? *Communications of the ACM*. October 2007/Vol. 50, No. 10.

Schindelhauer, Christian: Hashing, s. 195–203. Teoksessa: *Algorithms Unplugged*. Springer, Heidelberg. Editors Vöcking, Berthold – Alt, Helmut – Dietzfelbinger, Martin – Reischuk, Rüdiger – Scheideler, Christian – Vollmer, Heribert – Wagner, Dorothea. 2011.

Siltala, Raimo: Oikeustieteen tieteenteoria. Suomalaisen lakimiesyhdistyksen julkaisuja Asarja N:o 234. Vammala, 2003.

Spraul, Anton: *How Software Works: The Magic Behind the Encryption, CGI, Search Engines and Other Everyday Technologies*. No starch press, 2015.

Strowel, Alain: *Droit d'auteur et copyright, Divergences et convergences*. Bruylant and Paris, Bryssel, 1993.

Strowel, Alain – Derclaye, Estelle: *Droit d'auteur et numérique: logiciels, bases de données, multimédia: droit belge, européen et compare*. Bruylant, 2001.

Strowel, Alain: *Droit d'ateur et copyright. Convergences des droits, regulation deffërente des contrats Mélanges en l'honneur dy Professeur André Lucas*. LexisNexis, s. 699–717. Paris, 2014.

Strowel, Alain: *Reconstructing the Reproduction and Communication to the Public Rights: How to Align Copyright with Its Fundamentals*. Teoksessa: *Information law series: Copyright Reconstructed – Rethinking Copyright's Economic Rights in a Time of Highly Dynamic Technological and Economic Change*. Wolters Kluwer, Editor P.Bernt Hugenholtz. Volume 41, 2018.

Tusikov, Natasha: Chokepoints: Global Private Regulation on the Internet. University of California Press. 1st ed, 2016.

VerSteeg, Russ: The Roman Law Roots of Copyright. Maryland Law Review. Volume 59, Issue 2, 2000.

Vilanka, Olli: Linkittäminen, tekijänoikeus ja verkkojulkaiseminen. Lakimies 4/2006, s. 608–627.

Virallislähteet

Euroopan komission lehdistötiedote: Tekijänoikeusuudistukselle lopullinen hyväksyntä: Komissio tyytyväinen sopuun uusista digiajan tarpeisiin soveltuvista säännöistä. Luxemburg 15. huhtikuuta 2019.

Euroopan komission tutkimuksen ja innovoinnin pääosaston asiantuntijalausunto aiheesta ”Standardisation in the area of innovation and technological development, notably in the field of text and data mining”, 7.11.2014. (*”Euroopan komission tutkimuksen ja innovoinnin pääosaston asiantuntijalausunto 2014”*).

Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat,: The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market – Legal Aspects. PE 604.941, 2018. (*”Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: oikeudellinen analyysi 2018”*).

Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat,: The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market – Technical Aspects. PE 604.942, 2018. (*”Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: tekninen analyysi 2018”*).

Euroopan parlamentin tarkistukset 12. syyskuuta 2018 ehdotukseen Euroopan parlamentin ja neuvoston direktiiviksi tekijänoikeudesta digitaalisilla sisämarkkinoilla (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)). P8_TA(2018)0337 Tekijänoikeus digitaalisilla sisämarkkinoilla ***I. (*”Euroopan parlamentin muutosehdotus tekijänoikeusdirektiiviiluonnokseen 2018”*).

Euroopan yhteisöjen komissio: Vihreä kirja Tekijänoikeus ja lähioikeudet tietoyhteiskunnassa. Bryssel 19.7.1995. KOM(95) 382 lopull. (*”Euroopan yhteisöjen komissio – Vihreä kirja 1995”*).

European Copyright Society: General Opinion on the EU Copyright Reform Package. 24.1.2017.

Komission tiedonanto Euroopan parlamentille, neuvostolle, Euroopan talous- ja sosiaaliskomitealle ja alueiden komitealle. Digitaalisten sisämarkkinoiden strategia Euroopalle. COM (2015) 192, Bryssel 6.5.2015.

Komission tiedonanto - Tekijänoikeutta ja lähioikeuksia tietoyhteiskunnassa koskevan vihreän kirjan seuranta KOM/96/0568.

Kulttuuri- ja koulutusvaliokunnan lausunto oikeudellisten asioiden valiokunnalle ehdotuksesta Euroopan parlamentin ja neuvoston direktiiviksi tekijänoikeudesta digitaalisilla sisämarkkinoilla (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)). 4.9.2017.

Mietintö ehdotuksesta Euroopan parlamentin ja neuvoston direktiiviksi tekijänoikeudesta digitaalisilla sisämarkkinoilla (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)) Oikeudellisten asioiden valiokunta, 29.6.2018.

Sisämarkkina- ja kuluttajansuojavaliokunnan lausunto oikeudellisten asioiden valiokunnalle ehdotuksesta Euroopan parlamentin ja neuvoston direktiiviksi tekijänoikeudesta digitaalisilla sisämarkkinoilla (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)). 14.6.2017.

Talous- ja sosiaaliskomitean lausunto aiheesta: ”Ehdotus: Euroopan parlamentin ja neuvoston direktiivi tekijänoikeuden ja lähioikeuksien tiettyjen piirteiden yhdenmukaistamisesta tietoyhteiskunnassa”(1) (98/C 407/06)”.

Teollisuus-, tutkimus- ja energiavaliokunnan lausunto oikeudellisten asioiden valiokunnalle ehdotuksesta Euroopan parlamentin ja neuvoston direktiiviksi tekijänoikeudesta digitaalisilla sisämarkkinoilla (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)). 1.8.2017.

Triaille, Jean Paul – de Meeûs d’Argenteuil, Jerome – de Francquen, Amélie: Study on the legal framework of text and data mining (TDM). Euroopan Komission rahoittama tutkimus. Marraskuu 2014.

Internetlähteet

Borghi, Maurizio: "Text & Data mining". <https://www.copyrightuser.org/understand/exceptions/text-data-mining/>. Luettu 11.7.2019.

Corporate Europe Observatory: "Copyright Directive: how competing big business lobbies drowned out critical voices", 10.12.2018. <https://corporateeurope.org/en/2018/12/copyright-directive-how-competing-big-business-lobbies-drowned-out-critical-voices>. Luettu 15.2.2020.

The Economist: "The world's most valuable resource is no longer oil, but data", 6.5.2017. <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>. Luettu 19.9.2019.

European Alliance for Research Excellence I: "Global AI-Race". <http://eare.eu/assets/uploads/2018/06/Global-AI-Race.pdf>. Luettu 12.1.2019.

European Alliance for Research Excellence II: "Japan amends its copyright legislation to meet future demands in AI and Big Data", 3.9.2018. <http://eare.eu/japan-amends-tdm-exception-copyright/>. Luettu 12.1.2019.

European Alliance for Research Excellence's statement: "The Copyright reform must also be fair to European researchers and entrepreneurs", 30.8.2018. http://eare.eu/assets/uploads/2018/09/Statement-on-Copyright-TDM-from-European-Research-and-Innovation-community_30-Aug.pdf. Luettu 12.1.2019.

Forbes: "Here's Why Data Is Not The New Oil", 5.11.2018. <https://www.forbes.com/sites/bernardmarr/2018/03/05/heres-why-data-is-not-the-new-oil/#406f113d3aa9>. Luettu 19.9.2019.

Geeks for Geeks: "Machine Learning". <https://www.geeksforgeeks.org/getting-started-machine-learning/>. Luettu 27.2.2020.

Geeks for Geeks: "Searching algorithms". <https://www.geeksforgeeks.org/searching-algorithms/>. Luettu 3.1.2019.

Geeks for Geeks: "Quick sort". <https://www.geeksforgeeks.org/quick-sort/>. Luettu 31.12.2019.

Google blog: "Our new search index: Caffeine", 8.6.2010. <https://googleblog.blogspot.com/2010/06/our-new-search-index-caffeine.html>. Luettu 19.1.2020.

Google Haku: ”Miten Haun algororitmit toimivat”.
<https://www.google.com/intl/fi/search/howsearchworks/algorithms/>. Luettu 19.1.2020.

Google Haku: ”Miten Haku järjestää tietoa”.
<https://www.google.com/intl/fi/search/howsearchworks/crawling-indexing/>. Luettu 19.1.2020.

Google-hakukoneen hakutulos hakusanalle: "text and data mining".
https://www.google.com/search?rlz=1C1GCEB_enSE851FI852&ei=1qMIXqDeBoS-wkwWFxbagBA&q=text+and+data+mining&oq=text+and+data+mining&gs_l=psy-ab.3..015j0i22i3015.75934.75934..77043...0.1..0.85.85.1.....0....2j1..gws-wiz.....0i71.RvVBkEQvF6c&ved=0ahUKEwigoce5nZLnAhUE2KQKHYWiDUQQ4dUD-CAs&uact=5&safe=active&ssui=on. Luettu 20.1.2020.

Googlen kotisivut: <https://www.google.com/intl/en-GB/about/>. Luettu 1.12.2019.

Google Support: ”Vanhentuneen sisällön poistotyökalu”. <https://support.google.com/webmasters/answer/7041154?hl=fi>. Luettu 5.2.2020.

The Guardian: ”Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach”, 17.3.2018. <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>. Luettu 1.1.2020.

Hardesty, Larry: ”Automatic image retouching on your phone”, 1.8.2017.
<http://news.mit.edu/2017/automatic-image-retouching-phone-0802>. Luettu 15.12.2019.

IBM Marketing Cloud: ”10 Key Marketing Trends for 2017 and Ideas for Exceeding Customer Expectations”, 2017. <http://anyflip.com/vdoi/wpza/basic>. Luettu 19.1.2020.

Internetlivestats: <https://www.internetlivestats.com/google-search-statistics/>. Luettu 19.1.2020.

Jisc Report: ”Value and benefits of text mining”, 3.1.2020. <https://www.jisc.ac.uk/reports/value-and-benefits-of-text-mining>. Luettu 9.3.2020.

Kersting, Kristian: ”Machine Learning and Artificial Intelligence: Two Fellow Travelers on the Quest for Intelligent Behavior in Machines”, 19.11.2018. <https://www.frontiersin.org/articles/10.3389/fdata.2018.00006/full>. Luettu 27.2.2020.

Kretschmer, Martin – Margoni, Thomas: “Data mining: why the EU’s proposed copyright measures get it wrong”, 24.5.2018. <https://theconversation.com/data-mining-why-the-eus-proposed-copyright-measures-get-it-wrong-96743>. Luettu 22.10.2018. (*Kretschmer – Margoni 2018a*)

Kretschmer, Martin – Margoni, Thomas: “The Text and Data Mining exception in the Proposal for a Directive on Copyright in the Digital Single Market: Why it is not what EU copyright law needs”, 25.5.2018. <https://www.create.ac.uk/blog/2018/04/25/why-tdm-exception-copyright-directive-digital-single-market-not-what-eu-copyright-needs/>. Luettu 22.10.2018. (*Kretschmer – Margoni 2018b*)

Market Insider: ”The 5 most valuable US tech companies are now worth more than \$5 trillion after Alphabet’s record close”, 17.1.2020. <https://markets.businessinsider.com/news/stocks/most-valuable-tech-companies-total-worth-trillions-alphabet-stock-record-2020-1-1028826533>. Luettu 19.1.2020.

Marsden, Sam: “Search Engine Indexing”, 17.5.2018. <https://www.deepcrawl.com/knowledge/technical-seo-library/search-engine-indexing/>. Luettu 16.1.2020.

Mercury News: "Dictionary adds verb: to google", 6.7.2006. <https://web.archive.org/web/20070206065348/http://www.mercurynews.com/mld/mercurynews/business/14985574.htm>. Luettu 12.1.2019.

Michiganin yliopiston ja Googlen välinen sopimus koskien kirjojen kopioimista: <https://www.lib.umich.edu/sites/default/files/services/mdp/um-google-cooperative-agreement.pdf>. Luettu 28.12.2019.

Nevala, Meri-Tuuli: “Miten Googlen hakukone toimii?”, 21.3.2017. <https://www.havain.fi/miten-googlen-haku-toimii/>. Luettu 10.1.2020.

Niechai, Valerie: “The Vice and Virtue of PageRank (2019 edition)”, 30.4.2019. <https://www.link-assistant.com/news/google-page-rank-2019.html>. Luettu 24.1.2020.

Niemi, Riikka – Kallio, Laura – Rissanen, Jaakko: ”Miten hakukoneet toimivat”, 4.12.2000. <https://www.netlab.tkk.fi/opetus/s38118/s00/tyot/42/>. Luettu 1.3.2020.

The Official Google Search blog: “Giving you fresher, more recent search results”, 3.11.2011. <https://search.googleblog.com/2011/11/giving-you-fresher-more-recent-search.html>. Luettu 19.1.2020.

OpenForum Academy: ”How to unleash the innovative potential of text and data mining in the EU”, 22.10.2015. <http://www.openforumeurope.org/wp-content/uploads/2015/12/White-Paper-TDM-1.pdf>. Luettu 1.9.2019.

Oxford Dictionary: https://www.oxfordlearnersdictionaries.com/definition/american_english/google. Luettu 12.1.2019.

Samuelson, Pamela: “The EU’s Controversial Digital Single Market Directive – Part II: Why the Proposed Mandatory Text- and Data-Mining Exception Is Too Restrictive”, 12.7.2018. <http://copyrightblog.kluweriplaw.com/2018/07/12/eus-controversial-digital-single-market-directive-part-ii-proposed-mandatory-text-data-mining-exception-restrictive/>. Luettu 19.1.2020.

Towards Data Science: ”Data is Not the New Oil”, 11.2.2019. <https://towardsdatascience.com/data-is-not-the-new-oil-721f5109851b>. Luettu 19.9.2019.

The United States Copyright Office: “More Information on Fair Use”, 9.2019. <https://www.copyright.gov/fair-use/more-info.html>. Luettu 26.12.2019.

VentureBeat: ”Data is not the new oil”, 10.2.2019. <https://venturebeat.com/2019/02/10/data-is-not-the-new-oil/>. Luettu 19.9.2019.

Webopedia: ”Proxy cache”, https://www.webopedia.com/TERM/P/proxy_cache.html. Luettu 5.2.2020.

WIPO: ”Understanding Copyright and Related Rights”, 2016. https://www.wipo.int/edocs/pubdocs/en/wipo_pub_909_2016.pdf. Luettu 12.10.2019.

Zeybek, Bengi: ”Text and Data Mining in the Proposed Directive: Where do we stand”, 23.3.2018. <http://copyrightblog.kluweriplaw.com/2018/03/23/text-data-mining-proposed-directive-stand/>. Luettu 26.10.2019.

Äänitallenteet

Hugenholtz, Bernt: Owing Science: “Intellectual Property Rights as Impediments to Knowledge Sharing”. Communia konferenssissa Italiassa Turinissa 29.6.2001. Äänite saatavilla osoitteesta <https://communia-project.eu/node/289.html>. Kuunneltu 14.12.2019.

Oikeuskäytäntö

Euroopan unionin tuomioistuin

C-476/01, Yhteisöjen tuomioistuimen tuomio (viides jaosto), 29. huhtikuuta 2004 (“*Kapper*”). ECLI:EU:C:2004:261.

C-46/02, Yhteisöjen tuomioistuimen tuomio (suuri jaosto), 9. marraskuuta 2004 (“*Fixtures v Veikkaus*”). ECLI:EU:C:2004:694.

C-203/02, Yhteisöjen tuomioistuimen tuomio (suuri jaosto), 9. marraskuuta 2004 (“*The British Horseracing Board*”). ECLI:EU:C:2004:695.

C-338/02, Yhteisöjen tuomioistuimen tuomio (suuri jaosto), 9. marraskuuta 2004 (“*Fixtures v Svenska Spel*”). ECLI:EU:C:2004:696.

C-444/02, Yhteisöjen tuomioistuimen tuomio (suuri jaosto), 9. marraskuuta 2004 (“*Fixtures Marketing*”). ECLI:EU:C:2004:697.

C-36/05, Yhteisöjen tuomioistuimen tuomio (kolmas jaosto), 26. lokakuuta 2006 (“*Komissio v Espanja*”). ECLI:EU:C:2006:672.

C-5/08, Yhteisöjen tuomioistuimen tuomio (neljäs jaosto), 16. heinäkuuta 2009 (“*Infopaq I*”). ECLI:EU:C:2009:465.

Yhdistetyt asiat C-236/08 ja C-238/08, Unionin tuomioistuimen tuomio (suuri jaosto), 23. maaliskuuta 2010 (“*Google France ja Google*”). ECLI:EU:C:2010:159.

C-467/08, Unionin tuomioistuimen tuomio (kolmas jaosto), 21. lokakuuta 2010 (“*Padawan*”). ECLI:EU:C:2010:620.

Yhdistetyt asiat C-403/08 ja C-429/08, Unionin tuomioistuimen tuomio (suuri jaosto), 4. lokakuuta 2011 (“*Premier League*”). ECLI:EU:C:2011:631.

C-145/10, Unionin tuomioistuimen tuomio (kolmas jaosto), 1. joulukuuta 2011, (*"Painer"*). ECLI:EU:C:2011:798.

C-302/10, Unionin tuomioistuimen määräys (kolmas jaosto), 17. tammikuuta 2012 (*"Infopaq II"*). ECLI:EU:C:2012:16.

C-604/10, Unionin tuomioistuimen määräys (kolmas jaosto), 1. maaliskuuta 2012 (*"Football Dataco"*). ECLI:EU:C:2012:115.

C-314/12, Unionin tuomioistuimen tuomio (neljäs jaosto), 27. maaliskuuta 2014 (*"Public UPC Telekabel Wien"*). ECLI:EU:C:2014:192.

C-360/13, Unionin tuomioistuimen tuomio (neljäs jaosto), 5. kesäkuuta 2014 (*"Public Relations Consultants Association"*). ECLI:EU:C:2014:1195.

C-201/13, Unionin tuomioistuimen tuomio (suuri jaosto), 3. syyskuuta 2014 (*"Deckmyn"*). ECLI:EU:C:2014:2132.

C-30/14, Unionin tuomioistuimen tuomio (toinen jaosto), 15. tammikuuta 2015 (*"Ryanair"*). ECLI:EU:C:2015:10.

C-419/13, Unionin tuomioistuimen tuomio (neljäs jaosto), 22. tammikuuta 2015 (*"Art & All-posters"*). ECLI:EU:C:2015:27.

C-490/14, Unionin tuomioistuimen tuomio (toinen jaosto), 29. lokakuuta 2015 (*"Verlag Esterbauer"*). ECLI:EU:C:2015:735.

C-301/15, Unionin tuomioistuimen tuomio (kolmas jaosto), 16. marraskuuta 2016 (*"Soulier ja Doke"*). ECLI:EU:C:2016:878.

C-161/17, Unionin tuomioistuimen tuomio (toinen jaosto), 7. elokuuta 2018 (*"Renckhoff"*). ECLI:EU:C:2018:634.

C-310/17, Unionin tuomioistuimen tuomio (suuri jaosto), 13. marraskuuta 2018 (*"Levola Hengelo"*). ECLI:EU:C:2018:899.

C-469/17, Unionin tuomioistuimen tuomio (suuri jaosto), 29. heinäkuuta 2019 (*"Funke Medien NRW"*). ECLI:EU:C:2019:623.

C-476/17, Unionin tuomioistuimen tuomio (suuri jaosto), 29. heinäkuuta 2019 (”*Pelham*”).
ECLI:EU:C:2019:624.

C-516/17, Unionin tuomioistuimen tuomio (suuri jaosto), 29. heinäkuuta 2019 (”*Spiegel Online*”). ECLI:EU:C:2019:625.

Euroopan unionin tuomioistuimen julkisasiamiehen ratkaisuehdotukset

Szpunar, Maciej: asiassa C-469/17 *Funke Medien NRW*. ECLI:EU:C:2018:870.

Szpunar, Maciej: asiassa C-476/17 *Pelham*. ECLI:EU:C:2018:1002.

Belgialainen tuomioistuin

Ratkaisu 2007/AR/1730, *Google Inc. v. Copiepresse*.

Ratkaisu *Google Inc. v. Copiepresse, Cours d’ Appel (CA)/Hoven van Beroep (HvB) (Courts of Appeals) Bruxelles*, 9e ch. May 26, 2011, R No. 2011/2999 No. 817.

Ranskalainen tuomioistuin

Kassaatio tuomioistuin, täysistunto, 30. lokakuu 1987, 86-11.918.

Yhdysvaltalaiset tuomioistuimet

Ratkaisu *Harper & Row v. Nation Enterprises*, 471 U.S. 539 (1985).

Ratkaisu *Campbell v. Acuff-Rose Music, Inc.* No. 92-1292, 510 U.S. 569 (1994).

Ratkaisu *Kelly v. Arriba Soft Corp.*, 336 F.3d 811 (9th Cir. 2003).

Ratkaisu *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146 (9th Cir. 2007).

Ratkaisu *A.V. ex rel. Vanderhuy v. iParadigms, LLC*, 562 F.3d 630 (4th Cir.2009).

Ratkaisu *Authors Guild, Inc. v. HathiTrust*, 902 F. Supp. 2d 445, 460 (S.D.N.Y. 2012).

Ratkaisu *Authors Guild, Inc. v. Google, Inc.*, 954 F. Supp. 2d 282, 291 (S.D.N.Y. 2013).

Ratkaisu *Authors Guild v. HathiTrust*, 755 F.3d 87 (2nd Cir. 2014).

Ratkaisu *Authors Guild v. Google Inc.* No. 13–4829–cv, 804 F.3d 202 (2nd Circ. 16. lokakuuta 2015).

LYHENTEET

Bernin yleissopimus	Bernin yleissopimus kirjallisten ja taiteellisten teosten suojaamisesta (1886/1971)
BLFD	Biased, low-friction data
CULT	The Culture and Education Committee, Kulttuuri- ja koulutusvaliokunta
Digitaalisten sisämarkkinoiden strategia Euroopalle	Komission tiedonanto Euroopan parlamentille, neuvostolle, Euroopan talous- ja sosiaalikomitealle ja alueiden komitealle. Digitaalisten sisämarkkinoiden strategia Euroopalle. COM (2015) 192, Bryssel 6.5.2015
DRM	Digital Rights Management
EU	Euroopan unioni
Euroopan parlamentin muutosehdotus tekijänoikeusdirektiiviiluonnokseen	Euroopan parlamentin tarkistukset 12. syyskuuta 2018 ehdotukseen Euroopan parlamentin ja neuvoston direktiiviksi tekijänoikeudesta digitaalisilla sisämarkkinoilla (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)). P8_TA(2018)0337 Tekijänoikeus digitaalisilla sisämarkkinoilla ***I
Euroopan yhteisöjen komissio – Vihreä kirja	Euroopan yhteisöjen komissio: Vihreä kirja Tekijänoikeus ja lähioikeudet tietoyhteiskunnassa. Bryssel 19.7.1995. KOM(95) 382 lopull.
IMCO	The Committee on the Internal Market and Consumer Protection, Sisämarkkina- ja kuluttajansuojavaliokunta

ITRE	The Industry, Research and Energy Committee, Teollisuus-, tutkimus- ja energia-valiokunta
JURI	The Committee on Legal Affairs, Oikeudellisten asioiden valiokunta
Rooman yleissopimus	Rooman yleissopimus esittävien taiteilijoiden, äänitteiden valmistajien sekä radioyhtymien suojaamisesta (1961)
TDM	Text and data mining, tekstin- ja tiedonlouhinta
Tekijänoikeusdirektiivi	Euroopan parlamentin ja neuvoston direktiivi (EU) 2019/790, annettu 17 päivänä huhtikuuta 2019, tekijänoikeudesta ja lähioikeuksista digitaalisilla sisämarkkinoilla ja direktiivien 96/9/EY ja 2001/29/EY muuttamisesta
Tietokantadirektiivi	Euroopan parlamentin ja neuvoston direktiivi 96/9/EY, annettu 11 päivänä maaliskuuta 1996, tietokantojen oikeudellisesta suojasta
Tietoyhteiskuntadirektiivi	Euroopan parlamentin ja neuvoston direktiivi 2001/29/EY, annettu 22 päivänä toukokuuta 2001, tekijänoikeuden ja lähioikeuksien tiettyjen piirteiden yhdenmukaistamisesta tietoyhteiskunnassa
TPM	Technological Protection Measures
TRIPS-sopimus	Agreement on Trade-Related Aspects of Intellectual Property Rights, Sopimus teollis- ja tekijänoikeuksien kauppaan liittyvistä näkökohdista
UNESCO	United Nations Educational, Scientific and Cultural Organization, Yhdistyneiden

	Kansakuntien kasvatus-, tiede- ja kulttuurijärjestö
WCT	WIPO Copyright Treaty, Maailman henkisen omaisuuden järjestön tekijänoikeussopimus
WIPO	World Intellectual Property Organization, Maailman henkisen omaisuuden järjestö
WPPT	WIPO Performances and Phonograms Treaty, Maailman henkisen omaisuuden järjestön esitys- ja äänitesopimus
WTO	World Trade Organization, Maailman kauppajärjestö
Yleinen tietosuoja-asetus	Euroopan parlamentin ja neuvoston asetukset (EU) 2016/679, annettu 27 päivänä huhtikuuta 2016, luonnollisten henkilöiden suojelusta henkilötietojen käsittelyssä sekä näiden tietojen vapaasta liikkuvuudesta ja direktiivin 95/46/EY kumoamisesta

1 JOHDANTO

1.1 Johdatus aiheeseen

Miten maailmassa, jossa datan väitetään olevan arvokkaampaa kuin öljyn, tulisi säännellä datan ympärille rakentuvia teknologioita ja ilmiöitä?¹ Digitalisoituneen yhteiskunnan uudet teknologiat ja ilmiöt haastavat voimassa olevia lainsäädäntömalleja muuttumaan ja tekemään tilaa innovaatiolle. Tekoäly, talouskasvu, tiede ja kehitys ajavat sekä nykyisten lainsäädäntömallien että oppirakennelmien muuttamista. Samalla on kuitenkin katsottava pöydän toiselle puolelle ja huomattava vanhojen lainsäädäntömallien ja oppirakennelmien taustalla olevat periaatteet sekä ne arvot, joita oikeusnormien on ajateltu suojelevan.

Uusiin innovaatioihin liittyy myös riskejä. Esimerkiksi *Cambridge Analytica* -skandaali opetti, miten meistä kaiken aikaa kerättävää tietoa voidaan käyttää väärin poliittisiin tarkoituksiin.² Teknologiselle kehitykselle yhteiskunnassa asetettavat rajoitukset ovat perusteltavissa niin yksityisyydensuojalla, elinkeinovapaudella kuin itsemääräämisoikeudella. Teknologioiden haasteiden ja etujen määrittely ei kuitenkaan ole yksiselitteistä.

Cambridge Analytica -skandaalissa käytettiin nimittäin samaa analysointiteknologiaa, jota sovelletaan myös yhteiskuntaa hyödyttävillä tavoilla, kuten lääketieteen tutkimuksissa, liikenne-ruuhkien ennustamisessa sekä internetin mullistaneen hakukoneen toiminnassa. Tätä teknologiaa kutsutaan nimeltä tekstin- ja tiedonlouhinta ("TDM").³ Näin ollen ei voida mustavalkoisesti väittää jonkin teknologian olevan joko huonoa tai hyvää. Vaaditaan punnintaa innovaatioiden riskeistä ja eduista, minkä lisäksi tarvitaan joustavia lainsäädäntömalleja. Tämä tutkielma nostaa pinnalle yhden näistä datan ympärille kehittyneistä ja voimassa olevaa lainsäädäntöä

¹ *The Economist* 2017. Ilmaisun "Data is valuable than oil" tarkkaa alkuperää ei ole tiedossa, mutta väitetään, että brittiläinen matemaatikko *Clive Humby* olisi ollut vuonna 2006 ensimmäisiä henkilöitä, joka on käyttänyt kyseistä ilmaisua. Ilmaisun on kuitenkin herättänyt myös eriäviä mielipiteitä, joiden mukaan dataa ei voida suoraan rinnastaa öljyyn. Eroja löytyy muun muassa siitä, miten näitä kerätään, säilötään, hyödynnetään sekä minkälaisia resursseja näihin kauppatavaroihin on sidottuina. Katso lisää aiheesta *Forbes* 2018, *VentureBeat* 2019 sekä *Towards Data Science* 2019.

² Yhdysvaltain republikaanien poliittisessa kampanjassa kerättiin Facebook-profiileista valtavat määrät tietoa, joiden perusteella laadittiin psykologisia profiileja, joita hyödynnettiin presidenttiehdokkaan kampanjassa. Facebook-profiilien analysointi perustui myös tekstin- ja tiedonlouhinteknologiaan. Ks. lisää *The Guardian* 2018, osoitteesta: <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>.

³ *Kretschmer – Margoni* 2018a.

haastavista tekstin- ja tiedonlouhintateknologioista, nimittäin hakukonetoiminnan, ja tutkii, miten sitä on lähestytty oikeudellisesta näkökulmasta.

Hakukoneet mahdollistavat pääsyn lähes rajattomaan määrään tietoa ja ideoita, minkä vuoksi esimerkiksi *googlaamisesta* on muodostunut olennainen osa arkeamme.⁴ Googlen hakukone kuvastaa hyvin sitä arvoa, joka datalla on nykypäivänä. Vaikkei Google omista perinteisiä suuria teollisia tehtaita taikka massatuotantoa, on se kiistatta yksi maailman arvokkaimmista yrityksistä.⁵ Googlen hakutoiminnon avulla haetaan tietoa miljardeja kertoja päivittäin.⁶ Yhtiön arvo perustuu kehittyneeseen tekstin- ja tiedonlouhintateknologiaan, joka mahdollistaa hakukoneen käsittelemään yksittäisen hakusanan avulla valtavia datamassoja sekuntien murto-osissa.⁷

Termi tekstin- ja tiedonlouhinta on määritelty tuoreen tekijänoikeusdirektiivin⁸ 1 artiklan 2 kohdassa, jonka mukaan tekstin- ja tiedonlouhinnalla tarkoitetaan kaikkia automaattisia analysointimenetelmiä, joilla analysoidaan digitaalimuotoisia teoksia ja muuta aineistoa tiedon, muun muassa mallien, suuntausten tai korrelaatioiden tuottamiseksi. Analysoinnin suorittamiseksi osana tekstin- ja tiedonlouhintateknologiaa tehdään teknisiä kopioita⁹ analysoitavasta materiaalista.

⁴ Googlen käyttäminen ja yleisyys ovat johtaneet siihen, että verbi ”*googlata*” lisättiin Oxfordin englanninkieliseen sivistyssanakirjaan 15. kesäkuuta 2016. Ks. *Mercury News* -arkistosta uutinen osoitteesta: <https://web.archive.org/web/20070206065348/http://www.mercurynews.com/mld/mercurynews/business/14985574.htm>.

⁵ Googlen emoyhtiö Alphabetin arvo nousi 16.1.2020 yli biljoonaan dollariin. Ks. lisää *Market Insider* 2020.

⁶ Noin 3,5 miljardia hakua tehdään päivittäin Googlen hakukoneella. Ks. lisää Googlen hakustatistiikoista osoitteesta: <https://www.internetlivestats.com/google-search-statistics/>.

⁷ Jos haluamme esimerkiksi tietoa siitä, miten tekstin- ja tiedonlouhinta määritellään, niin voimme kirjoittaa sanan ”TDM” Googlen hakukoneeseen ja saada 340 miljoonaa hakutulosta ainoastaan 0,55 sekunnissa. Katso hakutulos hakusanalle ”*text and data mining*” osoitteesta: https://www.google.com/search?rlz=1C1GCEB_enSE851FI852&ei=1qMIXqDeBoS-wkwWFxbagBA&q=text+and+data+mining&oq=text+and+data+mining&gs_l=psy-ab.3..015j0i22i3015.75934.75934..77043...0.1..0.85.85.1.....0....2j1..gws-wiz.....0i71.RvVBkE-QvF6c&ved=0ahUKFwigoce5nZLnAhUE2KQKHYWiDUQQ4dUDCAs&uact=5&safe=active&ssui=on.

⁸ Euroopan parlamentin ja neuvoston direktiivi (EU) 2019/790, annettu 17 päivänä huhtikuuta 2019, tekijänoikeudesta ja lähioikeuksista digitaalisilla sisämarkkinoilla ja direktiivien 96/9/EY ja 2001/29/EY muuttamisesta.

⁹ Tässä tutkielmassa käytetään ilmaisuja kappaleen valmistaminen ja kopiointi synonyymeina toisistaan, vaikkakin tosiasiallisesti kappaleen valmistaminen on yläkäsite, jonka alle käsite kopiointi sisältyy. Valintaa perustellaan tutkielman selkeyttämisellä.

Analysointiteknologian yhteydessä otettavat kopiot ovat kuitenkin ongelmallisia tekijänoikeudellisesta näkökulmasta, nimittäin Euroopan unionin (”EU”) tekijänoikeuslainsäädännön lähtökohdaksi on asetettu tietoyhteiskuntadirektiivissä¹⁰ se, että tekijänoikeuden oikeudenhaltijalla on yksinoikeus valmistaa kappaleita teoksestaan. Yksinoikeuden myötä digitaalisessa ympäristössä on ilmennyt epävarmuutta siitä, miten tekijänoikeudella suojattua materiaalia voidaan hyödyntää tekstin- ja tiedonlouhintateknologiassa. Epävarmuus on liittynyt muun muassa siihen, missä laajuudessa oikeudenhaltijat voivat rajoittaa pääsyä materiaaliin sekä siihen, missä tilanteissa ja mitä tarkoitusta varten toimijoiden on sallittua valmistaa kopioita.¹¹

Lisäksi Euroopan unionissa on ollut havaittavissa tekijänoikeuslainsäädännön tasolla sirpaloitumista, sillä jäsenvaltiot ovat omaksuneet eri lähestymistapoja sen suhteen, miten sovittaa yhteen tekijänoikeus tekstin- ja tiedonlouhintateknologian kanssa. Tämä oikeudellinen epävarmuus tunnistettiin myös laadittaessa *Digitaalisten sisämarkkinoiden strategiaa Euroopalle*.¹² Kyseisen strategian myötä aloitettiin lainsäädäntöprosessi, jonka lopputuloksena syntyi tekijänoikeusdirektiivi, joka hyväksyttiin Euroopan unionin neuvostossa 14. huhtikuuta 2019¹³. Seuraavaksi jäsenvaltioiden on implementoitava direktiivi kansalliseen lainsäädäntöönsä 7. kesäkuuta 2021 mennessä.¹⁴

Tekijänoikeusdirektiivi lähtee liikkeelle siitä, että tekstin- ja tiedonlouhinnan yhteydessä otettavat kopiot kuuluvat tekijänoikeuden yksinoikeuden alaan, eikä kopioita voida näin ollen ottaa ilman oikeudenhaltijan suostumusta. Tekijänoikeusdirektiiviin sisällytettiin kuitenkin artiklat 3 ja 4, jotka asettavat implementoinnin jälkeen puitteet, joiden rajoissa voidaan tehdä kopioita tekstin- ja tiedonlouhinnan yhteydessä. Tilanteet, joissa tekstin- ja tiedonlouhintaa voidaan harjoittaa, jaetaan uudessa tekijänoikeusdirektiivissä tieteellisiin sekä muihin tarkoituseriin. Jaotelusta aiheutuu, että varsinkin kaupallisia tarkoituseriä varten suoritettu tekstin- ja tiedonlouhinta kohtaa lainsäädännön tasolla uusia rajoituksia. Näin ollen myös hakukonetoiminta,

¹⁰ Euroopan parlamentin ja neuvoston direktiivi 2001/29/EY, annettu 22 päivänä toukokuuta 2001, tekijänoikeuden ja lähioikeuksien tiettyjen piirteiden yhdenmukaistamisesta tietoyhteiskunnassa.

¹¹ Oikeudellinen epävarmuus on tunnistettu laajalti niin oikeuskirjallisuudessa kuin myös lainsäädäntövaiheessa. Katso esimerkiksi tekijänoikeusdirektiivi, resitaali 3 sekä *Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: oikeudellinen analyysi* 2018, s. 4–5.

¹² Komission tiedonanto Euroopan parlamentille, neuvostolle, Euroopan talous- ja sosiaalikomitealle ja alueiden komitealle. Digitaalisten sisämarkkinoiden strategia Euroopalle. COM (2015) 192, Bryssel 6.5.2015 (”*Digitaalisten sisämarkkinoiden strategia Euroopalle*”). Ks. erityisesti *Digitaalisten sisämarkkinoiden strategia Euroopalle*, s. 7.

¹³ *Euroopan komission lehdistötiedote* 2019.

¹⁴ Tekijänoikeusdirektiivin 29 artikla 1 momentti.

jonka taustalla vaikuttavat kaupalliset tarkoitusperät, joutuu uuteen epävarmaan oikeudelliseen toimintaympäristöön.

Törmäyskurssi tekijänoikeuden ja tekstin- ja tiedonlouhintateknologian välillä on kuitenkin ollut väistämätön, kun tekijänoikeuden vanhat, analogiseen maailmaan suunnitellut lainsäädännölliset rakenteet kohtaavat uudet innovatiiviset teknologiat. Se, että uusi innovaatio haastaa tekijänoikeuden perusoppirakenteita, ei ole kuitenkaan mikään uusi oikeudellinen ilmiö, vaan jo muutaman vuosikymmenen ajan on puhuttu *tekijänoikeuden digitaalisesta kriisistä*. Oikeudenhaltijoiden oikeudet ovat vahvistuneet teknologian kehittyessä. Kehitys on kuitenkin tapahtunut ottamatta huomioon tasapainottavia mekanismeja, jotka takaisivat sen, ettei yksinoikeuden korostuneella suojelulla olisi teosten uusia hyödyntämistapoja taikka teknologian kehitystä jarruttavaa vaikutusta.¹⁵ Tässä tutkielmassa tarkastelun kohteeksi nouseekin tekijänoikeussäätelyn vaikutus Euroopan unionissa hakukoneiden harjoittamaan kaupalliseen tekstin- ja tiedonlouhintaan.

1.2 Tutkimusongelma ja aiheen rajaus

Tässä tutkielmassa tarkastellaan analyttisesti ja järjestelmällisesti hakukoneiden suorittaman kaupallisen tekstin- ja tiedonlouhinnan juridista viitekehystä ja mahdollisia oikeudellisia haasteita digitalisoituneessa ympäristössä. Tutkielman tutkimusongelma on, mikä on tämän hetkinen oikeustila Euroopan unionissa hakukoneiden suorittaman kaupallisen tekstin- ja tiedonlouhinnan suhteen. Tutkimusongelmaan vastaaminen edellyttää sitä, että tutkimusongelma jaetaan kolmeen pienempään tutkimuskysymykseen.

Ensinnäkin on selvitettävä käytännön tasolla, mitä tarkoitetaan tekstin- ja tiedonlouhinnalla ja ymmärrettävä, miten koneoppimiseen perustuvat hakukoneet toimivat. Tämän jälkeen voidaan tuoda tarkasteluun tekijänoikeudellinen ulottuvuus. *Toisena* tutkimuskysymyksenä selvitetään tietotekniikan ymmärrystä soveltaen, mihin tekstin- ja tiedonlouhintateknologian ja tekijänoikeuden välinen jännite perustuu. Tutkimuskysymykseen vastaaminen edellyttää ymmärrystä tekijänoikeuden peruskäsitteistä, kuten tekijänoikeuden sisällöstä, tekijänoikeuden taloudellisesta ulottuvuudesta ja etenkin kappaleen valmistamisesta.

¹⁵ Salokannel 2013, s. 10.

Kolmanneksi tutkimuskysymykseksi on nostettu kysymys siitä, miten tekijänoikeus kohtasi tekstin- ja tiedonlouhintaan perustuvan teknologian oikeudellisella rajapinnalla ennen tekijänoikeusdirektiiviä. Analysointi keskittyy erityisesti Euroopan unionin tekijänoikeudelliseen normistoon, ja niihin tilanteisiin, joissa normiston mukaan voidaan rajoittaa tekijänoikeutta. Lisäksi kolmas tutkimuskysymys selventää tutkimusongelman taustalla olevaa laajempaa yhteiskunnallista ilmiötä liittyen tekijänoikeuden digitaaliseen kriisiin. Kolmanteen tutkimuskysymykseen vastaaminen valottaa, mihin tekijänoikeusdirektiivin säätämistarve perustui.

Peruskäsitteiden ja uudistamistarpeen ymmärtämisen jälkeen voidaan vastata itse tutkimusongelmaan, eli siihen, mikä on tämän hetkinen oikeustila Euroopan unionissa hakukoneiden suorittaman kaupallisen tekstin- ja tiedonlouhinnan suhteen. Tähän tutkimuskysymykseen vastaamiseksi on analysoitu erityisesti sitä, miten tekijänoikeusdirektiivin artikkelit 3 ja 4 muuttavat oikeustilaa kaupallisen tekstin- ja tiedonlouhinnan suhteen.

Jotta tutkimusongelmaan voidaan antaa vastaukseksi toimiva kokonaisuus, on tässä tutkielmassa tehty neljä laajempaa rajausta, jotka liittyvät henkilötietoihin, tietokantoihin, tekijänoikeusdirektiivin muuhun sisältöön sekä yleisölle välittämistä koskevaan yksinoikeuteen.

Ensinnäkin hakukoneiden suorittamaan kaupalliseen tekstin- ja tiedonlouhintaan liittyvät henkilötietojen näkökulmasta oikeudellisia ongelmia. Huomioitaessa tekstin- ja tiedonlouhinnassa käsiteltävät laajat tietomassat, on väistämätöntä, että koneoppimiseen perustuvassa prosessissa saatetaan myös käsitellä yksityishenkilöiden henkilötietoja. Vuonna 2016 voimaan tullut yleinen tietosuojasetus¹⁶ asettaa rajoituksia sille, miten automatisoidussa tekstin- ja tiedonlouhinnassa voidaan käsitellä henkilötietoja.¹⁷ Henkilötietoihin liittyvät näkökulmat muodostavat täysin oman aiheensa, eikä niitä ole tarkoituksenmukaista tarkastella enempää tässä tutkielmassa.

¹⁶ Euroopan parlamentin ja neuvoston asetus (EU) 2016/679, annettu 27 päivänä huhtikuuta 2016, luonnollisten henkilöiden suojelusta henkilötietojen käsittelyssä sekä näiden tietojen vapaasta liikkuvuudesta ja direktiivin 95/46/EY kumoamisesta (”yleinen tietosuojasetus”).

¹⁷ Euroopan komission tutkimuksen ja innovaation pääosaston asiantuntijaryhmä on lausunnossaan ”Standardisation in the area of innovation and technological development, notably in the field of Text and Data Mining 2014” käsitellyt kattavasti aihealuetta. Ks. *Euroopan komission tutkimuksen ja innovoinnin pääosaston asiantuntijalautsunto* 2014, s. 60–63.

Tarkasteltaessa – varsinkin hakukoneiden suorittamaan – tekstin- ja tiedonlouhintaan liittyviä oikeudellisia ongelmia, on keskeistä tunnistaa tietokantoihin¹⁸ liittyvä juridinen problematiikka. Olennainen osa tekstin- ja tiedonlouhinnassa käytettävästä materiaalista voi sijaita myös tietokannoissa. Tietokantojen tekijät voivat kuitenkin tekijänoikeuden¹⁹, sui generis -oikeuden²⁰ ja/tai sopimusoikeuden²¹ nojalla rajoittaa kopiointia näistä tietokannoista.²² Tietokantoihin liittyvä oikeudellinen viitekehys on kuitenkin siinä määrin laaja, että sivuteknisistä syistä toinen rajattava kokonaisuus käsittää tietokannat.²³

Kolmas keskeinen rajaus koskee tekijänoikeusdirektiiviä. Tutkielma keskittyy nimenomaisesti tekijänoikeusdirektiivin 3 ja 4 artikloihin, jolloin tarkastelun ulkopuolelle jää esimerkiksi lehdistulkaisujen suoja verkkokäytössä koskeva tekijänoikeusdirektiivin 15 artikla. Lisäksi tutkielma tutkii tekijänoikeusdirektiivin 3 ja 4 artikloita lähtökohtaisesti pelkästään kaupallisen

¹⁸ Tietokannan käsite määritellään tietokantadirektiivissä (Euroopan parlamentin ja neuvoston direktiivi 96/9/EY, annettu 11 päivänä maaliskuuta 1996, tietokantojen oikeudellisesta suojasta), jonka 1 artiklan mukaan tietokannalla tarkoitetaan teosten, tietojen ja muiden itsenäisten aineistojen kokoelmaa, joka on järjestetty järjestelmällisellä tavalla ja johon on elektronisesti tai muulla tavoin mahdollisesti pääsy. Lisäksi Euroopan unionin tuomioistuimien on oikeuskäytännössään tulkinnut laajasti tietokannan käsitettä ja näin ollen oikeudellista suoja saavat ”kaikenmuotoiset” tietokannat. Tietokannan käsitteen on ymmärrettävä tällöin tarkoittavan kaikkia kirjallisia, taiteellisia, musiikillisia tai vastaavia teoskokoelmia, tai tekstiä, ääntä, kuvia, lukuja, tosiasioita ja tietoja sisältäviä kokoelmia. Ks. lisää aiheeseen liittyen: asia C-490/14 *Verlag Esterbauer*, kohdat 13 ja 14 sekä asia C-444/08 *Fixtures Marketing*, kohta 20.

¹⁹ Tietokantadirektiivin 5 artikla määrittelee tilanteet, joissa tietokannan tekijä saa tekijänoikeudellista suoja. Suojan saaminen edellyttää kuitenkin omaperäisyyskriteerin täyttymistä, mikä merkitsee sitä, että tietokannan tekijän on ilmaistava tietokantaan sisällytettävien tietojen valitsemisen taikka järjestämisen kautta luovuuttaan omaperäisellä tavalla tekemällä vapaita ja luovia ratkaisuja. Omaperäisyyskriteeri ei kuitenkaan täyty, kun valmistamisen sanelevat tekniset seikat, säännöt tai rajoitteet, jotka eivät jätä tilaa luomisen vapaudelle. Ks. lisää asia C-604/10 *Football Dataco*, kohdat 38 ja 39.

²⁰ Sui generis -oikeus antaa suoja niissä tilanteissa, joissa omaperäisyyskriteeri ei täyty tietokannan kohdalla. Tietokantadirektiivin 7 artiklan nojalla voidaan nimittäin suojata mitä tahansa kokoelmaa dataa, informaatiota tai teoksia, joiden kerääminen, varmistaminen tai esittäminen ovat vaatineet huomattavaa investointia. Kategorisesti sui generis -oikeus on omankaltaisensa suojaus, jota arvioidaan erillään tekijänoikeussuojasta ja sen mahdollisesta olemassaolosta samassa tietokannassa. Katso tarkemmin aiheesta asia C-604/10 *Football Dataco*, kohta 27.

²¹ Tietokantoja voidaan myös säännellä sopimusoikeudellisesti. Tietyissä tilanteissa tietokantojen käyttö voi vaatia sopimusehtojen hyväksymistä, joissa voidaan rajoittaa tietokantojen käyttämistä tiettyihin toimintoihin, kuten tekstin- ja tiedonlouhintaan. Ks. lisää aiheesta *Borghini*, s. 4 sekä katso myös asia C-30/14 *Ryanair* ja erityisesti ratkaisun kohta 46.

²² Tietokannan suojaamisesta tekijänoikeuden tai sui generis -oikeuden nojalla voidaan tunnistaa neljä eri mahdollisuutta: se, että tietokanta saa 1) tekijänoikeussuojaa, 2) sui generis -suoja, 3) molempia suoja tai 4) tietokanta ei saa kumpaakaan suoja. Katso tarkemmin aiheesta asia C-604/10 *Football Dataco*, kohta 27 sekä katso myös *Ohisalo* 2017, s. 76.

²³ Tietokantojen ja tekstin- ja tiedonlouhinnan yhteensovittamista on tarkasteltu esimerkiksi yksityiskohtaisemmin seuraavissa lähteissä: *Salokannel* 2013; *Ducato – Strowel* 2018, s. 6–14; *Triaille – de Meeûs d’Argentuil – de Francquen* 2014, s. 33–41 sekä *Euroopan komission tutkimuksen ja innovoinnin pääosaston asiantuntijalausunto* 2014, s. 47–48. Lisäksi tietokantoihin liittyy Euroopan unionin tuomioistuimen oikeuskäytäntöä. Keskeisiä ratkaisuja ovat asia C-46/02 *Fixtures v Veikkaus*, asia C-203/02 *The British Horseracing Board*, asia C-338/02 *Fixtures v Svenska Spel* sekä asia C-444/02 *Fixtures Marketing*.

tekstin- ja tiedonlouhinnan näkökulmasta. On nimittäin huomioitava, että tekijänoikeusdirektiivin tekstin- ja tiedonlouhintaa koskevilla säännöksillä on myös vaikutuksia muuhun kuin kaupalliseen sektoriin, kuten tieteelliseen tutkimukseen taikka journalismiin²⁴.

Neljäs kokonaisuus, joka rajataan tutkielmassa tarkastelun ulkopuolelle, käsittää taloudellisen oikeuden kannalta toisen keskeisen yksinoikeuden – yleisölle välittämisen²⁵. Tekijänoikeusdirektiivin artikloiden 3 ja 4 sisältämät poikkeukset ulottuvat ainoastaan tilanteisiin, joissa tekijänoikeudella suojatusta materiaalista valmistetaan kopioita. Artiklat eivät ota kantaa siihen, miten tulisi tarkastella tilannetta, jossa osana tekstin- ja tiedonlouhintaprosessia tekijänoikeudella suojattua materiaalia välitetään yleisölle.²⁶ Näin ollen tutkielmassa keskitytään yksityiskohtaisemmin kappaleen valmistamista koskevaan yksinoikeuteen ja sen analysointiin.

1.3 Tutkimusmenetelmät ja käytettävät oikeuslähteet

Tutkittaessa, mikä on tämän hetkinen oikeustila Euroopan unionissa hakukoneiden suorittaman kaupallisen tekstin- ja tiedonlouhinnan suhteen, valikoituu tutkimusmenetelmiksi oikeusdogmaattinen eli lainopillinen tutkimus. Oikeusdogmaattisella tutkimuksella systematisoidaan ja tulkitaan voimassa olevan oikeuden sisältöä.²⁷ Systematisoinnilla viitataan siihen, että määritetään voimassa olevien oikeussääntöjen välisiä suhteita.²⁸ Tässä tutkielmassa systematisointi ilmenee siinä, kun määritellään tekijänoikeuslainsäädännössä esitetyn oikeudenhaltijan yksinoikeuden suhdetta oikeudesta tehtäviin poikkeuksiin ja rajoituksiin.

Voimassa olevan oikeuden tulkinnalla puolestaan viitataan menettelyyn, jossa annetaan tulkinnallinen merkityssisältö oikeussäännölle.²⁹ Tulkintaa käytetään tässä tutkielmassa, kun tarkastellaan tekijänoikeuden kannalta keskeisiä käsitteitä ja niiden merkityssisältöjä. Lisäksi keskei-

²⁴ *Kretschmer – Margoni* 2018b. On linjattu, että tekijänoikeusdirektiivi vaikuttaa myös journalistien sekä innovatiivisten yhtiöiden suorittamaan tekstin- ja tiedonlouhintaan.

²⁵ Tietoyhteiskuntadirektiivin 3 artiklassa on määritelty yleisölle välittämistä koskeva yksinoikeus yksinoikeudeksi sallia tai kieltää teosten langallinen tai langaton välittäminen yleisölle. Yksinoikeus käsittää myös teosten saattamisen yleisön saataviin siten, että yleisöön kuuluvilla henkilöillä on mahdollisuus saada teokset saataviinsa itse valitsemastaan paikasta ja itse valitsemanaan aikana.

²⁶ Ks. *Asia C-5/08 Infopaq I*, kohta 30. Lisäksi esimerkiksi *Depreeuw* on käsitellyt kattavasti aihetta. Katso lisää esimerkiksi *Depreeuw* 2014, s. 537–538.

²⁷ *Aarnio* 1989, s. 48.

²⁸ *Siltala* 2003, s. 384.

²⁹ *Siltala* 2003, s. 384.

senä osana oikeusdogmaattista tutkimusmetodia on oikeusperiaatteiden punninta ja niiden yhteensovittaminen.³⁰ Näin ollen tässä tutkielmassa esitetään punnintakannanottoja siitä, miten kaupallisen tekstin- ja tiedonlouhinnan sekä tekijänoikeuden taustalla olevat oikeusperiaatteet ovat sovitettavissa yhteen.

Teknologisen innovaation ollessa keskeinen osa tätä tutkielmaa on olennaista ottaa huomioon myös *Law and Technology* -tutkimusmetodi, joka yhdistää tieteellisessä tutkimuksessa oikeustieteen teknologiaan.³¹ Tutkimusmetodi on tarpeellinen, sillä perinteinen oikeudellinen analyysi ei kykene riittävässä määrin ymmärtämään oikeuden ja teknologian välistä yhteyttä. Lisäksi metodin avulla voidaan paremmin huomioida teknologisten muutosten vaikutuksia yhteiskunnan kannalta tärkeisiin tavoitteisiin ja arvoihin.³²

Tutkielmassa omaksuttu *Law and Technology* -tutkimusmetodi on luonteeltaan liberaalinen ja käsittää neljä eri osa-aluetta.³³ Tällä viitataan siihen, että ensinnäkin tutkimusmetodin avulla selvitetään teknologian ja lain välistä monimutkaista vuorovaikutusta, eli sitä miten tekstin- ja tiedonlouhintateknologia on ylipäänsä yhteydessä tekijänoikeuteen. Toisaalta tutkimusmetodin avulla pyritään tunnistamaan teknologian asettamat riskitekijät tekijänoikeudelle ja sen ytimeen sijoittuvalle taloudelliselle ulottuvuulle ("*law is technology*").

Kolmanneksi tutkimusmetodin tarkoituksena on tunnistaa tilanteita, joissa lain avulla voidaan kannustaa tiettyjen teknologioiden käyttöön tiettyjen arvojen suojelemiseksi ("*technology is law*"). Tutkimusmetodin neljäs osa-alue sisältää analysoinnin, jonka pyrkimyksenä on sovittaa normiston kehittyminen teknologiseen ympäristöön ("*is the legal rule scientific?*").³⁴ Neljäs osa-alue ilmenee tässä tutkielmassa annettavien *de lege ferenda* -suositusten³⁵ ja *de lege sententia* -suositusten³⁶ muodossa. Esitettävät suositukset ovat tulkintakannanottoja siitä, miten

³⁰ Hirvonen 2011, s. 24 sekä Siltala 2003, s. 384.

³¹ Cockfield 2004, s. 384.

³² Cockfield 2004, s. 384.

³³ Cockfield on esittänyt, että tutkimusmetodista voidaan esittää kaksi suuntausta. Ensinnäkin liberaali lähestymistapa, joka pyrkii tarjoamaan teknologiaherkkiä ja joustavia oikeudellisia ratkaisuja, jotka poikkeavat perinteisesti omaksutuista lähestymistavoista ja oppirakenteista. Toinen suuntaus on luonteeltaan enemmän konservatiivinen ja nojautuu enemmän perinteiseen oikeudelliseen analyysiin. Ks. lisää Cockfield 2004, s. 384–385.

³⁴ Cockfield 2004, s. 399–400.

³⁵ *De lege ferenda* -suositukset ovat lainvalmistelulle suunnattuja lainsäädännöllisiä ratkaisumalleja siitä, millainen lain tulisi olla. Ks. lisää Siltala 2003, s. 131.

³⁶ *De lege sententia* -suosituksilla viitataan tutkimukseen, jonka tarkoituksena on vaikuttaa Euroopan unionin tuomioistuimien tuottamaan oikeuskäytäntöön. Ks. lisää Siltala 2003, s. 132.

hakukoneiden suorittaman kaupallisen tekstin- ja tiedonlouhinnan sekä tekijänoikeuden yhteensovittamista olisi mahdollista kehittää oikeudellisesti.

Tässä tutkielmassa hyödynnetään myös oikeusvertailevia näkökulmia, kun analysoidaan kvalitatiivisesti kaupallisen tekstin- ja tiedonlouhinnan oikeudellista viitekehystä.³⁷ Vertailukohtia haetaan yhdysvaltalaisesta järjestelmästä. Eurooppalaista järjestelmää on mielekästä vertailla yhdysvaltalaiseen järjestelmään useasta syystä. Ensinnäkin molemmissa järjestelmissä teknologinen kehitys on korkeatasoista ja samat teknologiset keksinnöt aiheuttavat haasteita lainsäädännölle. Toiseksi *Digitaalisten sisämarkkinoiden strategiassa Euroopalle* on nostettu tavoitteeksi kilpailukyvyyn parantaminen digitaalisella akselilla³⁸, ja Yhdysvallat on yksi merkittävimmistä Euroopan unionin kilpailijoista. Oikeusvertailun tarkoituksena ei ole tässä tutkielmassa tuottaa normatiivisia tuloksia, vaan oikeusvertailua käytetään ainoastaan syventämään ja parantamaan laadullisesti oikeudellista argumentaatiota.³⁹ Näin ollen tutkielma ei ole tutkimusmetodisesti oikeusvertaileva.

Jotta voidaan vastata, mikä on tämän hetkinen oikeustila, tulee hyödyntää voimassa olevaa lainsäädäntöä⁴⁰, sen valmisteluaineistoa, oikeuskäytäntöä sekä oikeuskirjallisuutta. Teknologinen ulottuvuus edellyttää myös ymmärrystä tietotekniikasta ja tietyistä tietotekniikan periaatteista. Tämän vuoksi tässä tutkielmassa hyödynnetään myös tietotekniikkaan liittyvää kirjallisuutta. Oikeustilan laadullisessa arvioinnissa hyödynnetään erityisesti asiantuntijalausuntoja, oikeuskirjallisuutta sekä oikeusvertailevia argumentteja. Sen lisäksi, että kannanottoja peilataan tietotekniisiin, oikeusvertaileviin ja tekijänoikeuden doktriinin peruseriaatteisiin, perustellaan tässä tutkielmassa esitettäviä kannanottoja myös yhteiskunnallisilla tavoitteilla sekä reaalisilla argumenteilla⁴¹.

1.4 Kirjallisuuskatsaus ja tutkielman tarpeellisuus

Jotta tässä tutkielmassa voidaan tarkastella yhteiskunnallisiin ilmiöihin sidoksissa olevaa oikeudellista ongelmaa, on olennaista ymmärtää aihealueen kannalta relevantteja oikeudellisia käsitteitä. *Pilan ja Torremansin kirja*⁴² on toiminut keskeisenä lähteenä, kun on ollut tarpeen

³⁷ Husa 2013, s. 32–34.

³⁸ *Digitaalisten sisämarkkinoiden strategia Euroopalle*, s. 4.

³⁹ Husa 2013, s. 34.

⁴⁰ Lain määritelmä ymmärretään tässä tutkielmassa laajaksi sisältäen Euroopan unionin oikeuden, ihmisoikeudet, perusoikeudet ja oikeusperiaatteet.

⁴¹ Ks. esim. *Aarnio* 1989, s. 217. *Aarnio* on määritellyt, että esimerkiksi reaaliset argumentit ja moraaliset näkökulmat voidaan laajassa mielessä tulkita oikeuslähteiksi, jotka kelpaavat juridisten tulkintaperusteiden tueksi.

⁴² *Pila – Torremans* 2019.

saada tietoa tekijänoikeuteen, kappaleen valmistamiseen ja tekijänoikeuden rajoittamiseen liittyvistä käsitteistä voimassa olevan lainsäädännön ja oikeuskäytännön valossa.

Triailien, de Meeûs d'Argenteuilin ja de Francquen tutkimus⁴³ tekstin- ja tiedonlouhinnan oikeudellisesta viitekehyksestä tarjoaa kattavan ymmärryksen siitä, mikä on ollut tekstin- ja tiedonlouhinnan asema ennen tekijänoikeusdirektiiviä. Tämän tutkielman tarkoituksena on jatkaa oikeustilan tarkastelua tekijänoikeusdirektiivin voimaantulon jälkeen. Johtuen tekijänoikeusdirektiivissä olevasta jaottelusta tieteelliseen ja muuhun tekstin- ja tiedonlouhintaan, on tässä tutkielmassa haluttu erityisesti korostaa kaupallista näkökulmaa, jota ei *Triailien, de Meeûs d'Argenteuilin ja de Francquen* tutkimuksessa ole otettu lähtökohdaksi.

Jotta kaupalliseen tekstin- ja tiedonlouhintaan ei kohdistettaisi lainsäädännön tasolla tarpeettomia rajoitteita, käytetään tässä tutkielmassa konkreettista esimerkkiä kaupallisen tekstin- ja tiedonlouhinnan soveltamiskohteesta, ja näin ollen myös korostetaan kaupallisen tekstin- ja tiedonlouhinnan tarpeellisuutta. Tästä johtuen tässä tutkielmassa keskitytään erityisesti hakukoneiden toimintaan. Tärkeäksi kokonaisuudeksi tässä tutkielmassa muodostuu sen ymmärtäminen, miten hakukoneet käytännössä toimivat ja miten niiden toiminta on yhteydessä tekijänoikeuteen. Tämän näkökulman mahdollistamiseksi tutkielmassa hyödynnetään tietotekniikkaan keskittyvää aineistoa sekä erityisesti *Spraulin* tutkimusta⁴⁴ siitä, miten hakukoneet toimivat. Monitieteisen näkökulman tuomisella tutkielmaan pyritään siihen, että voidaan tunnistaa käytännön tasolla ne tilanteet, joihin liittyy tekijänoikeuden näkökulmasta haasteita.

*Strowelin*⁴⁵ ja *Depreeuwn*⁴⁶ tutkimukset ovat puolestaan keskeisiä tiedonlähteitä, kun pyritään ymmärtämään tutkimusongelman taustalla olevaa laajempaa ilmiötä, eli tekijänoikeuden taloudellisen ulottuvuuden murrosta digitaalisella aikakaudella. Tässä tutkielmassa on tavoitteena jatkaa heidän tutkimusta arvioimalla tuoreen tekijänoikeusdirektiivin vaikutuksia tekijänoikeuden taloudellisen ulottuvuuteen.

Muutostarpeesta esitettävälle kannanotoille on saatu argumentatiivista tukea *Sagin* tutkimuksesta⁴⁷ tekstin- ja tiedonlouhinnan sekä koneoppimisen haasteista yhdysvaltalaisessa ympäris-

⁴³ *Triaille – de Meeûs d'Argenteuil – de Francquen* 2014.

⁴⁴ *Spraul* 2015.

⁴⁵ *Strowel* 2018, *Strowel* 1993, *Strowel – Derclaye* 2001 sekä *Strowel* 2014.

⁴⁶ *Depreeuw* 2014.

⁴⁷ *Sag* 2019.

tössä. Vaikka tämän tutkielman tarkoituksena on tarkastella asiaa Euroopan unionin näkökulmasta, on hyödyllistä tunnistaa, miten toisessa oikeuskulttuurissa on käsitelty kyseistä vuorovaikutussuhdetta ja onko siellä mahdollisesti omaksuttu näkökulmia, joita voitaisiin hyödyntää Euroopan unionissa.

1.5 Tutkielman rakenne

Tutkielma jakautuu seitsemään lukuun, jotka muodostavat oikeudellisen arvioinnin siitä, miten hakukoneiden suorittamaa kaupallista tekstin- ja tiedonlouhintaa on lähestytty tekijänoikeudellisesta näkökulmasta Euroopan unionissa.

Ensimmäinen luku johdattelee niihin oikeudellisiin haasteisiin, jotka liittyvät hakukoneiden suorittaman kaupallisen tekstin- ja tiedonlouhinnan yhteensovittamiseen tekijänoikeuden kanssa digitaalisessa ympäristössä. Tämän lisäksi luvussa rajataan tutkimuksessa esitetty aihe, esitellään tutkimusmenetelmät ja tutkielman rakenne sekä esitellään kirjallisuuskatsauksen muodossa aihealueeseen liittyvää aikaisempaa olennaista tieteellistä tutkimusta. *Toisessa* luvussa perehdytään tekstin- ja tiedonlouhintaan. Tarkoituksena on selventää, mitä tekstin- ja tiedonlouhinta tarkoittaa ja miten sitä voidaan soveltaa nyky-yhteiskunnassa. Luku tarjoaa teknologisen lähestymistavan ja siinä kuvataan, miten hakukoneet toimivat.

Tutkielman *kolmannessa* luvussa keskitytään tekijänoikeuteen. Luku perehtyy ensin tekijänoikeuden käsitteen sisältöön ja sen taustalla vaikuttaviin periaatteisiin. Tämän jälkeen luku syvenyy ymmärtämään, mikä yhteys tekstin- ja tiedonlouhinnalla on tekijänoikeuteen, ja tarkastelee siihen liittyvää oikeustilaa ennen tekijänoikeusdirektiiviä. Lisäksi luvussa tarkastellaan tekijänoikeutta osana digitaalista kehityspolkua. Luvun tarkoituksena on näin ollen myös selvittää, miten teknologinen kehitys haastaa tekijänoikeuden oppirakenteita.

Tekijänoikeuden ja teknologian välistä jännitettä analysoidaan tarkemmin *neljännessä* luvussa, jossa käsitellään niitä edellytyksiä, joiden perusteella tekijänoikeutta voidaan rajoittaa. *Viides* luku keskittyy analysoimaan sitä, mikä on nykyinen oikeustila Euroopan unionissa hakukoneiden suorittaman kaupallisen tekstin- ja tiedonlouhinnan suhteen. *Kuudennessa* luvussa annetaan kannanottoja siitä, miten kaupallisen tekstin- ja tiedonlouhinnan oikeudellista ympäristöä tulisi kehittää Euroopan unionissa. Tutkielman päättää *seitsemäs* luku, johon on koottu yhteen tässä tutkielmassa esitetyt johtopäätökset.

2 TEKSTIN- JA TIEDONLOUHINTA

2.1 Määritelmä

Big datan aikakaudella tietoa on kerättävissä suunnattomat määrät mitä moninaisimmista lähteistä, ja tiedon määrä kasvaa koko ajan kohti ääretöntä. Tämän tietomassan käsittelyyn, järjestykseen ja havainnointiin eivät perinteiset analogiset analysointimenetelmät ole riittäviä. Samalla on myös tunnistettava, ettei tiedon arvo perustu siihen, että sitä tarkasteltaisiin sen alkuperäisessä kontekstissa eristyksessä muusta tiedosta – vaan päinvastoin – tarvitaan teknologiaa, joka louhii tietoa ja kykenee hyödyntämään näissä äärettömissä tietomassoissa piilevän arvon yhdistelemällä tietoa uudella tavalla, ja tuottamaan näin louhitusta tiedosta esimerkiksi uusia korrelaatioita, väittämiä, malleja, suuntauksia taikka havaintoja. Tätä teknologiaa kutsutaan tekstin- ja tiedonlouhinnaksi.⁴⁸ Lisäksi tekstin- ja tiedonlouhintateknologia pystyy analysoimaan tietoa murto-osassa siitä ajasta, mikä ihmisillä menisi samaan analysointiin.⁴⁹

Tekstin- ja tiedonlouhintaa käytetään laajalti esimerkiksi tieteellisessä ja lääketieteellisessä tutkimuksessa, liike-elämässä sekä kansallisen turvallisuuden kehittämisessä. Tieteen kehittyminen kasvattaa entisestään tämän teknologian potentiaalia ja sitä voidaan etenevissä määrin hyödyntää parantamaan talouskasvua ja kilpailukykyä.⁵⁰

Tekstin- ja tiedonlouhinnalla on myös olennainen yhteys tekoälyyn ja sen kehittämiseen, sillä tekstin- ja tiedonlouhintateknologiaa hyödynnetään koneoppimisessa.⁵¹ Koneoppimisella viitataan tilasto- ja ohjelmointitekniikoiden rykelmään, jonka avulla tietokoneet voivat itse opettaa itseään. Koneoppiminen perustuu siihen, että tietokoneet parantavat suoritustaan jokaisella suorituskerralla käytössä olevan datan perusteella ilman nimenomaista päivitystä.⁵² Koneoppiminen on arvokasta etenkin silloin, kun käsillä oleva tehtävä sisältää periaatteita, jotka ovat vaikeasti selitettäviä tai muutoin äärimmäisen monimutkaisia.⁵³

⁴⁸ Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: tekninen analyysi 2018, s. 1 sekä Triaille – de Meeûs d'Argentuil – de Francquen 2014, s. 8–19.

⁴⁹ Jisc Report 2020.

⁵⁰ Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: tekninen analyysi 2018, s. 2.

⁵¹ Koneoppiminen luokitellaan yhdeksi tekoälyn alaryhmistä. Ks. lisää aiheesta Kersting 2018.

⁵² Katso lisää koneoppimisen määritelmästä osoitteesta <https://www.geeksforgeeks.org/getting-started-machine-learning/>.

⁵³ Sag, 2019, s. 4–7.

Koneoppimisen hyödyntämisen rajat ovat mittavat. Koneoppimista voidaan esimerkiksi hyödyntää täysin automatisoidussa kuvanmuokkauksessa⁵⁴ tai kvalitatiivisen ja luonnollisen kielen käsittelytyökaluna kriminologiassa ennustamaan jengiväkivaltaa Twitter-viestien perusteella.⁵⁵ Aikakaudella, jolla 2,5 triljoonaa tavua dataa tuotetaan päivittäin⁵⁶, muodostuu välttämättömäksi se, että myös ihmisillä on pääsy tähän suureen tietomäärään. Tärkeitä työkaluja tähän tietoon käsiksi pääsemiseen ovat internetin hakukoneet, joiden toiminta perustuu juuri tekstin- ja tiedonlouhintateknologiaa hyödyntävään koneoppimiseen. Seuraavaksi tässä luvussa syvennytään ymmärtämään, miten tekstin- ja tiedonlouhintateknologia toimii ja miten se soveltuu hakukoneiden toimintaan.

Tekstin- ja tiedonlouhintateknologian käyttäminen mitä moninaisimpiin tarkoituksiin ja prosesseihin aiheuttaa sen, ettei sille ole tunnistettavissa mitään tiettyä tieteellistä ja teknologista menetelmää.⁵⁷ Näin ollen eri soveltamistarkoituksiin voidaan käyttää eri tekstin- ja tiedonlouhintamenetelmiä. Esimerkiksi teknologinen konteksti, analysoinnin tarkoitus, kovalevyn laatu, kustannukset, valittu ohjelmisto sekä käytettävissä oleva aika vaikuttavat menetelmän valintaan.⁵⁸ Tekstin- ja tiedonlouhintamenetelmät voidaan kuitenkin karkeasti määritellä teknologioiksi, joissa tuotetaan informaatiota ja uutta tietoa suuresta määrästä digitaalista materiaalia, kuten tekstistä, datasta, äänistä, kuvista, muista elementeistä taikka näiden yhdistelmistä automatisoiduilla ohjelmistoilla.⁵⁹ Tekstin- ja tiedonlouhintateknologiasta voidaan erottaa seuraavat kolme⁶⁰ vaihetta:

1. Pääsy materiaaliin;
2. Kopiointi⁶¹; sekä

⁵⁴ Hardesty 2017.

⁵⁵ Patton – Rambow – Auerbach – Li – Frey 2018. Katso tekstin- ja tiedonlouhinnan muista soveltamiskohteista Sag 2019, s. 4–7.

⁵⁶ Vuonna 2017 IBM Marketing Cloudin tekemässä tutkimuksessa arvioitiin, että ylipäänsä 90 % maailman datasta on luotu lyhyessä ajassa – kahden vuoden sisällä. Ks. lisää IBM Marketing Cloud 2017.

⁵⁷ Termin tieteellisen kontekstin vuoksi sille on havaittavissa useampia määritelmiä. Tutkielman rajatun sivumäärän vuoksi kaikkia määritelmiä ei käydä tässä tutkielmassa läpi. Ks. eri määritelmistä tarkemmin esimerkiksi Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: tekninen analyysi 2018, s. 4 sekä Triaille – de Meeûs d’Argentuïl – de Francquen 2014, s. 8–19.

⁵⁸ Triaille – de Meeûs d’Argentuïl – de Francquen 2014, s. 28.

⁵⁹ Triaille – de Meeûs d’Argentuïl – de Francquen 2014, s. 17.

⁶⁰ Esimerkiksi Triaille – de Meeûs d’Argentuïl – de Francquen ovat määritelleet tekstin- ja tiedonlouhinnan viisiportaiseksi prosessiksi. Ks. lisää Triaille – de Meeûs d’Argentuïl – de Francquen 2014, s. 28. Tässä tutkielmassa on käytetty Rosatin esittämää kolmiportaista mallia, sillä se on luonteeltaan yksinkertaisempi ja kyseisen mallin vaiheet ovat paremmin sovitettavissa yhteen tekijänoikeuden kappaleen valmistamista koskevan oikeuden kanssa. Katso lisää kolmiportaisesta mallista Rosati 2019, s. 7–14.

⁶¹ Englanninkielisestä termistä *extraction* käytetään tässä tutkielmassa suomennosta kopioiminen.

3. Louhinta.

Esitettävä kolmiportainen malli ei aina ole kaikissa menetelmissä kronologinen, vaan se on luonteeltaan enemmän havainnollistava.⁶² Tekstin- ja tiedonlouhinnassa ei ole välttämättä erotettavissa selkeää siirtymää vaiheesta toiseen, vaan vaiheet voivat tapahtua osittain, limittäin ja osin myös samanaikaisesti.⁶³

Ensimmäinen vaihe tekstin- ja tiedonlouhintateknologiassa käsittää pääsyn tutkittavaan materiaaliin, eli toisin sanoen analysoitavan syöttömateriaalin identifioinnin.⁶⁴ Digitaalisessa muodossa oleva materiaali voi koostua teoksista, erikseen kerätystä datasta, tai jo olemassa olevissa tietokannoissa⁶⁵ olevasta järjestetystä datasta.⁶⁶ Tällöin tekstin- ja tiedonlouhintaa harjoittava toimija tarvitsee pääsyn tiettyyn mediaan tai tietokantaan. On olemassa tiettyjä toimijoita, jotka tarjoavat automaattisesti pääsyn tietokannoissaan olevaan materiaaliin. Esimerkiksi Wikipediassa sijaitseva tieto on suoraan koneluettavassa muodossa. Tällaiseen materiaaliin, johon on vapaa pääsy, liittyy toisaalta myös riski siitä, että se voi olla viallista.⁶⁷ Ongelmaksi tekstin- ja tiedonlouhinnan kannalta muodostuu etenkin se, jos pääsyä tietokantaan taikka mediaan on rajoitettu.

Ennen kuin tarkastellaan tarkemmin, miten pääsyä materiaaliin voidaan rajoittaa, on vastattava kysymykseen siitä, miksi pääsyä on ylipäänsä rajoitettava. Rajoittamisen taustalla vaikuttaa

⁶² Tekstin- ja tiedonlouhintateknologian eri vaiheita on esitetty useassa eri tieteellisessä tutkimuksessa. Ks. esimerkiksi *Sag* 2019, s. 49–62; *Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: tekninen analyysi* 2018, s. 4–7; *Rosati* 2019, s. 8; sekä *Triaille – de Meeüs d’Argen-tuil – de Francquen* 2014, s. 28.

⁶³ Esimerkiksi pääsy materiaaliin sekä kopioiminen voivat hyvin tapahtua jo samassa vaiheessa teknistä prosessia.

⁶⁴ *Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: tekninen analyysi* 2018, s. 4–7. Katso myös tarkemmin käytettävistä tekstin- ja tiedonlouhintateknologioista *Han – Kamber – Pei* 2012.

⁶⁵ Vaikkakin tässä kohtaa tutkielmaa viitataan tietokantoihin, ei tutkielmassa oteta erikseen tarkasteluun tietokantoihin liittyvää oikeudellista sääntelyä.

⁶⁶ Tässä tutkielmassa tarkastellaan ainoastaan digitaalisessa muodossa olevaa analysointimateriaalia. Tietyissä tilanteissa on kuitenkin mahdollista myös käyttää fyysisessä muodossa olevaa materiaalia. Tällöin tekstin- ja tiedonlouhintaa harjoittavalla toimijalla tulee olla joko omassa hallussaan printtikopiot taikka pääsy kolmannen hallussa oleviin printtikopioihin. Luonteeltaan tämä vaihe on kuitenkin enemmän sopimus- ja esineoikeudellinen kuin immateriaalioikeudellinen. Kirjastot ja arkistot voivat esittää tiettyjä vaatimuksia ja ehtoja koskien vahingonkorvauksia, korvaussuorituksia, vakuutuksia, velvoitteita sekä erityisesti heidän vakuuttamista mahdollisten tekijänoikeusloukkauksien varalta. Ks. lisää *Sag* 2019, s. 47–48. Esimerkiksi *Google Books* -ratkaisuihin Google oli tehnyt kirjastojen kanssa sopimuksia, jotka mahdollistivat valtavan fyysisten kirjojen kopioimisen. Ks. lisää esimerkiksi Michiganin yliopiston ja Googlen välisestä sopimuksesta osoitteessa <https://www.lib.umich.edu/sites/default/files/services/mdp/um-google-cooperative-agreement.pdf>.

⁶⁷ *Levendowski* on tutkinut ansiokkaasti, miten tekijänoikeuden ja tekoälyn kehittymisen välillä oleva kitka edesauttaa kehitystä, jossa tekoäly joutuu hyödyntämään virheellistä dataa. Hän kuvaa tällaista dataa termillä ”*biased, low-friction data (BLFD)*”. Ks. lisää aiheesta *Levendowski* 2018, s. 589.

tekstin- ja tiedonlouhinnan eri sidosryhmien asemat ja sidosryhmien jakautuminen kahteen. Ensimmäinen sidosryhmä koostuu julkaisijoista, jotka ovat myös oikeudenhaltijoita, ja toinen sidosryhmä tekstin- ja tiedonlouhinnan harjoittajista, joita voivat olla esimerkiksi niin tieteelliset tutkijat, kulttuuriperintölaitokset, teknologiafirmat, dataa hallinnoivat yritykset, kuin myös uutispalveluiden tarjoajat. Riippuen julkaisijan toiminnan luonteesta, sen suorittamista investoinneista sekä intresseistä voi julkaisijalla olla painava tarve rajoittaa materiaaliin pääsyä.⁶⁸

Tämän vuoksi osa julkaisijoista on ottanut käyttöön erilaisia lisensointijärjestelmiä, jotka voivat rajoittaa sitä, missä määrin materiaalia voidaan hyödyntää.⁶⁹ Toisaalta osa julkaisijoista tarjoaa vapaan pääsyn materiaaliin ja sallii myös tekstin- ja tiedonlouhinnan ilman lisenssiä.⁷⁰ Pääsyä materiaaliin vaikeuttaa myös se, että tietyissä tilanteissa on haastavaa tulkita, mikä on oikea taho myöntämään lisenssin tekstin- ja tiedonlouhintaa varten.⁷¹ Mikäli julkaisijan suostumus vaaditaan, tulee sen olla nimenomainen ja etukäteen annettu.⁷² Julkaisijoiden jaukautuneet intressit ja näkökannat lisensointijärjestelmiä kohtaan ovat aiheuttaneet oikeudellista epävarmuutta sen suhteen, mitä materiaalia on lähtökohtaisesti ylipäänsä mahdollista hyödyntää tekstin- ja tiedonlouhinnassa.⁷³

⁶⁸ Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: tekninen analyysi 2018, s. 3–4 sekä Euroopan komission tutkimuksen ja innovoinnin pääosaston asiantuntijalausunto 2014, s. 13.

⁶⁹ Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: oikeudellinen analyysi 2018, s. 4 ja 20.

⁷⁰ Esimerkiksi Elsevier tarjoaa kaikille sisäänkirjautuneille palvelunsa käyttäjille mahdollisuuden käyttää tietokantansa artikkeleita ja kirjoja tekstin- ja tiedonlouhintaa varten. Oxford University Press sekä Springer antavat tietokantansa käyttöön ilman lisenssiä ei-kaupallista tekstin- ja tiedonlouhintaa varten. Ks. Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: tekninen analyysi 2018, s. 3–4 sekä Euroopan komission tutkimuksen ja innovoinnin pääosaston asiantuntijalausunto 2014, s. 13.

⁷¹ Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: tekninen analyysi 2018, s. 4.

⁷² Euroopan unionin oikeuskäytännössä tämä on määritelty seuraavissa ratkaisuissa asia C-301/15 *Soulier ja Doke*, kohta 34 sekä asia C-314/12 *UPC Telekabel Wien*, kohdat 24 ja 25. Varsinkin orvot teokset tuottavat omat haasteensa. Nimittäin myös orvot teokset voivat saada suojaa joko tekijänoikeuden taikka lähioikeuksien nojalla. Ongelmallista on kuitenkin, kun näiden teoksien oikeudenhaltijat eivät ole tiedossa eikä näin ollen suostumusta voida hankkia. Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: tekninen analyysi 2018, s. 4 sekä Rosati 2019, s. 9.

⁷³ Aihealueesta on laajalti kirjoitettu oikeudellisessa kirjallisuudessa ja kannanottoja on myös otettu huomioon lainvalmisteluesitöissä. Katso esimerkiksi Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: tekninen analyysi 2018, s. 2–4; Euroopan komission tutkimuksen ja innovoinnin pääosaston asiantuntijalausunto 2014, s. 15–24; Triaille – de Meeûs d’Argenteuil – de Francquen 2014, s. 18–27; sekä OpenForum Academy 2015. Lisäksi myös Euroopan parlamentin eri valiokunnat kommentoivat tekijänoikeusdirektiiviä sen valmisteluvaiheessa sekä esittivät näkemyksensä lisensointiehtoihin liittyen, katso esimerkiksi seuraavat lausunnot: Sisämarkkina- ja kuluttajansuojavaliokunnan lausunto; Teollisuus-, tutkimus- ja energiavaliokunnan lausunto; Kulttuuri- ja koulutusvaliokunnan lausunto sekä Kansalaisvapauksien sekä oikeus- ja sisäasioiden valiokunnan lausunto.

Lisensointijärjestelyiden lisäksi julkaisijat voivat rajoittaa pääsyä materiaaliin niin kutsutuilla teknologisilla rajoituksilla (Technological Protection Measures, ”TPM”).⁷⁴ Rajoitukset voivat olla erilaisia, mutta niiden tarkoituksena on joko estää täysin pääsy materiaaliin taikka rajata pääsy vain tietyille toimijoille. Esimerkkejä TPM-rajoituksista ovat teknologiat, jotka vaativat käyttäjää tunnistautumaan taikka varmentavat käyttäjän IP-osoitteen.⁷⁵ Lisäksi tekijänoikeusdirektiivin 3 ja 4 artikloilla on vaikutusta siihen, miten pääsyä materiaaliin voidaan mahdollisesti rajoittaa. Tekijänoikeusdirektiiviin liittyvää kokonaisuutta tarkastellaan jäljempänä yksityiskohtaisemmin luvussa 5.

Mikäli materiaaliin on saatu pääsy – joko sen takia, että siihen on vapaa pääsy tai oikeudenhaltijalta on saatu lisenssi, eikä pääsyä ole myöskään rajoitettu TPM-rajoituksilla – voidaan siirtyä tekstin- ja tiedonlouhinnan toiseen vaiheeseen, jossa materiaali muunnetaan koneluettavaan muotoon ja mahdollisesti ladataan tarpeellisille alustoille. Tähän vaiheeseen sisältyy myös tarvittavan materiaalin kopiointia.⁷⁶ Kopiointiin liittyy monesta eri lainsäädännöllisestä näkökulmasta ongelmia.⁷⁷ Tässä tutkielmassa tarkastellaan kuitenkin kopiointia ainoastaan tekijänoikeuden näkökulmasta. Tähän ongelmaan palataan tarkemmin luvussa 4.

Kolmas vaihe, louhinta, sisältää analyttisen prosessoinnin, jonka tarkoituksena on tuottaa kerätystä ja kopioidusta materiaalista niin kutsuttuja empiirisiä havaintoja, eli muuntaa tekstiä dataksi ja löytää uutta tietoa.⁷⁸ Louhintavaiheeseen voi sisältyä tekstin- ja tiedonlouhintateknologiasta riippuen myös kerätyn materiaalin esikäsittelyä kahdessa eri alatasossa. Ensimmäiseen alatasoon sisältyy materiaalin jäsentelyä sekä puhdistamista virheistä ja epäolennaisesta tiedosta.⁷⁹

Kun haluttu data on saatu oikeaan muotoon, voidaan siirtyä toiseen alatasoon, joka voi sisältää esimerkiksi sanaistamista (”tokenization”)⁸⁰, synonyymien tunnistamista, sääntöperusteista lähestymistä, tilastollista lähestymistä, luokittelua taikka sana-analyysia.⁸¹ Tapahtuva

⁷⁴ Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: oikeudellinen analyysi 2018, s. 4 ja 20.

⁷⁵ Tekijänoikeusdirektiivi, resitaali 16.

⁷⁶ Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: oikeudellinen analyysi 2018, s. 5.

⁷⁷ Edellä rajauksia koskevassa luvussa 1.2 on käsitelty tarpeellisilta osin muita näkökulmia.

⁷⁸ Sag 2019, s. 57–59.

⁷⁹ Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: tekninen analyysi 2018, s. 6 sekä Rosati 2019, s. 13–14.

⁸⁰ Kyseinen termi on määritelty tarkemmin jäljempänä luvussa 2.2.6.

⁸¹ Rosati 2019, s. 14.

analysointi voi tapahtua joko täysin automatisoidusti tai vaatia ihmisen toimintaa.⁸² Esikäsittelyn jälkeen dataa voidaan analysoida. Analysoinnin tavoitteena on yhdistellä dataa uudella tavalla, jotta voidaan tunnistaa uusia malleja, suuntauksia tai korrelaatioita.⁸³

Edellä kuvatun kolmiportaisen prosessin mukaisesti eri tietolähteissä olevasta materiaalista voidaan tuottaa uutta informaatiota. Seuraavaksi tutkitaan, miten tekstin- ja tiedonlouhintateknologiaa hyödynnetään hakukonetoiminnassa. Esimerkkinä käytetään maailman tunnetuinta hakukonetta – Googlea.⁸⁴

2.2 Google-hakukoneen toiminnan logiikka

2.2.1 Yleisesti

Viimeisen kahdenkymmenen vuoden aikana hakukoneet ovat edesauttaneet tieteen, tutkimuksen sekä yhteiskuntamme rakenteiden kehitystä, joka on ollut mullistavaa. Hakukoneet ovat auttaneet niin yrityksiä, tutkimuslaitoksia kuin myös yksilöitä pääsemään käsiksi rajattomaan määrään tietoa ja ideoita. Tämä on mahdollistanut sellaisen tiedon tutkimisen, johon aikaisemmin pääsy on ollut vaikeaa tai jopa mahdotonta. Lienee itsestäänselvää, että tämä on edesauttanut innovointia, tieteellisiä läpimurtoja sekä muutoksia yhteiskunnan perinteisissä rakenteissa.

Google tulee varmasti monella ensimmäisenä mieleen, kun mainitaan hakukoneet. Google on vaikuttanut pysyvällä tavalla siihen, miten tarkastelemme tänä päivänä maailmaa. Google ei ole nykyään pelkästään yritys tai internetissä oleva yksinkertainen palvelu, vaan se on verbi, ilmiö

⁸² Jos toiminta on täysin automatisoitua, tulee varsinkin tekijänoikeudellisesta näkökulmasta tarkasteltavaksi se, kuinka paljon automatisoidussa prosessissa on hyödynnetty materiaalissa olevaa alkuperäistä ilmaisua. Jos prosessointi on puolestaan merkittävässä määrin vaatinut ihmisen toimintaa, on todennäköistä, että tämän henkilön omat käsitykset, asenteet, tavoitteet ovat voineet vaikuttaa prosessointiin ja siihen, missä määrin alkuperäistä ilmaisua on hyödynnetty. *Sag* 2019, s. 57–59.

⁸³ Katso lisää aiheesta *Sag* 2019, s. 59.

⁸⁴ Nykypäivänä Google dominoi maailmanlaajuisesti hakukonemarkkinoita ja siitä on tullut jo eräänlainen synonyymi kaikille muille hakukoneille. *Tusikov* 2016, s. 116.

ja miljardien eurojen arvoinen digitaalinen kokonaisuus, joka on digitalisoitunut lähestulkoon jokaisen elämäntilanteemme.⁸⁵ Sanakirjamääritelmän mukaan googlaus tarkoittaa sanojen kirjoittamista Googlen hakukoneeseen tarkoituksena löytää tietoa jostakin tai jostakusta.⁸⁶ Tämä määritelmä ei kuitenkaan riitä antamaan ymmärrystä siitä, miten hakukoneet soveltavat tekstin- ja tiedonlouhintaa tai miten ne ovat yhteydessä tekijänoikeuteen.

Tässä luvussa 2.2 on tarkoituksena siirtyä yksityiskohtaisemmalla tasolla tutkimaan, miten hakukoneeseen syötettyjen sanojen avulla voidaan löytää uutta tietoa sekä mihin peruseräisiin hakukonetoimintoihin ylipäänsä kytkeytyy. Näihin kysymyksiin vastaaminen tietotekniikan avulla auttaa vastaamaan *Law and Technology* -tutkimusmetodin ensimmäiseen osa-alueeseen, eli siihen, miten tekstin- ja tiedonlouhintateknologia on ylipäänsä yhteydessä tekijänoikeuteen. Lisäksi tässä luvussa tuotettua tietoa hakukoneiden toimintalogiikasta hyödynnetään *Law and Technology* -tutkimusmetodin muihin osa-alueisiin vastattaessa jäljempänä tässä tutkielmassa.

Hakukoneiden toimintalogiikkaa tarkasteltaessa on syytä aluksi poistaa yleinen virheellinen ennakkokäsitys siitä, että hakusanan syöttäminen hakukoneeseen tarkoittaisi sitä, että hakukone kävisi läpi kaikki maailman internetsivut suorittaessaan hakua. Todellisuudessa haku suoritetaan jo valmiina olevasta indeksistä.⁸⁷ Seuraavissa luvuissa käsitellään sitä, mitä indeksillä tarkoitetaan ja miten data on koottu tehokkaalla tavalla indekseihin. Lisäksi tutkitaan, miten binäärihaku toimii sekä minkälaisia algoritmeja hakukoneet hyödyntävät.

Ennen kuin syvennyttään tietotekniikan maailmaan ja hahmottamaan hakukoneiden toimintalogiikkaa, avataan vielä muutamia aihealueen kannalta keskeisiä termejä. Indeksit kuuluvat tietokannan (*”data collection”*) määritelmän alle ja tietokannat koostuvat arvoista (*”record”*), joiden yksilöimiseen käytetään niin kutsuttuja avaimia (*”key”*). Haku indeksistä perustuu siihen, että algoritmi hakee arvoa, joka vastaa sille annettua avainta. Haun suorittamisessa onkin tavoitteena löytää oikea arvo. Olennaista soveltuvien hakumenetelmien valinnassa on valita menetelmä, joka mahdollistaa hakujen nopean toteuttamisen.⁸⁸ Pidemmittä puheita, siirrytään määrittelemään ensiksi yleisellä tasolla luvuissa 2.2.2–2.2.3, miten hakuja tehdään tietokannoista sekä miten arvot ja avaimet linkittyvät toisiinsa. Tämän jälkeen luvussa 2.2.4 siirrytään

⁸⁵ Googlen tavoitteena onkin organisoida maailmassa oleva tieto ja saattaa se universaalisti kaikkien saataville ja käytettäväksi. Katso Googlen nettisivut osoitteesta <https://www.google.com/intl/en-GB/about/>.

⁸⁶ Katso määritelmä Oxford Dictionary -sanakirjasta osoitteesta https://www.oxfordlearnersdictionaries.com/definition/american_english/google.

⁸⁷ Niemi – Kallio – Rissanen, osoitteessa <https://www.netlab.tkk.fi/opetus/s38118/s00/tyot/42/>.

⁸⁸ *Spraul* käyttää esimerkkinä arvojen ja avaimien välisestä yhteydestä sanakirjaa. Avain vastaa sitä sanaa, jota etsitään sanakirjasta, kun puolestaan arvo vastaa sanakirjassa olevaa määritelmää. Ks. lisää *Spraul* 2015, s. 146.

tarkemmin käsittelemään hakukoneiden toimintaa. Aloitetaan kuitenkin ensiksi datan järjestämisestä.

2.2.2 Datan järjestäminen

Tehokkaan dataa sisältävän järjestelmän lähtökohta on, että järjestelmän hyödyntämä data on järjestyksessä. Tilannetta havainnollistamaan voidaan käyttää esimerkkinä yksittäisen teoksen etsimistä kirjastosta. Kirjojen järjestely aihepiireittäin ja aakkosittain, auttaa löytämään halutun teoksen nopeammin kuin tilanteessa, jossa kaikki tuhannet kirjaston kirjat käytäisiin yksittäin järjestyksessä läpi. Sama logiikka pätee myös tietokantahakuihin ja niiden hakuperiaatteisiin.⁸⁹

Esimerkinmukaista hakumenetelmää, jossa käydään jokainen kirja järjestyksenmukaisesti läpi, kutsutaan perättäishauksi (*“sequential search”*). Tietokoneohjelmissa käytettävät haut saattavat kuitenkin käsittää miljoonia arvoja, minkä vuoksi perättäishaun toteuttaminen ei ole tehokasta eikä edes mahdollista. Tämän takia on välttämätöntä järjestää tietokannat tiettyyn järjestykseen. Tätä kutsutaan nimellä lajittelu (*“sorting”*). Lajittelu on mahdollista toteuttaa usealla tavalla, mutta tässä tutkielmassa otetaan käsitteeseen valintalajittelu (*“selection sort”*) sekä pikalajittelu (*“quick sort”*).⁹⁰ Lähempään tarkasteluun otettavien lajittelumenetelmien taustalla vaikuttavat erilaiset toimintalogiikat. Tutkielmassa käsitellään näitä molempia menetelmiä, jotta saadaan ymmärrys siitä, miksi toinen menetelmä soveltuu paremmin tietokantahakuihin. Seuraavaksi siirrytään käsittelemään molempia menetelmiä ja niiden soveltumista tietokantahakuihin.

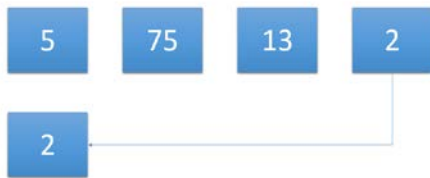
2.2.2.1 Valintalajittelu

Valintalajittelussa on tarkoituksena etsiä taulukosta pienin alkio (*“element”*), joka kopioidaan uuteen taulukkoon ensimmäiseksi alkioksi. Tämän jälkeen etsitään seuraavaksi pienin alkio, joka kopioidaan samaiseen taulukkoon toiseksi alkioksi. Näin jatketaan, kunnes kaikki alkupe-
räisen taulukon alkiot on käyty läpi ja kopioitu pienimmästä suurimpaan uuteen taulukkoon.⁹¹ Alla on havainnollistava kuvaus alkioiden lajittelusta suuruusjärjestykseen valintalajittelulla.

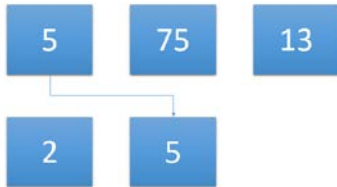
⁸⁹ Spraul 2015, s. 146.

⁹⁰ Spraul 2015, s. 146.

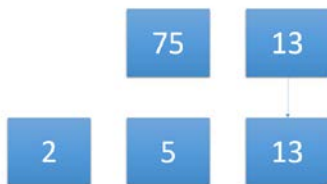
⁹¹ Spraul 2015, s. 146.



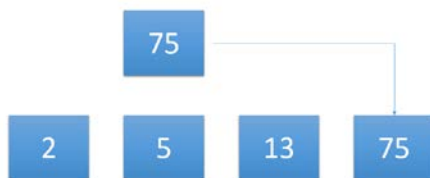
Kuvio 1.



Kuvio 2.



Kuvio 3.



Kuvio 4.

Kuviossa 1 valitaan pienin alkio, eli numero 2, ja kopioidaan se alla olevalle riville ensimmäiseksi alkioksi. Kuviossa 2 jatketaan samaa menetelmää ja kopioidaan seuraavaksi pienin alkio, eli numero 5. Kuviossa 3 kopioidaan jälleen pienin alkio, eli numero 13. Samaa toimintaperiaatetta jatketaan kuviossa 4 vielä viimeisimmän alkion, eli numeron 75, kohdalla. Tämän jälkeen on saatu uusi taulukko, jossa alkioit 5, 75, 13, 2 ovat uudessa järjestyksessä pienimmästä suurimpaan, eli: 2, 5, 13, 75. Valintajaoittelu tuottaa toivotussa järjetyksessä olevan taulukon, mutta edellä esitetystä käy myös ilmi, että järjestely vaati jokaisen alkion läpikäymistä ja siirtämistä. Kun kyseessä on suurempi taulukko, vaiheiden määrä kasvaa nopeasti huomattavan

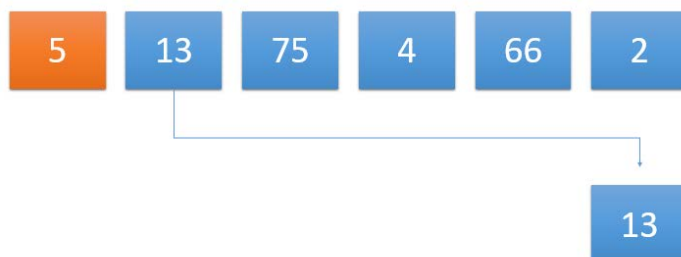
suureksi, eikä jaottelu ole tehokasta. Tämän takia tarvitaan lajittelumenetelmä, joka tuottaa vähemmällä vaiheilla halutun lopputuloksen. Pikalajittelu täyttää tämän vaatimuksen, ja sitä käsitellään seuraavaksi.

2.2.2.2 Pikalajittelu

Pikalajittelu on valintalajittelua tehokkaampi algoritmi, joka pystyy lajittelemaan koko tietokantaa jokaisen vaiheen aikana jakamalla ongelman pienemmiksi osaongelmiksi. Pikalajittelussa vaiheiden vähäisempi määrä ja tehokkuus ovatkin edesauttaneet sitä, että pikajaottelua suositaan käytettäväksi tietokoneohjelmissa.⁹²

Pikalajittelun toimintalogiikka perustuu siihen, että joukosta valitaan tietty alkio vertailukohdaksi ("pivot"), jonka suhteen joukko jaetaan vertailukohtaa suurempiin ja pienempiin alkioihin. Alkiot, jotka ovat vertailukohtaa pienempiä siirretään taulukon alkuun ja sitä suuremmat taulukon loppuun. Tämän jälkeen on käsillä tilanne, jossa vertailukohdaksi valitun alkion kummallakin puolella on kaksi järjestämätöntä joukkoa. Järjestämättömiä joukkoja kutsutaan alijoukoiksi. Näistä alijoukoista voidaan edelleen valita uudet vertailukohdat, joiden mukaan alkiot jaetaan jälleen kahteen joukkoon: vertailukohtaa suurempiin ja vertailukohtaa pienempiin alkioihin. Uuden vertailukohdan avulla joukon jakamista jatketaan niin kauan kunnes alijoukoissa ei ole enempää kuin yksi alkio. Tällöin alkuperäinen joukko alkioita on saatu lajiteltua järjestykseen.⁹³

Alla kuvioden 5–11 avulla on havainnollistettu pikalajittelua. Vertailukohdan valitsemiselle on olemassa eri menetelmiä ja alla esitettävässä esimerkissä on päädytty menetelmään, jossa vertailukohdaksi valitaan aina joukon ensimmäinen alkio.⁹⁴



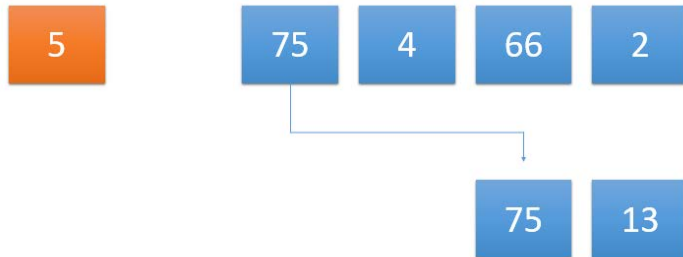
⁹² Spraul 2015, s. 147.

⁹³ Spraul 2015, s. 147–149.

⁹⁴ Vertailukohdaksi voidaan valita joukon ensimmäinen taikka viimeinen alkio. Vertailukohta voidaan myös valita satunnaisesti taikka valita joukon mediaani. Ks. lisää vertailukohdan valinnasta *Geeks for Geeks* -nettisivulta osoitteesta <https://www.geeksforgeeks.org/quick-sort/>.

Kuvio 5.

Vertailukohdaksi on valittu joukon ensimmäinen alkio, numero 5 (merkitty oranssilla). Seuraavaksi jaetaan alkiot vertailukohdan molemmille puolille. Seuraava alkio joukossa, numero 13, siirretään taulukon oikeaan reunaan, sillä se on suurempi kuin 5.



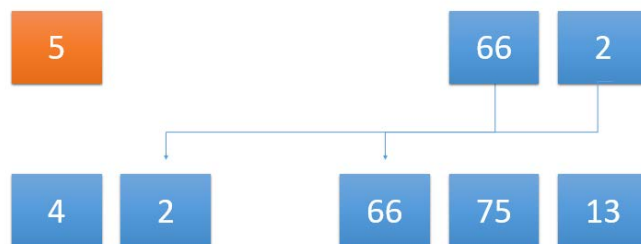
Kuvio 6.

Myös seuraava alkio, eli numero 75, on vertailukohtaa suurempi, joten sekin siirretään taulukossa oikealle.



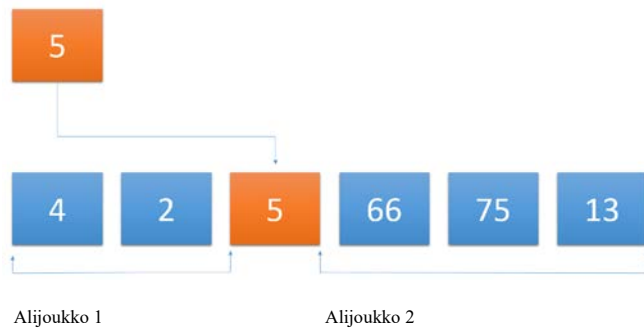
Kuvio 7.

Järjestyksessä seuraava alkio, numero 4, onkin pienempi kuin vertailukohta. Tällöin alkio siirretään taulukossa vasempaan reunaan.



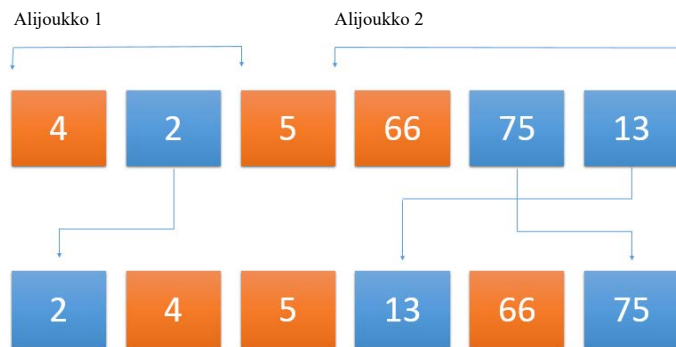
Kuvio 8.

Tässä kuviossa lajitellaan loput alkiot, 66 ja 2, samalla logiikalla omille paikoilleen taulukkoon.



Kuvio 9.

Tämän jälkeen siirretään vertailukohta tyhjälle paikalle ja saadaan uusi taulukko, joka ei kuitenkaan ole vielä järjestelty – vaikkakin se on järjestellymmässä muodossa kuin alkuperäinen. Vertailukohtan vasemmalle puolelle muodostuu alijoukko 1 ja oikealle puolelle alijoukko 2. Koska molemmissa alijoukoissa on enemmän kuin yksi alkio, tulee jakamista jatkaa.



Kuvio 10.

Valitaan alijoukosta 1 vertailukohdaksi sen ensimmäinen alkio, numero 4 (merkitty oranssilla), sekä valitaan alijoukosta 2 vertailukohdaksi sen ensimmäinen alkio, numero 66 (merkitty oranssilla). Uuden vertailukohtan valitsemisen jälkeen on alijoukossa 1 vain yksi vertailtava alkio ja se on numero 2. Siirretään tämä vertailukohtan 4 vasemmalle puolelle, sillä se on vertailukohtaa pienempi. Alijoukossa 2 on puolestaan jäljellä kaksi vertailtavaa alkioita uuden vertailukohtan valitsemisen jälkeen. Alkio, numero 75, on suurempi kuin vertailukohta, joten se siirretään vertailukohtan oikealle puolelle. Alkio, numero 13, on pienempi, joten se siirretään puolestaan vertailukohtan vasemmalle puolelle.



Kuvio 11.

Viimeisimmästä kuviosta nähdään, ettei alijoukoissa 1–3 ole enää enempää alkioita kuin yksi, mikä tarkoittaa, että lajittelu on loppuunsaatettu ja joukko on järjestetty.

Edellä käyty esimerkki pikalajittelusta osoittaa, että kyseisen lajittelumenetelmän avulla voidaan lajitella joukko dataa järjestykseen verrattain tehokkaalla tavalla. Täten myös tietokoneohjelmat käyttävät usein kyseistä menetelmää datansa järjestämiseen.⁹⁵ Seuraavaksi tarkastellaan, miten järjestellystä datasta voidaan löytää tiettyjä arvoja.

2.2.3 Binäärihaku

Kun data on saatu järjestettyä, pystyvät tietokoneohjelmien hakualgoritmit (*“searching algorithms”*) tekemään tietokannasta hakuja ja löytämään haluttuja arvoja.⁹⁶ Yksi tunnettu hakualgoritmi on binäärihaku (*“binary search”*), jota voidaan käyttää taulukkoon, joka on edellä esitetyllä tavalla järjestetty suuruusjärjestykseen. Binäärihaun nimi tulee siitä, että haun tarkoituksena on aina valita kahdesta vaihtoehdosta ja päätyä lopulta tilanteeseen, jossa on vain yksi vaihtoehto jäljellä.⁹⁷ Seuraavaksi havainnollistetaan yksinkertaistetun esimerkin avulla kuviossa 12–16 miten binäärihaku toimii.



Kuvio 12.

Kuviossa nähdään järjestetty joukko. Havainnollistavan esimerkin tarkoituksena on löytää arvo, jonka avain on 33 (merkitty nuolella). Otetaan lähtökohdaksi, ettei joukosta tiedetä muuta kuin, että se sisältää avaimen 33 ja että avaimet ovat joukossa suuruusjärjestyksessä.



Kuvio 13.

Ensimmäisessä vaiheessa kuviossa 13 tarkastellaan joukon keskellä olevaa arvoa. Huomataan, että kyseisen arvon avain ei ole 33, vaan 59, joka on suurempi kuin etsitty avain. Tämän yhden avaimen valinta ei ainoastaan eliminoi kyseistä avainta, vaan samalla voidaan myös tehdä päätelmä, että keskellä valitusta avaimesta katsottuna kaikki avaimet oikealla voidaan jättää myös huomioimatta, sillä ne ovat suurempia kuin 33.



Kuvio 14.

Seuraavassa vaiheessa tarkastellaan neljättä arvoa oikealta. Kyseisen arvon avain on 27. Tämä on puolestaan pienempi kuin 33. Tiedetään siis, että avain 33 on jossain avaimien 27 ja 59 välissä.

⁹⁵ Spraul 2015, s. 146.

⁹⁶ Ks. lisää *Geeks for Geeks* -nettisivulta osoitteesta: <https://www.geeksforgeeks.org/searching-algorithms/>.

⁹⁷ Spraul 2015, s. 151.



Kuvio 15.

Valitaan seuraavaksi viides arvo oikealta. Tämä paljastuu arvoksi, jonka avain on 47.



Kuvio 16.

Koska tiedetään, että avaimien 27 ja 47 välissä on vain yksi avain, voidaan tehdä päätelmä siitä, että etsitty avain löytyy näiden avaimien välistä. Jos kyseisellä arvolla ei olisi ollut avainta 33, olisi tämä tarkoittanut sitä, ettei kyseisessä taulukossa olisi ollut laisinkaan etsittyä avainta.

Jokainen vaihe binäärihaussa eliminoi puolet arvoista, minkä takia binäärihaku on tehokkaampi kuin aikaisemmin luvussa 2.2.2 esitelty perättäishaku, jossa jokainen avain käydään läpi. Jos tutkittavan tietokannan määrä kaksinkertaistetaan, tarkoittaa se perättäishaussa sitä, että vaiheiden määrä kaksinkertaistuu. Binäärihaussa tämä tarkoittaa kuitenkin vain yhden vaiheen lisäämistä. On siis selkeää, että binäärihakuja käytetään suurissa tietokannoissa, tietokoneohjelmissa ja etenkin hakukoneissa.⁹⁸

2.2.4 Tietokannan muistirakenne ja indeksointi

Edellä on käsitelty sitä, miten data tulisi järjestää ja miten voidaan tehokkaasti hakea avaimilla tiettyjä arvoja tietokannoista. Tämän lisäksi tulee kuitenkin hahmottaa, miten muisti kannattaa allokoita, jotta tietokoneohjelma voi suorittaa tehokkaasti hakuja tietokannasta. Seuraavaksi käsitellään kolme eri tapaa allokoita muistia ja tutkitaan, mikä tapa sopii parhaiten hakukoneiden toimintaan.

Ensinnäkin on mahdollista hyödyntää kiinteää muistia (*fixed-size storage*), jossa allokoitaan jokaisella arvolla sama määrä muistia. Kiinteän muistin käyttäminen sopii tilanteisiin, joissa data on aina samankokoista taikka yksittäiset arvot ovat kooltaan pieniä. Esimerkkinä voidaan käyttää luottokortteja. Luottokortin numerosarja on kooltaan aina 16 numeroa. Näin ollen on siis helppoa tallentaa jokainen numero samankokoiselle muistipaikalle. Nimet luottokorteissa puolestaan ovat pituudeltaan eriäviä ja kirjainyhdistelmät poikkeavat toisistaan. Jos esimerkiksi luottokortissa enimmäismerkkimäärä sukunimelle olisi 15 merkkiä, sukunimi *Virtanen* käyt-

⁹⁸ *Spraul* 2015, s. 152.

täisi siitä kahdeksan merkkiä jättäen seitsemän merkkiä tyhjiksi, kun taas sukunimi *Kangaskorhonen* käyttäisi 14 merkkiä jättäen ainoastaan yhden merkin tyhjäksi. Nimien tallentamiseen vaadittava muistimäärä on kuitenkin vähäinen, eivätkä myöskään tyhjät merkit aiheuta paljoa hukkatilaa. Näin ollen nimet luottokorteissa tallennetaan kiinteänkokoisille muistipaikoille.⁹⁹

Toiseksi, muisti voi olla myös räätälöity juuri oikeankokoiseksi vastaamaan datan todellista muistin tarvetta. Tällöin puhutaan muuttuvasta muistista (*“variable-size storage”*). Otetaan esimerkiksi joukko MP3-tiedostoissa olevia musiikkikappaleita. Perinteinen radiosta kuuluva pop-kappale voi olla kooltaan 3MB, kun puolestaan jokin pitkä akustinen rock-kappale voi olla jopa 20MB kokoinen. Saman kiinteän muistin käyttäminen 3MB sekä 20MB kokoisiin kappaleisiin aiheuttaisi sen, että tallennustilaa jäisi paljon käyttämättä. Tästä johtuen musiikkikappaleiden tallentamiseen hyödynnetään muistia, joka on kooltaan juuri sen kokoinen mitä kappale vaatii.¹⁰⁰

Muuttuva muisti mahdollistaa tilan tehokkaan hyödyntämisen, mutta kiinteään muistiin perustuva allokointi on puolestaan tarpeen, jotta tietokoneohjelmat voivat tehdä tehokkaasti hakutoimintoja. Jos kaikki tiedostot tietokannassa ovat samankokoisia, voi tietokoneohjelma löytää nopeasti tietyn arvon tietystä sijainnista. Tämä johtuu siitä, että muistin sijainnit identifioidaan numeerisilla osoitteilla, muistipaikoilla (*“address”*). Jokainen tavu digitaalisessa muistissa, joko tietokoneen keskusmuistissa, muistitikulla taikka kiintolevyasemalla, pystytään tarkalleen paikantamaan muistipaikkansa perusteella. Jos tietokoneen keskusmuisti on esimerkiksi 3GB:n kokoinen, ovat kaikki tavut numeroituja nolasta kolmeen biljoonaan. Joukkiot kiinteään muistiin tallennettuja arvoja tallennetaan vierekkäin, mikä helpottaa näiden löytämistä. Otetaan esimerkiksi tietokanta, jossa on 10 arvoa, joista jokainen on 20 tavun kokoinen ja tietokanta alkaa numerosta 1000. Tämä tarkoittaa tällöin sitä, että ensimmäinen arvo löytyy muistipaikasta 1000, toinen puolestaan 1020, kolmas 1040 ja niin edelleen. Jokainen arvo voidaan löytää yksinkertaisesti laskemalla seuraavalla kaavalla – kerrotaan järjestysluku 20:llä ja lisätään 1000. Näin tietokoneohjelma voi löytää jokaisen arvon tietokannasta, joka on rakennettu kiinteälle muistille.¹⁰¹

Arvojen löytäminen nopeasti on keskeistä esimerkiksi edellä käsitellylle binäärihauulle. Ilman kiinteitä muistipaikkoja ainoa tapa löytää arvo tietystä muistipaikasta on käydä koko tietokanta

⁹⁹ *Spraul* 2015, s. 152.

¹⁰⁰ *Spraul* 2015, s. 152.

¹⁰¹ *Spraul* 2015, s. 152.

läpi arvo kerrallaan. On selvää, että tällainen perättäishaku on epätehokasta. Lähtökohtaisesti valitseminen kiinteän muistin ja muuttuvan muistin välillä tarkoittaa siis valintaa tehokkaan hakutoiminnan ja tehokkaan säilyttämisen välillä. On olemassa kuitenkin menetelmä, joka yhdistää nämä kaksi ominaisuutta. Tätä menetelmää kutsutaan indeksoinniksi. Seuraavaksi tässä tutkielmassa keskitytään yksityiskohtaisemmin tähän kolmanteen tietokannan muistirakenteseen.¹⁰²

Indeksi on taulukko, johon on koottu arvojen avaimia ja muistipaikkoja. Muistipaikat ovat tallennettuja binäärinumeroina kiinteällä määrällä tavuja. Koska muistipaikat ovat kiinteän kokoisia, voidaan muistipaikat ja avaimet tallentaa yhdessä kiinteään muistiin rakentuvaan indeksiin, josta voidaan etsiä tehokkaasti tiettyä arvoa esimerkiksi binäärihaulla. Loput jokaisen arvon datasta on tallennettu muuttuvaan muistiin. Indeksoinnissa data siis jaetaan kahteen eri paikkaan: kiinteille muistipaikoille sekä muuttuville muistipaikoille. Indeksien taulukosta löytyvät kiinteille muistipaikoille tallennetut avaimet ja osoitteet. Arvot on puolestaan allokoitu muuttuville muistipaikoille eli muistia on varattu juuri sen verran kuin on tarpeen.¹⁰³

Indeksointi mahdollistaa myös sen, että itse indeksi ja arvot tallennetaan eri muistilaitteisiin. Indeksi voi siis sijaita esimerkiksi tietokoneen keskusmuistissa ja arvot kiintolevyllä. Koska ainoastaan itse indeksiä käytetään hakutoiminnassa, ovat haut nopeita eivätkä vaadi kuin vähimmäismäärän keskusmuistista. Tietokanta voi myös sisältää useita indeksejä eri hakukriteereille, ja koska arvo on omassa muistissaan, johon viitataan pelkällä muistipaikalla, ei useiden indeksien käyttäminen vaikuta merkittävästi tietokannan muistin määrään.¹⁰⁴ Muistirakenteen nopeuden ja tehokkuuden vuoksi indeksiä käytetään myös internetin hakukoneissa. Jotta hakukoneen indeksi toimisi vielä tehokkaammin ja joustavasti, käytetään indeksien rakentamisessa hajautusfunktioita, joita käsitellään seuraavaksi.

2.2.5 Hajautusfunktiot

Tietokannan käyttötarkoitus asettaa myös vaatimuksia sille, pitääkö tietokantaan tehdä muutoksia. Puhutaan staattisista ja dynaamisista tietokannoista. Staattisella tietokannalla viitataan tietokantaan, joka muuttuu harvoin. Esimerkiksi kun tekstinkäsittelyohjelman oikolukutoi-

¹⁰² Spraul 2015, s. 153–154.

¹⁰³ Spraul 2015, s. 153–154.

¹⁰⁴ Spraul 2015, s. 153–154.

minto ehdottaa muutosehdotusta virheellisesti kirjoitetulle sanalle, perustuvat oikolukutoiminnon ehdotukset staattiseen tietokantaan.¹⁰⁵ Internetin hakukoneet tarvitsevat puolestaan dynaamisen tietokannan, sillä hakujen pitää olla ajan tasalla. Esimerkiksi Googlen *Caffeine*-indeksiin lisätään satojatuhansia gigatavuja dataa päivittäin.¹⁰⁶

Tarvitaan tietorakenne, jonka sisältöä voidaan muuttaa jatkuvasti päivittämällä, poistamalla tai lisäämällä dataa.¹⁰⁷ Kuitenkin jos poistot ja lisäykset ovat yleisiä, datan uudelleen lajitteluun voi mennä huomattavasti aikaa, mikä puolestaan heikentää haun suorittamisen tehokkuutta. Tietorakenne, joka kykenee mukautumaan tällaisiin lisäyksiin ja poistoihin, on hajautustaulu (*“hash table”*). Yksi yleisimmistä hajautustauluista on hajautusfunktio (*“hash function”*).¹⁰⁸

Hajautusfunktioiden toiminta perustuu siihen, että määritellään tietty matemaattinen kaava, eli hajautusfunktio, jolla kaavalle annettu arvo voidaan muuttaa tietyllä välillä olevaksi kokonaisluvuksi. Tätä lukua kutsutaan termillä hajautusarvo (*“hash value”*). Hajauttamisen tarkoituksena on ikäänkuin tiivistää alkuperäinen arvo. Jotta hajautusarvoa voidaan pitää pysyvänä kuvauksena alkuperäisestä arvosta, edellytetään, että hajautusfunktion tulee toimia joka kerta samalla tavalla. Lisäksi hajautusfunktion tulisi antaa mahdollisimman tasaisesti eri kokonaislukuja.¹⁰⁹

Havainnollistamiseksi esitetään tilanne, jossa hajautusfunktio tuottaa aina hajautusarvon, joka vastaa kokonaislukua välillä 0–7. Tätä voidaan hyödyntää tallennettaessa arvoja hajautustauluksoon, jossa on korttipaikkoja kahdeksalle arvolle. Korttipaikka (*“slot”*) on paikka, johon arvo voidaan tallentaa. Jotta arvo voidaan tallentaa hajautustauluksoon, tulee arvon avain tiivistää. Tiivistäminen määrittää puolestaan sen, mitä korttipaikkaa voidaan käyttää. Oletetaan, että ollaan tallentamassa MP3-tiedostomuodossa olevia kappaleita ja kyseisten kappaleiden nimet toimivat avaimina.¹¹⁰ Hajautustaulukkojen toimintaa havainnollistetaan alla olevalla esimerkillä.

Kappaleen nimi	Hajautusarvo
----------------	--------------

¹⁰⁵ *Spraul* 2015, s. 154.

¹⁰⁶ Ks. *Google blog* 2010, osoitteesta: <https://googleblog.blogspot.com/2010/06/our-new-search-index-caffeine.html>.

¹⁰⁷ *Cormen – Leiserson – Rivest – Stein* 2009, s. 253.

¹⁰⁸ *Spraul* 2015, s. 20.

¹⁰⁹ *Schindelbauer* 2011, s. 195–200.

¹¹⁰ *Spraul* 2015, s. 154.

<i>Hard Rock Hallelujah</i>	3
<i>Aikuinen nainen</i>	2
<i>Aamu</i>	1
<i>Tässä talossa</i>	2

Kuvio 17.

Kappaleiden nimiä ja niiden hajautusarvoja esittävässä kuviossa on hajautusfunktiolla tiivistetty kappaleiden nimet hajautusarvoiksi. Tässä esimerkissä hajautusfunktio perustuu yksinkertaisesti siihen, että lasketaan kappaleen nimen sisältämien sanojen kappalemäärä yhteen. Kappaleen nimessä ”*Hard Rock Hallelujah*” on kolme sanaa, joten sen hajautusarvoksi tulee 3, kappaleissa ”*Aikuinen nainen*” ja ”*Tässä talossa*” on kaksi sanaa, joten niiden hajautusarvoiksi tulee 2. Kappaleessa ”*Aamu*” on vain yksi sana, joten sen hajautusarvo on 1.

0	0		
1	1	<i>Aamu</i>	Muistipaikka
2	1	<i>Aikuinen nainen</i>	Muistipaikka
3	1	<i>Hard Rock Hallelujah</i>	Muistipaikka
4	0		
5	0		
6	0		
7	0		

Kuvio 18.

Kuvio esittää hajautustaulukkoa, johon on lisätty kolme ensimmäistä kappaletta kuvioista 17. Ensimmäinen kolumni esittää kokonaislukuja välillä 0–7, eli hajautustaulukon korttipaikkoja. Toisessa kolumnissa puolestaan esitetään tavuina, onko kyseinen korttipaikka vapaana vai ei. Jos korttipaikka on varattu, merkitään se numerolla 1. Jos se on puolestaan vapaa, merkitään se numerolla 0. Kolmas kolumni esittää kappaleiden nimiä eli avaimia. Neljäs kolumni sisältää tiedon muistipaikasta eli tiedon siitä, missä loput datasta on.

Hajautusfunktion toiminta ei itsessään vaadi etsimistä, vaan avaimen syöttäminen hajautusfunktion paljastaa, missä etsitty arvo sijaitsee. Jos kyseisessä korttipaikassa ei ole arvoa, voidaan heti tunnistaa, ettei kyseinen tietokanta sisällä ollenkaan kyseistä arvoa. Hajautustaulukko mahdollistaa myös sen, ettei sen sisältämää dataa tarvitse lajitella. Tämän takia se soveltuu hyvin dynaamisiin tietokantoihin, kuten hakukoneiden tietokantoihin, jotka vaativat jatkuvaa päivittämistä.¹¹¹

¹¹¹ *Spraul* 2015, s. 155–156.

Kuten kuviosta 18 huomataan, siihen ei ole lisätty kappaletta ”*Tässä talossa*”, koska kyseisen kappaleen hajautusarvoksi muodostui sama kuin ”*Aikuinen nainen*” -nimisen kappaleen. Hajautusfunktion käyttäminen ei siis takaa sitä, että saataisiin aina erilaiset hajautusarvot kaikille avaimille. Käsillä olevaa tilannetta, jossa kahdella (tai useammalla) eri avaimella on sama hajautusarvo, kutsutaan kollisioksi (”*collision*”). Kollisiotilanteita varten tarvitaan sääntöjä. Yksinkertainen esimerkki säännöstä on, että kollisiotilanteen sattuessa tulee käyttää paikkaa, joka on seuraavaksi vapaana.¹¹²

0	0		
1	1	<i>Aamu</i>	Muistipaikka
2	1	<i>Aikuinen nainen</i>	Muistipaikka
3	1	<i>Hard Rock Hallelujah</i>	Muistipaikka
4	1	<i>Tässä talossa</i>	Muistipaikka
5	0		
6	0		
7	0		

Kuvio 19.

Kuviossa havainnollistetaan, miten voidaan hyödyntää sääntöä kollisiotilanteessa. Korttipaikka 2 on jo varattuna kappaleelle ”*Aikuinen nainen*”. Joten säännön mukaan kappale ”*Tässä talossa*”, sijoitetaan seuraavalle vapaalle korttipaikalle. Koska myös paikka 3 on varattu, sijoitetaan kappale korttipaikalle 4.

Kuten kuviosta 19 on nähtävissä, ratkaisi säännön käyttäminen kollisiotilanteen. Seurauksena on kuitenkin se, hajautustaulukon käyttö vaikeutuu. Kollisiosäännön käyttäminen aiheuttaa sen, ettei tietyn arvon löytäminen ole enää yksivaiheinen prosessi. Jokainen haku alkaa siitä, että tarkistetaan se korttipaikka, johon hajautusarvo viittaa. Tämän jälkeen haussa tarkistetaan jokainen korttipaikka, kunnes löydetään muistipaikka, jossa on oikea avain eli kappaleen nimi. Jos haussa päädytään tyhjiin korttipaikkaan, tarkoittaa tämä sitä, ettei kyseistä kappaletta, eli avainta, ole kyseisessä hajautustaulukossa.¹¹³

Kollisiot voivat myös aiheuttaa sen, että kappale on kaukana siitä korttipaikasta, jossa sen pitäisi hajautusarvon mukaan olla. Esimerkiksi tilanteessa, jossa kappale ”*Aikuinen nainen*” poistettaisiin taulukosta ja merkittäisiin kyseinen korttipaikka vapaaksi numerolla 0, ja tulisi tilanne, jossa haluttaisiin etsiä kappaletta ”*Tässä talossa*” hajautusarvolla 2, ei kappaletta löytyisi, koska haussa tarkistettaisiin ensiksi hajautusarvon 2 mukainen paikka, joka on merkitty vapaaksi. Haku loppuu tähän ja tulkitaan, ettei kyseistä kappaletta ole myöskään kyseisessä taulukossa.

¹¹² *Spraul* 2015, s. 155–156.

¹¹³ *Spraul* 2015, s. 155–156.

Tämän takia poistettaessa jokin kappale tulee merkitä sen korttipaikka edelleen varatuksi. Tällaista korttipaikkaa, joka on oikeasti tyhjä, mutta on kuitenkin merkitty käytössä olevaksi, nimitetään hautakiveksi (*"tombstone"*).¹¹⁴

Vaikka hautakivien käyttäminen mahdollistaa sen, että hakuja voidaan edelleen suorittaa datan poistamisen jälkeen, käyttävät hautakivet turhaan muistia. Lisäksi niiden käyttäminen ruuhkauttaa taulukkoja. Tämän takia hajautustaulukot tulee tietyn ajoin uudelleen hajauttaa (*"re-hash"*). Tilanteessa, jossa tietty prosentti hajautustaulukon korttipaikoista on käytössä, luodaan uusi isompi hajautustaulukko. Kuvatun kaltaisessa uudessa hajautustaulukossa vanhat hajautusarvot hajautetaan uudelleen.¹¹⁵

Kun sovelletaan edellä esitettyä Googlen hakukoneeseen ja sen laajaan indeksiin, voidaan hakukoneen käyttäjän syöttämiä hakusanoja kutsua avaimiksi, muistipaikkoja URL-osoitteiksi¹¹⁶ ja internetsivuja arvoiksi.¹¹⁷ Googlen indeksin tarkka koko ei ole tiedossa, mutta sen arvellaan olevan 100 beetatavun kokoinen. On selvää, että tarvitaan kehittyneitä hakumenetelmiä, jotta tällaisesta valtavasta indeksistä pystytään tekemään hakuja. Googlen indeksi on luonteeltaan invertoitu indeksi (*"inverted index"*). Se on siis taulukko, joka koostuu tekstielementeistä sekä osoittimista, jotka kertovat mitkä tiedostot sisältävät kyseisiä elementtejä. Tekstielementit ovat sanaistettuja (*"tokenize"*), jolla tarkoitetaan sitä, että sanat on tiivistetty niiden ydinsisältöön.¹¹⁸

Vaikka edellä esitetyt tekniikat havainnollistavat sitä, miten laajasta indeksistä voidaan tehdä hakuja ja miten indeksi toimii, ne eivät kuitenkaan selitä, miten tällainen massiivinen indeksi on alun perin koottu.¹¹⁹ Seuraavaksi perehdytään siihen, miten Google-hakukoneen indeksi luodaan ja sen jälkeen tutkitaan, mihin hakujen tarkkuus perustuu.

2.2.6 Googlen algoritmit

Kuten aikaisemmin on jo mainittu, käsillä oleva tiedon määrä nyky-yhteiskunnassa on ääretön. Edellä mainittuja menetelmiä tarvitaan, jotta tätä tietomassaa voidaan tehokkaalla tavalla käsi-

¹¹⁴ Spraul 2015, s. 155–156.

¹¹⁵ Spraul 2015, s. 155–156.

¹¹⁶ URL tulee englanninkielen sanoista *"Uniform Resource Locator"*.

¹¹⁷ Spraul 2015, s. 157.

¹¹⁸ Marsden 2018.

¹¹⁹ Spraul 2015, s. 157.

tellä internethauissa. Näiden menetelmien lisäksi Google hyödyntää toiminnassaan myös lukemattomia algoritmeja, joihin kätkeytyy sen miljardien arvoisen liiketoiminnan salaisuus. Yksi tärkeä algoritmien ryhmä liittyy Googlen indeksin kokoamiseen.

Google käyttää niin kutsuttuja googlebotteja ("robot") internetin lukemiseen, eli indeksinsä rakentamiseen. Nämä kyseiset googlebotit ovat indeksointirobotteja, joiden automatisoitu toiminta perustuu täysin algoritmeihin. Googlebottien suorittama indeksointiprosessi alkaa jo aikaisemmin indeksoiduista internetosoitteista. Aloittaessaan tietyltä internetsivulta ne laativat listan kaikista niistä internetosoitteista, jotka löytyvät kyseiseltä internetsivulta. Näitä linkkejä käytetään puolestaan löytämään lisää internetosoitteita uusille internetsivuille. Näin googlebotit keräävät laajaa internetosoitteiden listaa, ja lopputuloksena on valtava indeksi, jota ne indeksoivat jatkuvasti keräten uusia ja päivitettyjä internetsivuja talteen. Kun googlebotti on löytänyt uuden internetsivun, poimitaan internetsivuilta hakusanoja eli avaimia, joiden perusteella kyseinen internetsivu lisätään indeksissä vastaavien avaimien kohdalle.¹²⁰

Osana tätä indeksointiprosessia indeksointirobotit tekevät kopion hakukoneen omalle palvelimelle jokaisen löytämänsä internetsivun HTML-koodista. Kyseisiä kopioita pidetään palvelimella, kunnes kopio korvataan uudella kopiolla, joka luodaan, kun indeksointirobotti käy uudestaan kyseisellä internetsivulla.¹²¹ Varsinkin tämä vaihe hakukoneen toiminnassa on tämän tutkielman kannalta keskeinen ja siihen syvennyttään tarkemmin luvussa 3.1.

Googlebotin algoritmi päättää, miten sivustoja indeksoidaan, kuinka usein ja kuinka monta sivua kultakin sivustolta haetaan.¹²² Googlebotit eivät kuitenkaan voi kerätä internetosoitteita indeksiin kaikilta internetsivuilta, sillä internetosoitteet voivat olla internetsivun oman hakukoneen takana. Esimerkiksi uutissivustot eivät tarjoa suoraan internetosoitteita vanhoihin artikkeleihinsa, vaan uutissivustot tarjoavat pääsyn niiden omaan arkistoon, jota voidaan selata uutissivuston oman hakukoneen avulla. Tätä tärkeää materiaalia, johon ei ole pääsyä suoraan internetosoitteilla, kuvataan englanninkielisellä termillä *deep web*. Jotta tällaista materiaalia voidaan hyödyntää, käytetään sivukarttoja ("sitemap"), joiden avulla sivuston omistaja voi valita, miten Google indeksoi internetsivun.¹²³ Tiivistäen voidaan esittää, että Googlen indeksi rakentuu googlebottien laatimille kopioille ja niistä saataville tiedoille.

¹²⁰ Ks. *Google Support* osoitteesta <https://support.google.com/webmasters/answer/70897?hl=fi> sekä *Nevala* 2017.

¹²¹ *Peguera* 2012, s. 174 sekä <https://www.google.com/intl/fi/search/howsearchworks/crawling-indexing/>.

¹²² Ks. <https://www.google.com/intl/fi/search/howsearchworks/crawling-indexing/>.

¹²³ *Spraul* 2015, s. 157 sekä <https://www.google.com/intl/fi/search/howsearchworks/crawling-indexing/>.

Hakukoneen käyttäjälle ei kuitenkaan riitä se, että Google löytää indeksistään äärettömän joukon hakutuloksia, eli arvoja. Tarvitaan lisäksi hakukoneoptimointiin erikoistuneita algoritmeja, jotta googlebottien keräämästä laajasta indeksistä Googlen hakukone onnistuu löytämään hakijalle hänen haluamansa vastauksen, eli hakutuloksena listauksen internetosoitteita eri internet-sivuille.

Hakukoneoptimointi perustuu siihen, että algoritmit ensinnäkin analysoivat sitä, *mitä hakija yrittää hakea* sekä sitä, *mitä hänelle pitäisi näyttää*. Sanojen analysointi on yksi keskeisimmistä keinoista optimoida hakuja.¹²⁴ Sanojen analysoinnissa hyödynnetään kielellisiä malleja, jotka auttavat korjaamaan kirjoitusvirheitä sekä löytämään synonyymejä. Algoritmit tarkastelevat myös sitä, minkä tyyppistä tietoa hakija on etsimässä. Tämän analysoinnissa tulkitaan, onko hakusana sisällöltään laaja tai rajattu taikka sisältääkö hakusana kuvaavia ilmaisuja. Tietyt sanat ovat linkitettyjä toisiinsa Googlen indeksissä. Sanat pyritään ottamaan kokonaisuutena huomioon. Esimerkiksi hakusana ”*punainen*” tuottaa täysin erilaisen hakutuloksen kuin mitä sanayhdistelmä ”*punainen mekko*” tuottaa.¹²⁵

Toiseksi hakukoneoptimoinnissa hyödynnetään tuoreusalgoritmeja¹²⁶ ja sitä, mitä muut käyttäjät ovat hakeneet. Jos hakija käyttää hakukoneessa hakusanoja, joita on käytetty tietyllä aikavälillä useasti, tulkitsee algoritmi, että hakija on kiinnostunut tuoreesta sisällöstä. Esimerkiksi jalkapallon EM-karsintojen aikaan hakusanalla ”*huuhkaja*” hakutulokset saattoivat olla enemmän kytkeytyneinä jalkapalloon kuin suureen eurooppalaiseen pöllölajiin.¹²⁷

Sanojen analyysin jälkeen algoritmi on muodostanut avaimia, joilla se käy läpi indeksiansa ja se pyrkii löytämään vastaavuuksia.¹²⁸ Yhtenä kriteerinä vastaavuuksien löytämisessä on hakusanojen esiintymistiheys sekä paikat, missä hakusanat esiintyvät internetsivulla.¹²⁹ Nykyjärjestelmässä tämä ei kuitenkaan riitä, sillä internetsivuja on miljardeja eikä tietyn sanan esiintyminen tietyllä internetsivulla takaa vielä sitä, että sivu olisi relevantti. Lisäksi tämä mahdollistaisi järjestelmän hyväksikäyttämisen kirjoittamalla toistuvasti tietyn hakusanan internetsivulle, mikä

¹²⁴ Ks. <https://www.google.com/intl/fi/search/howsearchworks/algorithms/>.

¹²⁵ Spraul 2015, s. 159 sekä <https://www.google.com/intl/fi/search/howsearchworks/algorithms/>.

¹²⁶ Lue lisää tuoreusalgoritmistä osoitteesta <https://search.googleblog.com/2011/11/giving-you-fresher-more-recent-search.html>.

¹²⁷ Ks. lisää osoitteesta <https://www.google.com/intl/fi/search/howsearchworks/algorithms/>.

¹²⁸ Ks. esimerkki binäärihauista.

¹²⁹ Jos hakusana löytyy esimerkiksi internetsivun otsikosta, tulkitaan se paremmaksi tulokseksi kuin että hakusana löytyisi leipätekstistä. Ks. lisää osoitteesta <https://www.google.com/intl/fi/search/howsearchworks/algorithms/>.

nostaisi hakutuloksen sijaa. Googlen hakukoneen tulee siis mitata muutakin kuin sitä, kuinka monta kertaa tietty hakusana toistuu internetsivulla. Yhtenä ratkaisuna tähän ongelmaan on hyödynnetty ulkoisia linkkejä (*”inbound links”*), eli sitä kuinka paljon muut internetsivut linkittävät kyseiselle internetsivulle.¹³⁰

Myös sijainnilla on merkitystä. Käytettäessä hakusanaa kampaaja, algoritmit optimoivat hakijalle hakutuloksiin kampaamoita, jotka ovat hakijan lähetyvillä. Hakuoptimoinnissa on myös vaikutusta sillä, mitä käyttäjä on aikaisemmin hakenut sekä millä internetsivuilla hän on aikaisemmin käynyt.¹³¹ Kun algoritmit ovat löytäneet indeksistä vastaavuuksia, järjestee sen PageRank-algoritmi hakutulokset hyödyllisyysjärjestykseen.¹³² Edellä esitetystä käy ilmi, että informaation löytäminen Googlen hakukoneen avulla perustuu useiden eri tarpeisiin suunniteltujen algoritmien toimintaan. Hakukoneiden toimintaperiaatteet ovat myös linkittyneitä luvussa 2.1 esitettyihin tekstin- ja tiedonlouhinnan vaiheisiin.

Sovellettaessa näitä tekstin- ja tiedonlouhinnan vaiheita (1. pääsy materiaaliin, 2. kopiointi ja 3. louhinta) Googlen hakukoneen toimintaan esitetään tutkielmassa seuraavia johtopäätöksiä. Googlen indeksin laatimisessa on ensimmäisenä vaiheena pääsy materiaaliin, mihin liittyy erityisesti lisensoinnin ja TPM-rajoitusten kannalta oikeudellisia epävarmuustekijöitä. Materiaaliin käsiksi pääsemisen jälkeen siirrytään tekstin- ja tiedonlouhinnan toiseen vaiheeseen, eli kopiointiin. Indeksointirobotit valmistavat kopioita lukemistaan internetsivuista ja nämä kopiot tallennetaan hakukoneen palvelimelle muistiin. Varsinkin näiden kopioiden valmistaminen on tekijänoikeudellisesti ongelmallinen tilanne.

Kolmannen vaiheen, eli louhimisen, arviominen on tekstin- ja tiedonlouhinnan kannalta haastavaa. Louhimisen tarkoituksena on tuottaa käytettävästä materiaalista uutta informaatiota hyödyntämällä analysointityökaluja. Tarkasteltaessa hakukoneiden toimintaa voidaan linkittää indeksin rakentaminen ja sen ajan tasalla pitäminen tähän vaiheeseen. Googlen algoritmit poimivat kopioidusta materiaalista informaatiota indeksiinsä. Indeksoinnin toteuttamiseksi on materiaali esikäsiteltävä oikeaan muotoon hajautusfunktioiden avulla. Esikäsitelyssä hyödynnetään sanaistamista, synonyymien tunnistamista, luokittelua sekä analyysia. Koska Googlen algoritmeista ei liikesalaisuuksiin liittyvistä syistä ole juurikaan tietoa saatavilla, ei tutkielmassa voida tarkalleen esittää, mitä analysointivaihe tarkalleen käsittää eikä voida tehdä päätelmiä siitä,

¹³⁰ *Spraul* 2015, s. 158.

¹³¹ Ks. lisää osoitteesta <https://www.google.com/intl/fi/search/howsearchworks/algorithms/>.

¹³² Ks. lisää aiheesta *Niechai* 2019.

onko kyseisessä vaiheessa käsillä myös kopioiden valmistamista. Luonteeltaan toiminta on kuitenkin pitkälle automatisoitua.

Louhinnan jälkeen on saatu tekstin- ja tiedonlouhintateknologiaa käyttäen aikaan tehokas indeksi, joka hakukoneoptimoinnilla tuottaa hakukoneen käyttäjälle hakutuloksia.¹³³ Kuten aikaisemmin tässä tutkielmassa on todettu, ei tekstin- ja tiedonlouhinnassa välttämättä ole havaittavissa selkeitä rajoja eri vaiheiden välillä. Hakukoneiden toiminnasta on tunnistettavissa kaikki kolmiportaisen mallin vaiheet, mutta varsinkin kopioinnin ja louhinnan suhteen on selvää, että vaiheet ovat keskenään limittäisiä hakukonetoiminnassa. Tässä luvussa 2 saavutetun tietoteknisen ymmärryksen hakukoneiden toimintalogiikasta sekä tekstin- ja tiedonlouhinnasta avulla voidaan vastata *Law and Technology* -tutkimusmetodin ensimmäiseen osa-alueeseen. Hakukonetoiminta on yhteydessä tekijänoikeuteen, kun indeksointirobotit tekevät kopioita internetsivuista osana indeksointiprosessia. Seuraavassa osassa tätä tutkielmaa selvitetään, mitä riskitekijöitä hakukonetoiminnasta aiheutuu tekijänoikeudelle sekä sen ytimeen ulottuvalle taloudelliselle ulottuvuudelle.

¹³³ Hakukoneiden toimintaan sisältyy myös viimeiseksi vaihe, jossa indeksoinnin ja hakukoneoptimoinnin avulla saadut hakutulokset listataan ja esitetään hakukonetta käyttävälle henkilölle. Tämä toteutetaan tekemällä lyhyitä tekstiotetteita, eli kopioita, hakutuloksista ja esittämällä ne listatussa muodossa haun tehneelle hakukoneen käyttäjälle. Tutkielmassa tekstin- ja tiedonlouhintateknologiaa kuvaamaan valittu kolmiportainen malli ei tunnista enää vaihetta, jossa hakutulokset esitetään yleisölle. Lisäksi tutkielman rajauksia koskevassa osiossa 1.2 todettiin, että yleisölle välittämistä koskeva yksinoikeus jätetään tämän tutkielman ulkopuolelle. Tästä johtuen tässä tutkielmassa ei tarkastella enempää kyseiseen vaiheeseen liittyviä tekijänoikeudellisia haasteita.

3 TEKIJÄNOIKEUS DIGITALISOITUVASSA EUROOPASSA

3.1 Tekijänoikeuden ja tekstin- ja tiedonlouhinnan välinen jännite

3.1.1 Tekijänoikeuden sisältö

Edellä on käsitelty, mitä vaiheita tekstin- ja tiedonlouhintaan sisältyy sekä tarkasteltu yksityiskohtaisemmin hakukoneiden analysointiprosessin eri vaiheita. Hakukoneiden toimintalogiikasta on ilmennyt, että osana tekstin- ja tiedonlouhintaa otetaan kopioita analysoitavasta materiaalista. Tässä luvussa 3 tarkastellaan, miksi nämä tekstin- ja tiedonlouhinnan aikana otettavat kopiot ovat ongelmallisia oikeudellisesta näkökulmasta. Lisäksi tutkitaan *Law and Technology* -tutkimusmetodin toista osa-aluetta eli analysoidaan, mitä riskitekijöitä liittyy tekstin- ja tiedonlouhintateknologiaan sovitettaessa sitä yhteen tekijänoikeuden ja etenkin sen taloudellisen ulottuvuuden kanssa.

Tämän tiedon selvittämistä täydentää myös tutkielman oikeusdogmaattinen tutkimusote. Ymmärrys vallitsevasta oikeustilasta saavutetaan systematisoimalla, tulkitsamalla sekä punnitsemalla voimassa olevaa lainsäädäntöä ja oikeuskäytäntöä Euroopan unionissa.

Ensiksi on kuitenkin määriteltävä, mitä tarkoitetaan tekijänoikeudella ja mitä sisältyy tekijänoikeuden taloudelliseen ulottuvuuteen. Tämän jälkeen voidaan asettaa luvussa 2 saatu tietotekniikan ymmärrys hakukoneista tekijänoikeudelliseen kontekstiin. Kokonaiskuvan ymmärtämiseksi on kuitenkin mentävä vielä pintaa syvemmälle. Tällä hetkellä käsillä oleva jännite tekijänoikeuden ja tekstin- ja tiedonlouhinnan välillä on osa suurempaa oikeudellista ilmiötä – tekijänoikeuden digitaalista kriisiä. Luvussa 3.2 tarkastellaan, minkälaisia haasteita tekijänoikeuteen ja teknologiaan liittyy digitaalisessa ympäristössä.

Tekijänoikeus on juridisesti luotu konstruktio, joka perustuu sekä kansalliseen lainsäädäntöön että kansainvälisiin sopimuksiin.¹³⁴ Euroopan unionin alueella on huomioitava myös sen oma

¹³⁴ Useiden direktiivien taustalla vaikuttaa kansainvälisten sopimusten sisältämien oikeuksien ja velvoitteiden implementoiminen. Keskeisiä kansainvälisiä sopimuksia ovat Bernin yleissopimus kirjallisten ja taiteellisten teosten suojaamisesta (1886/1971) ("Bernin yleissopimus"); Rooman yleissopimus esittävien taiteilijoiden, äänitteiden valmistajien sekä radioyhtymien suojaamisesta (1961) ("Rooman yleissopimus"); Maailman kauppajärjestön ("WTO") sopimus teollis- ja tekijänoikeuksien kauppaan liittyvistä näkökohdista ("TRIPS-sopimus"); Yhdistyneiden Kansakuntien kasvatus-, tiede- ja kulttuurijärjestön ("UNESCO") yleismaailmallinen tekijänoikeussopimus; sekä Maailman henkisen omaisuuden järjestön ("WIPO") tekijänoikeussopimus ("WCT") ja WIPO:n esityksensä äänitesopimus ("WPPT"). Lisäksi viime vuosien aikana EU on myös allekirjoittanut WIPO:n Pekingin sopimuksen koskien audiovisuaalisia esityksiä sekä WIPO:n Marrakeshin sopimuksen julkaistujen teosten saatavuuden helpottamisesta sokeiden, heikkonäköisten tai muulla tavoin lukemisesteisten hyväksi.

merkittävä tekijänoikeudellinen normisto, joka on suurimmaksi osaksi harmonisoitua.¹³⁵ Tekijänoikeus tarkoittaa oikeudenhaltijan, eli tekijän, oikeutta määrätä teoksensa käytöstä sekä oikeutta tulla mainituksi sen tekijänä. Tekijänoikeudella voidaan suojella kirjallisuutta, artistisia luomuksia, kuten kirjoja, musiikkia, maalauksia, veistoksia, elokuvia sekä teknologiaan perustuvia teoksia, kuten tietokoneohjelmia ja tietokantoja. Toisin kuin keksintöjä suojaavat juridiset konstruktiot¹³⁶, tekijänoikeus ja sen lähioikeudet suojaavat vain itse idean ilmaisumuotoa eivätkä itse ideaa. Tekijänoikeudessa puhutaan idea-ilmaisu -jaottelusta.¹³⁷ Tekijänoikeudella suojatut teokset ilmaisevat luovuutta sen mitä moninaisimmissa muodoissa ja tekijänoikeus antaa yksinoikeussuojan itsenäisen luomistyön omaperäiselle teokselle, jonka luonnollinen henkilö on luonut tai useampi luonnollinen henkilö ovat luoneet.¹³⁸

Tekijänoikeuden yksinoikeussuojalla viitataan yleisesti tekijänoikeuden taloudelliseen ulottuvuuteen. Varsinkin Euroopan unionissa tekijänoikeuden taloudellinen ulottuvuus ilmenee sekä kappaleen valmistamisena että yleisölle välittämisenä. Tietoyhteiskuntadirektiivistä löytyy näiden molempien yksinoikeuksien normatiivinen perusta, jota on tarkennettu oikeuskäytännössä.¹³⁹ Taloudellisten oikeuksien taustalla vaikuttaa tarve turvata henkinen luomistyö sekä taata tekijöille ja esittäville taiteilijoille asianmukainen korvaus teostensa käyttämisestä.¹⁴⁰ Tämän tutkielman kannalta varsinkin kappaleen valmistamista koskeva yksinoikeus vaatii tarkempaa tutkimista.

Euroopan unionin tasolla kappaleen valmistamista koskeva yksinoikeus kuuluu tekijänoikeuden ja sen lähioikeuksien ytimeen, sillä kappaleen valmistamisen ansiosta oikeudenhaltija voi

¹³⁵ Keskeinen tekijänoikeuteen liittyvä EU-oikeudellinen normisto koostuu tietokantadirektiivistä 1996/9/EY, direktiivistä sähköisestä kaupankäynnistä 2000/31/EY, tietoyhteiskuntadirektiivistä 2001/29/EY, direktiivistä tekijänoikeuden ja tiettyjen lähioikeuksien suojan voimassaoloajasta 2006/116/EY, tietokoneohjelmadirektiivistä 2009/24/EY, tekijänoikeusdirektiivistä 2019/790/EU sekä Euroopan unionin tuomioistuimen oikeuskäytännöstä. Ks. lisää aiheesta *Pila – Torremans* 2019, s. 225.

¹³⁶ Esimerkiksi patentit.

¹³⁷ *WIPO* 2016, s. 4–6. Tekijänoikeudessa on havaittavissa luonnonoikeudellinen ajattelu. Tekijänoikeuden yhteydessä puhutaan dualistisesta perinteestä eli todellisuuden kaksijakoisuudesta, jossa idea ja ilmaisu erotetaan toisistaan. *Locken* omistusoikeuden teoria ilmentää selkeästi tätä ajattelua. Hänen mukaansa ideat ovat peräisin Jumalalta ja ne kuuluvat yhteisesti kaikille yhteiskunnassa. Kukaan ei siis voi omistaa ideoita. Ilmaisuu puolestaan saa suojaa, sillä se vaatii tekijältänsä työtä. Myös *Hughes* on korostanut, että omistajuus henkisestä omaisuudesta vaatii epämiellyttävää työtä. Ks. lisää *Hughes* 1988, s. 6–25 sekä yleisesti aiheesta *John Locke* teoksessa *Second Treatise of Government* 1982.

¹³⁸ *Pila – Torremans* 2019, s. 222.

¹³⁹ Merkittäviä lähteitä taloudellisen oikeuden arvioinnissa ovat myös Bernin yleissopimus ja WIPO:n tekijänoikeussopimus (WCT) sekä esitys- ja äänitesopimus (WPPT). Näitä ei kuitenkaan tarkemmin tässä tutkielmassa tarkastella, sillä sekä tietoyhteiskuntadirektiivi että tekijänoikeusdirektiivi on näiden pohjalta säädetty.

¹⁴⁰ Tietoyhteiskuntadirektiivi, resitaalit 9 ja 10.

sallia tai kieltää kappaleen valmistamisen suojatusta teoksesta.¹⁴¹ Estämällä tällainen kappaleen valmistaminen voidaan edesauttaa myöhempien käyttötapahuumien valvomista.¹⁴² On kuitenkin keskeistä selvittää ensiksi, mitä teoksella tarkoitetaan.¹⁴³ Teoksen sisältö on määritelty Bernin yleissopimuksessa¹⁴⁴ sekä oikeuskäytännössä, ja tulkinnan mukaan tekijänoikeutta voidaan soveltaa sellaiseen aineistoon, jota pidetään omaperäisenä siinä mielessä, että se on tekijänsä luovan henkisen työn tulos.¹⁴⁵

Aineisto voidaan määritellä tekijän luovan henkisen työn tulokseksi silloin, kun se pystyy kuvastamaan tekijänsä yksilöllisyyttä esimerkiksi siten, että tekijä on teoksen toteutuksessa kyennyt ilmaisemaan luovaa kykyään tekemällä vapaita ja luovia ratkaisuja.¹⁴⁶ Toisaalta teokseksi voidaan luokitella vain elementtejä, joita voidaan pitää ilmauksena henkisestä työstä.¹⁴⁷ Haku-koneiden suorittaman tekstin- ja tiedonlouhinnan aikana indeksointirobotit vierailevat miljardeilla internetsivuilla ja on ilmeistä, että osa internetisivuilla olevasta materiaalista täyttää teoksen kriteeristön. Indeksointitarkoitusta varten indeksointirobotit ottavat myös kopioita tästä materiaalista. Näin ollen on keskeistä tarkastella, miten indeksointirobottien tekemät kopiot nähdään tekijänoikeudellisessa kontekstissa. Tutkielmassa tarkastellaan seuraavaksi kappaleen valmistamista koskevaa artiklaa tarkemmin.

Kappaleen valmistamista koskevan yksinoikeuden juridinen sisältö on määritelty tietoyhteiskuntadirektiivin 2 artiklassa, jota Euroopan unionin tuomioistuin on myös tulkinnut laajasti. Artikla on luonteeltaan kappaleen valmistamista koskeva yleissäännös, jolla pyritään harmonisointiin horisontaalilla tasolla.¹⁴⁸ Artikla määrittelee yksinoikeuden laajuuden ja mitä seikkoja

¹⁴¹ Tietoyhteiskuntadirektiivin 2 artiklan mukaan kappaleen valmistamista koskeva oikeus kattaa niin tekijöiden teokset kuin myös esitysten tallenteet, äänitteet, elokuvien alkuperäiskappaleet ja kopiot sekä yleisradio-organisaatioiden lähetysten tallenteet.

¹⁴² Euroopan yhteisöjen komissio: Vihreä kirja Tekijänoikeus ja lähioikeudet tietoyhteiskunnassa. Bryssel 19.7.1995. KOM(95) 382 lopull. ("Euroopan yhteisöjen komissio – Vihreä kirja 1995"), s. 49.

¹⁴³ Teoksen käsitteen määrittelyyn on liittynyt tulkinnallisuutta. Verrattain tuoreessa ratkaisussa asia C-310/17 *Levola Hengelo* Euroopan unionin tuomioistuin totesi, ettei elintarvikkeen makua voida pitää tekijänoikeussuojaa saavana teoksena. Tätä perusteltiin sillä, että elintarvikkeen maun täsmällinen ja objektiivinen yksilöiminen on ongelmallista. Näin ollen on haastavaa erottaa maku muista samantapaisista elintarvikkeista. Ks. lisää asia C-310/17 *Levola Hengelo*, kohta 43.

¹⁴⁴ Bernin yleissopimuksen 2 artiklasta ilmenee, että tietyt aineistot saavat tekijänoikeudellista suojaa teoksina, jos niitä voidaan pitää henkisen luomistyön tuotteina.

¹⁴⁵ Asia C-5/08 *Infopaq I*, kohdat 34 ja 35.

¹⁴⁶ Asia C-145/10 *Painer*, kohdat 87-89, asia C-469/17 *Funke Medien NRW*, kohta 19 sekä yhdistetyt asiat C-403/08 ja C-429/08 *Premier League*, kohta 98.

¹⁴⁷ Asia C-310/17 *Levola Hengelo*, kohta 37, asia C-5/08 *Infopaq I*, kohta 39 sekä yhdistetyt asiat C-403/08 ja C-429/08 *Premier League*, kohta 159.

¹⁴⁸ Ennen tietoyhteiskuntadirektiiviä käytettiin puhuttiin vertikaalista harmonisoinnista. Tällä viitataan siihen, että harmonisointi tehtiin kansallisella tasolla aina koskien tiettyä aihealuetta, kuten esimerkiksi ohjelmistoja taikka tietokantoja. Ks. *Ohly* 2009, s. 214 ja 232.

tulee ottaa huomioon, kun arvioidaan, onko oikeudenhaltijalla oikeus sallia taikka kieltää kappaleen valmistamiseksi tunnistettu toiminta. Näitä seikkoja käydään seuraavaksi läpi.

Ensinnäkin oikeudenhaltijalla on yksinoikeus *joko kieltää taikka sallia kappaleen valmistaminen teoksista, joihin henkilöllä on ollut joko suoraan taikka välillisesti pääsy*. Esimerkiksi valokuvan ottaminen maalauksesta käsittää tilanteen, jossa henkilöllä on ollut suora pääsy teokseen. Jos tilanteena on puolestaan se, että henkilö toteuttaa veistoksen ainoastaan toisen henkilön kuvauksen perusteella, on kyseessä tilanne, jossa henkilöllä on välillinen pääsy teokseen.¹⁴⁹ Edellä luvussa 2.2.6 on esitetty, että hakukoneiden indeksointirobotit lukevat miljardeja internetsivuja indeksointivaiheen aikana. Vaikka oikeustila on epäselvä sen suhteen, missä laajuudessa lisensointijärjestyillä ja TPM-rajoituksilla voidaan pääsyä materiaaliin rajoittaa, voidaan kuitenkin tulkita, että indeksointiroboteilla on ainakin suora pääsy siihen materiaaliin, josta ne ottavat indeksointitarkoitusta varten kopioita. Näin ollen ensimmäinen kriteeri täyttyy.

Toiseksi yksinoikeuden haltijalla on oikeus *kieltää taikka sallia tilapäisesti tai pysyvästi tapahtuva kappaleen valmistaminen*. Tilapäisellä kappaleen valmistamisella tarkoitetaan sitä, että kopio tuhoetaan automaattisesti ilman, että ihmisen tulee erikseen poistaa kopio. Pysyvä kappaleen valmistaminen sen sijaan tarkoittaa sitä, että ihmisen on erikseen tuhottava kopio.¹⁵⁰ Myöskin tämä vaihe täyttyy hakukoneiden kohdalla. Ennen kuin indeksointirobottien keräämää materiaalia voidaan esikäsitellä ja indeksoida, valmistetaan siitä kopioita hakukoneen palvelimelle. Kopiot ovat palvelimella niin pitkään, kunnes ne korvataan uusilla kopioilla.

Kolmanneksi yksinoikeus kattaa *millä tahansa keinolla ja missä muodossa tahansa tapahtuvan kappaleen valmistamisen*. Tietoyhteiskuntadirektiivi ei sisällä nimenomaista määritelmää, mitä kappaleen valmistamisella tarkoitetaan, vaan se sisältää ainoastaan teknologianeutraalin määritelmän, jota digitaaliset teknologiat ovat huomattavasti laajentaneet ilman lainsäätäjän vaikutusta.¹⁵¹ Oikeuskäytännössä on todettu, että kappaleen valmistamista koskevan yksinoikeuden suojan on oltava ulottuvuudeltaan laaja.¹⁵² Teknologianeutraalisuudella viitataan siihen, ettei kappaleen valmistamista koskeva yksinoikeus ole riippuvainen siitä, missä mediassa

¹⁴⁹ Pila – Torremans 2019, s. 279.

¹⁵⁰ Pila – Torremans 2019, s. 280.

¹⁵¹ Deepreuw 2014, s. 239. Teknologianeutraaliin määritelmään liittyviä haasteita tarkastellaan tarkemmin luvussa 3.2.

¹⁵² Asia C-5/08 Infopaq I, kohta 43.

taikka ulottuvuudessa kopio on toteutettu.¹⁵³ Päinvastoin kappaleen valmistaminen kattaa monia eri tilanteita, muotoja ja ilmenemistapoja sekä erilaisia hyväksikäyttötilanteita.¹⁵⁴ Täten se käsittää niin kääntämisen, muuntelun sekä käyttämisen uuteen tarkoitukseen, kuten parodiaan taikka kasvojen tunnistukseen.¹⁵⁵ Laaja määritelmä kattaa myös osana hakukoneiden indeksointiprosessia tehtävät digitaaliset ja tekniset kopiot, jolloin myös kolmas kappaleen valmistamista koskeva kriteeri täyttyy hakukonetoiminnassa.

Neljänneksi kappaleen valmistamista koskeva yksinoikeus kattaa *sekä kokonaan että osittain tapahtuvan kappaleen valmistamisen*. Ratkaisussa *Infopaq I* Euroopan unionin tuomioistuin on määritellyt, että silloin on kyse osittain tapahtuvasta kappaleen valmistamisesta, jos kyseinen osittainen kopio sisältää sellaisen teoksen osatekijän, joka ilmentää sellaisenaan tekijänsä henkistä luomistyötä.¹⁵⁶ Sama on vahvistettu ratkaisussa *Premier League*.¹⁵⁷ Osana hakukoneiden suorittamaa tekstin- ja tiedonlouhintaa valmistetaan kopioita kokonaisista internetsivuista hakukoneen palvelimelle. Näin ollen myös viimeinen kriteeri täyttyy hakukonetoiminnan kohdalla.

Hakukoneiden suorittama tekstin- ja tiedonlouhinta on siis toimintaa, johon ulottuu tietoyhteiskuntadirektiivin 2 artiklassa määritelty kappaleen valmistamista koskeva yksinoikeus. Tekstin- ja tiedonlouhinta ei kuitenkaan sellaisenaan ole tekijänoikeuden loukkaus, mutta kyseisen teknologian puitteissa tehtävät kopiot ovat haasteellisia tekijänoikeudellisessa mielessä. Oikeuskirjallisuudessa sekä Euroopan unionin eri toimielimissä on esitetty kantoja siitä, miten tätä tekijänoikeuden ja tekstin- ja tiedonlouhinnan välistä jännitettä kuuluisi tarkastella. Seuraava alaluku tarkastelee näitä kantoja.

¹⁵³ Asia C-419/13 *Art & Allposter*, kohdat 42 ja 43. Kyseisessä ratkaisussa katsottiin, että suojatun teoskappaleen muuntelu korvaamalla alusta toisella, eli tekijänoikeudella suojatun teoksen kuvan siirtäminen paperijulisteesta canvas-kankaalle, saattoi jopa parantaa jäljennöksen kestävyyttä ja kuvanlaatua julisteeseen nähden ja lopputuloksena saattoi olla teos, joka on jopa uskollisempi teoksen alkuperäiskappaleelle. Täten Euroopan unionin tuomioistuin katsoi, että kyseinen teoskappaleen muuntelu täytti kappaleen valmistamista koskevan määritelmän ja oli täten toimintaa, johon on sidoksissa oikeudenhaltijan yksinoikeus kappaleen valmistamiseen. *Strowel* kyseenalaistaa vahvasti näkemyksen siitä, että kaikkia kopioita tulisi kohdella samanarvoisesti. Hän esittää, että sen sijaan että poikkeusten listaa kasvatettaisiin entisestään, tulisi sen sijaan tarkastella uudestaan kappaleen valmistamista koskevaa käsitettä sekä sen loukkaamista. *Strowel* 2018, s. 206.

¹⁵⁴ *Depreeuw* 2014, s. 192.

¹⁵⁵ *Pila – Torremans* 2019, s. 280. Katso erityisesti asia C-145/10 *Painer*, kohta 99, jonka mukaan sen jälkeen kun teosta voidaan pitää teoksena, sen saama suoja ei saa olla rajoitetumpaa kuin minkään muun teoksen suoja.

¹⁵⁶ Ratkaisussa *Infopaq I* Euroopan unionin tuomioistuin katsoi, että tietojenkeruun aikana suoritettu toimi, joka koostuu yhdestätoista sanasta koostuvan suojatun teoksen otteen tallentamisesta tietokoneen muistiin sekä tällaisen otteen tulostamisesta, voidaan katsoa käsittävän kappaleen valmistamiseksi määriteltävän toimenpiteen, jos näin esitettyjen osatekijöiden voidaan katsoa ilmentävän kyseisen teoksen tekijän henkistä luomistyötä. Ks. Asia C-5/08 *Infopaq I*, kohdat 48 ja 51.

¹⁵⁷ Yhdistetyt asiat C-403/08 ja C-429/08 *Premier League*, kohta 159.

3.1.2 Tarkempi oikeudellinen tarkastelu

Kretschmer ja *Margoni* kyseenalaistavat perustellusti tulisiko tekstin- ja tiedonlouhintaa ylipäätensä tarkastella tekijänoikeuden näkökulmasta ja ettei tekstin- ja tiedonlouhintaa välttämättä tulisi edes luokitella toiminnaksi, joka loukkaisi tekijänoikeutta. Tekstin- ja tiedonlouhinta teknologian tavoitteena ei ole käyttää hyväksi tekijänoikeudella suojattua materiaalia, vaan se keskittyy arvioimaan mitä tietoon on varastoitu sekä tekemään tämän perusteella johtopäätöksiä.¹⁵⁸ Esimerkiksi hakukoneiden tuottamien hakutulosten ja hakukoneindeksoinnin tavoitteena ei ole hyödyntää teoksien alkuperäistä ilmaisua, vaan järjestää internetissä olevaa tietoa ja tarjota hakukoneen käyttäjille pääsy siihen.

Myös Euroopan unionin tasolla on varovaisesti otettu kantaa asiaan. *Infopaq I* -ratkaisussa Euroopan unionin tuomioistuin on nimenomaisesti ilmaissut, että vähäinen muutamista sanoista koostuva kopiointi ei kuulu tekijänoikeuden loukkauksen alaan.¹⁵⁹ Lisäksi on todettu, ettei analysointiprosessin aikana tapahtuva *vähäinen* indeksointiprosessi myöskään kuulu tekijänoikeuden loukkauksen alaan. Tätä on perusteltu juurikin sillä, että tekijänoikeus ei suojaa dataa itsessään vaan sen omaperäistä esittämismuotoa.¹⁶⁰ Näin ollen lyhyiden katkelmien ei katsota loukkaavan tekijänoikeutta, mutta puolestaan pitkien katkelmien etenevissä määrin kyllä.

Lisäksi Euroopan parlamentin sisäasioiden poliitikan yksikkö on todennut, ettei vähäisten kopioiden taikka indeksoinnin tulisi olla toimintaa, joka loukkaisi kappaleen valmistamista koskevaa yksinoikeutta.¹⁶¹ Myös tässä tutkielmassa suhtaudutaan kriittisesti tulkintaan siitä, että laajat kopioinnit käsittäisivät automaattisesti tekijänoikeuden loukkauksen. Hakukoneiden kohdalla on kyse valtavista indekseistä, jotka vaativat miljardien internetsivujen kopioimista. Kuitenkaan pelkän kopioinnin laajuuden perusteella ei voida perustellusti argumentoida, onko tiedon omaperäistä esittämismuotoa loukattu. On nimittäin huomioitava, että kyseessä on tekno-

¹⁵⁸ *Kretschmer – Margoni* 2018a ja *Kretschmer – Margoni* 2018b.

¹⁵⁹ Asia C-5/08 *Infopaq I*, kohdat 54–55.

¹⁶⁰ *Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: oikeudellinen analyysi* 2018, s. 6.

¹⁶¹ Lausunnossa esitetään, että asiaa tulisi ennemmin tarkastella yleisölle välittämistä koskevan normiston näkökulmasta eikä toimintana, joka loukkaisi kappaleen valmistamisen yhteydessä yksinoikeutta. Yleisölle välittämistä koskeva kokonaisuus on kuitenkin rajattu tämän tutkielman tarkastelun ulkopuolelle. Ks. esimerkiksi lisää aiheesta *Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: oikeudellinen analyysi* 2018, s. 7. Tosin *Triaille, de Meeûs d'Argentuil* ja *de Francquen* ovat esittäneet, ettei kyseessä ole yleisölle välittämistä, sillä kyseisiä teknisen prosessin aikana tehtyjä kopioita ei itsessään esitetä yleisölle. Yleisölle esitetään ainoastaan tekstin- ja tiedonlouhinta teknologian avulla tuotettu uusi informaatio. Ks. *Triaille – de Meeûs d'Argentuil – de Francquen* 2014, s. 33.

logia, joka perustuu tietomassan laadulliseen analysointiin. Kopioinnin laajuus ei saisi olla ratkaiseva tekijä, jonka ympärille tekijänoikeuden ja tekstin- ja tiedonlouhinnan välisen suhteen arvioinnin tulisi keskittyä.

Oikeuskäytäntö on vaikuttanut erityisesti tällä hetkellä vallitsevaan tulkintaan siitä, missä laajuudessa kappaleen valmistamista koskevaa yksinoikeutta tulkitaan. Tämän hetkisen oikeustilan taustalla vaikuttaa kuitenkin myös laajempi yhteiskunnallinen ilmiö sekä historiallinen kehityspolku. Seuraavaksi käsitellään näitä tarkemmin, jotta saadaan monipuolisempi ymmärrys kappaleen valmistamisen ja tekstin- ja tiedonlouhinnan välisen jännitteen taustalla olevista vaikuttimista.

3.2 Tekijänoikeuden digikriisi

3.2.1 *Analogisesta maailmasta digitaaliseen ympäristöön*

Kappaleen valmistamiseen liittyvä problematiikka kuuluu osaksi laajempaa yhteiskunnallista kehitystä, jota kuvataan termillä *tekijänoikeuden digikriisi*. Puhe tekijänoikeuden digitaalisesta kriisistä ei ole kuitenkaan mikään tuore tai tiettyyn maanosaan¹⁶² rajattu ilmiö. Se, miten taloudellisten oikeuksien asema hahmotetaan tekijänoikeuslainsäädännössä sekä tekijänoikeuden yleisten periaatteiden ja tavoitteiden valossa, on puhututtanut ympäri maailmaa siitä asti, kun internetin ja digitalisaatiokehityksen vallankumous on alkanut.¹⁶³

Tekijänoikeuden vahva sidonnaisuus yhteiskuntaan on aiheuttanut sen, että tekijänoikeuden alkuajoista lähtien on käyty punnintaa niin tekijänoikeuden oikeudenhaltijoiden oikeuksien, kolmansien osapuolien oikeuksien kuin myös julkisen edun välillä.¹⁶⁴ Painokoneen keksimisen jälkeen teknologiset keksinnöt ja ilmiöt ovat aika ajoin haastaneet tarkastelemaan uudestaan

¹⁶² Myös Yhdysvalloissa *Samuelson* on esittänyt vastaavanlaista kritiikkiä yhdysvaltalaisista tekijänoikeuslainsäädäntöä kohtaan. Hänen mukaansa se ei ole ottanut riittävässä määrin huomioon teknologian kehitystä. Lisäksi hän on esittänyt, että tekijänoikeuteen kuuluvaa taloudellisen oikeuden ulottuvuutta tulisi tarkastella uudelleen sekä etenkin kappaleen valmistamiseen liittyvä problematiikka vaatii uudelleentarkastelua. Ks. lisää *Samuelson* 2007, s. 2–4.

¹⁶³ *Strowel* 2018, s. 203.

¹⁶⁴ Euroopan unionissa 2000-luvulla näiden kilpailevien oikeuksien ja etujen välistä jännitettä on valtiosääntöoikeudellisella tasolla luonnehdittu perusoikeuksien kollisioksi, jota punnittaessa on huomioitava niin immateriaali-oikeuksien suojaaminen, sananvapaus, koulutukseen liittyvät oikeudet, yhteiskunnallinen integroituminen, osallistuminen yhteiskunnalliseen keskusteluun ja päätöksentekoon sekä elinkeinovapaus. Ks. *Torremans* 2019, s. 221–222.

tekijänoikeuteen liittyviä doktriineja. Doktriinien uudelleentarkastelu on tärkeää, sillä tekijänoikeuden doktriinit voivat suuresti vaikuttaa tiedon liikkumiseen ja jakamiseen sekä teknologian kehittymiseen yhteiskunnassa.¹⁶⁵

Ennen internetin vallankumousta saman ilmiön ovat käynnistäneet esimerkiksi satelliitti, kopiokone, VCR, radio, nauhuri, valokuvaus sekä printtimedia.¹⁶⁶ Teknologian kehitys on saanut tuomarit ja lainsäätäjät kohtaamaan tilanteita, joissa he joutuvat sovittamaan uusia ongelmia lainsäädännölliseen ympäristöön, joka ei ole tarpeeksi joustava. Voimassa olevien oikeussääntöjen puuttuminen ajaakin heidät ratkaisemaan uusia ongelmia tutkimalla, miten aikaisemmat menettelytavat, kohtuus, taloudellinen tehokkuus sekä keskeiset oikeusperiaatteet voitaisiin saada oikeudenmukaisella tavalla sovitettua yhteen.¹⁶⁷ Harkinnassa haetaankin yleisesti tulkinta-apua niin valtiosäännöstä kuin myös tekijänoikeuden historiallisesta kehityksestä ja tekijänoikeuden taustalla vaikuttavista perusperiaatteista.¹⁶⁸

Esimerkiksi *Hughenholz* on tarkastellut jo vuonna 1996 sitä, millaisia muutospaineita internetin käytön yleistymisen on tuonut mukanaan sekä sitä onko verkossa tapahtuvaa kommunikaatiota pystytty sisällyttämään sen hetkiseen oikeusjärjestelmään. Hän on esittänyt jo tuolloin, että tekijänoikeusjärjestelmä vaatii suuria muutostöitä, jotta se pystyisi vastaamaan digitalisaation muutospaineesiin. Järjestelmän uudistamista hänen mielestään ei olisi tullut ulottaa pelkästään suojelemaan oikeudenhaltijoita, vaan myös niitä subjekteja, jotka oikeutetusti käyttivät teoksia. Hän on tunnistanut jo miltei 25 vuotta sitten sen, ettei vanhaan mediaan suunniteltuja käsitteitä voine rajattomasti soveltaa teknologioihin ja tilanteisiin, joita ei ole ollut edes olemassa alkuperäisissä soveltamistilanteissa. Esimerkiksi kappaleen valmistamiseen liittyvän konseptin soveltaminen alkuperäistä tarkoitusta laajemmin voi johtaa oikeudenhaltijoiden ylisuojelemiseen.¹⁶⁹

Tänä päivänä nämä muutospainemat ovat johtaneet siihen, että teosten digitalisoitumisen yhteydessä on alettu puhua *tekijänoikeuden digitaalisesta kriisistä*¹⁷⁰. Kriisi ilmenee paradoksaalisin

¹⁶⁵ *Price* 1998, s. 443.

¹⁶⁶ *Versteeg* 2000, s. 522.

¹⁶⁷ *Versteeg* 2000, s. 523.

¹⁶⁸ *Leaffer* 2014, s. 5–6.

¹⁶⁹ *Hughenholz* 1996, s. 18.

¹⁷⁰ *Strowel* 2018, s. 203 sekä *Samuelson* 2007.

seurauksin. Yhtäältä käsillä on tilanne, jossa tekijänoikeuden tehokkuus ja oikeutus ovat heikkenemässä. Oikeudenhaltijoiden on vaikea kontrolloida luvattomien kopioiden leviämistä. Lisäksi piraatit ovat siirtäneet toimintaansa matalan lainsäädännön maihin, siviili- sekä rikostyöntäntöönpanotoimet ovat tehottomia ja Digital Rights Management -tekniikoita (”DRM”) hakkeroidaan systemaattisesti. On havaittavissa, että digitaaliseen maailmaan on haastavaa soveltaa analogiseen maailmaan suunniteltuja suojamekanismeja.¹⁷¹

Toisaalta on havaittavissa kehitys, jossa yksinoikeuden soveltamisalaa on laajennettu merkittävästi. Tekijänoikeudellisen kontrollin alle on siirtynyt uusia kohteita, kuten lyhyet otteet tai käyttäjien luoma materiaali. Lisäksi kontrollia on vahvistettu tekijänoikeuteen liittyvillä korvauksilla, veroilla, tekijänoikeuden kestoa pidentämällä sekä oikeussuojakeinojen määrää lisäämällä.¹⁷²

Tekstin- ja tiedonlouhintaa koskevan juridisen viitekehyksen taustalta onkin havaittavissa suurempi ongelma, joka on sidoksissa tekijänoikeuden peruseriaatteisiin ja niiden määrittelyyn. Digitalisoituminen on mahdollistanut sen, että muutamalla hiiren napsautuksella on mahdollista tehdä täydellisiä kopioita tekijänoikeudella suojatusta materiaalista samalla, kun kustannukset voivat olla mahdollisesti täysin olemattomia. Näitä kopioita on mahdollista jakaa maailmanlaajuisesti sekuntien murto-osissa. Teknologinen kehitys on kuitenkin johtanut siihen tilanteeseen, että yhä useampi nykypäivän teknologia sisältää vaiheita, joissa on välttämätöntä ottaa kopioita, käyttää niitä ja tietyissä määrin myös levittää niitä.¹⁷³

Kopioiden valmistaminen osana teknologista prosessia on siis osa laajempaa ilmiötä. Tämä on johtanut siihen, että kappaleen valmistaminen on määritelty laajasti teknologianeutraalin käsitteen avulla. Lainsäädännöllisteknisesti tämä ei kuitenkaan ole ollut onnistunut ratkaisu. Seuraavassa luvussa tarkastellaan, mikä on tarkalleen johtanut teknologianeutraalin määritelmän käyttöönottoon sekä mitä haasteita teknologianeutraalin kappaleen valmistamista koskevaan määritelmään sisältyy.

¹⁷¹ Strowel 2018, s. 203–204. Myös *Pila* ja *Torremans* ovat todenneet, että tekijänoikeuden ja lähioikeuksien täytäntöönpanoon ja hyödyntämiseen liittyvät haasteet digitaalisessa ympäristössä ovat tällä hetkellä merkittäviä lainsäädännöllisestä näkökulmasta. Katso lisää aiheesta *Pila – Torremans* 2019, s. 221.

¹⁷² Strowel 2018, s. 203–204.

¹⁷³ Strowel 2018, s. 204–205.

3.2.2 Kappaleen valmistaminen digitaalisessa ympäristössä

Tietoyhteiskuntadirektiivin taustalla on vaikuttanut Euroopan yhteisöjen komission linjaus, jonka mukaan kopioinnin on katettava kaikki tunnetut tai tulevaisuudessa keksittävät kopiointimenetelmät.¹⁷⁴ Näin ollen kopioinnin käsite on määritelty laajaksi ja siihen katsotaan myös sisältyvän satunnaiset ja hetkelliset toimet, jotka ovat osa tavallista elektronista järjestelmää, kuten esimerkiksi tiedonsiirtoa internetissä.¹⁷⁵ Myös Euroopan talous- ja sosiaalikomitean näkemyksen mukaan tietoyhteiskuntaa voitaisiin edistää vain, mikäli oikeudenhaltijat voivat valvoa teostensa kaupallista online-käyttöä, sillä online-käyttö käsittää paljolti väliaikaista ja lyhytaikaista kopiointia.¹⁷⁶ Tietoyhteiskuntadirektiivistä on nähtävissä ajatus siitä, että toimien laaja määrittely takaisi paremmin oikeusvarmuuden Euroopan unionin sisämarkkinoilla.¹⁷⁷

Tarve laajasta teknologianeutraalista määritelmästä ei ole kuitenkaan syntynyt tyhjistä tietoyhteiskuntadirektiivin säätämisen aikana, vaan tarkempi tarkastelu osoittaa sen olevan osa isompaa uudistusta. Tietoyhteiskuntadirektiivin säätämiseen on vaikuttanut nimittäin pyrkimys saattaa Euroopan unionissa tekijänoikeudellinen suoja samalle tasolle kuin mille se oltiin asetettu jo tietokoneohjelmadirektiivissä¹⁷⁸, vuokraus- ja lainausoikeusdirektiivissä¹⁷⁹, satelliittidirektiivissä¹⁸⁰, direktiivissä tekijänoikeuden ja tiettyjen lähioikeuksien suojan voimassaoloajan yhdenmukaistamisesta¹⁸¹ sekä tietokantadirektiivissä¹⁸². Lisäksi tietoyhteiskuntadirektiivissä on haluttu sovittaa tekijänoikeuden periaatteet ja säännöt tietoyhteiskuntaan.¹⁸³ Etenkin tietokoneohjelmadirektiivin ja tietokantadirektiivin sisältämät suojamekanismit ovat olleet merkittäviä säädettäessä tietoyhteiskuntadirektiiviä.

¹⁷⁴ Euroopan yhteisöjen komissio – Vihreä kirja 1995, s. 50.

¹⁷⁵ Komission tiedonanto - Tekijänoikeutta ja lähioikeuksia tietoyhteiskunnassa koskevan vihreän kirjan seuranta KOM/96/0568, s. 9.

¹⁷⁶ Talous- ja sosiaalikomitean lausunto aiheesta ”Ehdotus: Euroopan parlamentin ja neuvoston direktiivi tekijänoikeuden ja lähioikeuksien tiettyjen piirteiden yhdenmukaistamisesta tietoyhteiskunnassa”(1) (98/C 407/06), kohta 3.3, s. 2.

¹⁷⁷ Tietoyhteiskuntadirektiivi, resitaali 21.

¹⁷⁸ Euroopan parlamentin ja neuvoston direktiivi 2009/24/EY, annettu 23 päivänä huhtikuuta 2009, tietokoneohjelmien oikeudellisesta suojasta.

¹⁷⁹ Euroopan parlamentin ja neuvoston direktiivi 2006/115/EY, annettu 12 päivänä joulukuuta 2006, vuokraus- ja lainausoikeuksista sekä tietyistä tekijänoikeuden lähioikeuksista henkisen omaisuuden alalla.

¹⁸⁰ Neuvoston direktiivi 93/83/ETY, annettu 27 päivänä syyskuuta 1993, tiettyjen satelliitin välityksellä tapahtuvaan yleisradiointiin ja kaapeleitse tapahtuvaan edelleen lähettämiseen sovellettavien tekijänoikeutta sekä lähioikeuksia koskevien sääntöjen yhteensovittamisesta.

¹⁸¹ Neuvoston direktiivi 93/98/ETY, annettu 29 päivänä lokakuuta 1993, tekijänoikeuden ja tiettyjen lähioikeuksien suojan voimassaoloajan yhdenmukaistamisesta.

¹⁸² Euroopan parlamentin ja neuvoston direktiivi 96/9/EY, annettu 11 päivänä maaliskuuta 1996, tietokantojen oikeudellisesta suojasta.

¹⁸³ Tietoyhteiskuntadirektiivi, resitaali 20.

Tietokoneohjelmadirektiivissä on määritelty kappaleen valmistamista koskevat suojaimekanismit laajoiksi ja niillä on pyritty kattamaan eri hyödyntämistilanteet. Samaa laajaa ja ylisuojelevaa ratkaisua on hyödynnetty myös tietokantadirektiivissä. Laajat suojaimekanismit ovat merkinneet tietokantadirektiivissä, että siinä on otettu perustavanlaatuinen siirtymä kappaleen valmistamisessa: kappaleen valmistamista koskeva oikeus ei enää ulottunut pelkästään suojaimefunktion, jonka tarkoituksena on ollut suojella teoksia niiden hyödyntämiseltä. Oikeus laajennettiin kattamaan kaikki kappaleen valmistamista koskevat toimet. Samassa tulkintaa on laajennettu kattamaan myös teoksista otetut väliaikaiset kopiot, ottamatta kuitenkaan huomioon altistavatko nämä kopiot tulevaisuudessa teoksen mahdolliselle hyödyntämiselle.¹⁸⁴

Tietokoneohjelmadirektiivistä omaksuttua tulkintaa siitä, että kappaleen valmistamista koskeva yksinoikeus tulisi tulkita laajasti, on laajennettu entisestään, kun tulkinta on siirretty digitaaliseen ympäristöön tietokantadirektiivissä. *Depreeuw* on kritisoinut sitä, että tietokantadirektiivissä oleva kappaleen valmistamista koskeva käsite on rakentunut periaatteiden ja hyödyntämissmallien varaan, joita ei voida suoraan soveltaa digitaalisessa ympäristössä. Kappaleen valmistamista koskeva yksinoikeus tulisikin palauttaa taloudellisen oikeuden peruseriaatteisiin.¹⁸⁵

Alun perin yksinoikeuden tarkoituksena on ollut suojata teoksia kappaleen valmistamiselta, jonka tarkoituksena on ollut hyödyntää teosta. Tällä hetkellä tilanne on kuitenkin se, että digitaalisten kopioiden kohdalla arvioitaessa ei oteta huomioon minkälainen arvo hyödyntämisellä on ja onko hyödyntämisellä ylipäänsä taloudellista arvoa. Tietoyhteiskuntadirektiivissä oleva säännös kappaleen valmistamisesta ei ole muuttunut digitalisoitumisen myötä, mutta yksinoikeuden soveltamisalaa on laajennettu teknologisessa ympäristössä tapahtuneiden muutosten seurauksena. Muutos on tapahtunut oikeusvarmuuden kustannuksella¹⁸⁶ sekä johtanut oikeudenhaltijoiden ylisuojeluun.

Strowel kyseenalaistaa vahvasti näkemyksen siitä, että kaikkia kopioita tulisi kohdella samankaltaisesti. Hän esittää, että poikkeuslistan tarpeettoman kasvattamisen sijasta tulisi tarkastella uudelleen kappaleen valmistamista koskevaa käsitettä. Hänen mukaansa tekijänoikeuden suhde

¹⁸⁴ *Depreeuw* 2014, s. 495–496.

¹⁸⁵ *Depreeuw* 2014, s. 241 ja 495.

¹⁸⁶ *Depreeuw* 2014, s. 496.

reaalimaailmaan heikkenee muutoin entisestään. Analogisia ja digitaalisia kopioita ei voida rinnastaa keskenään, sillä niillä on aivan erilainen kaupallinen asema.¹⁸⁷ Myös *Deepreuw* on esittänyt, että analogisten ja digitaalisten kopioiden välillä on tehtävä ero. Kun arvioidaan kopioiden sidonnaisuutta suojattuun teokseen, on keskeistä tarkastella, ovatko kopiot pysyviä, perinteisiä, fyysisiä kopioita vai digitaalisia, aineettomia, huomaamattomia ja/tai lyhytaikaisia kopioita.¹⁸⁸

Lisäksi yhdysvaltalaisen tuomioistuimen ratkaisussa *Authors Guild v Google, Inc*¹⁸⁹ on tarkasteltu tekstin- ja tiedonlouhintateknologian suhdetta tekijänoikeuden tarkoitukseen. Ratkaisussa on tulkittu, että Google Books:in¹⁹⁰ toiminnan tarkoituksena on laittaa saataville merkittävää tietoa itse kirjoista ja antaa hakijalle mahdollisuus löytää teoksia hakusanojen ja mahdollisten mielenkiinnon kohteiden perusteella, kun Google Books kopio alkuperäisiä tekijänoikeudella suojattuja kirjoja. Suurten tietomäärien kopiointia ja analysointia voidaan pitää kohtuulliseen käyttöön kuuluvaksi, sillä se on luonteeltaan uudistavaa.¹⁹¹

Vastaavan tulkintatavan siirtäminen Euroopan unioniin olisi suositeltavaa. Tämä nimittäin mahdollistaisi sen, että arvioitaessa sitä, onko käsillä tekijänoikeuden loukkaus, arvioitaisiin käsittääkö kappaleen valmistaminen toimintaa, joka käyttää teosta itse teoksena. On ilmeistä, ettei tekstin- ja tiedonlouhinnan kohdalla voida puhua siitä, että teknologia hyödyntäisi tekijänoikeudella suojattua teosta itse teoksena. Sen tarkoituksena ei ole käyttää hyväksi sen omaperäistä ilmenemismuotoa. Etenkin hakukoneiden kohdalla kappaleen valmistamisen taustalla on tarkoitus indeksoinnin ja optimointialgoritmien avulla järjestää internetissä oleva ääretön tietomassa ja mahdollistaa hakukonetta käyttäville henkilöille pääsy tähän informaatioon. Kyseinen käyttötarkoitus on kaukana alkuperäisen teoksen ilmenemismuodosta.

¹⁸⁷ *Strowel* 2018, s. 206.

¹⁸⁸ *Deepreuw* 2018, s. 238–239.

¹⁸⁹ Katso tarkemmin yhdysvaltalaisen valitustuomioistuimen ratkaisu *Authors Guild v. Google, Inc.* 804 F.3d 202 (2d Cir. 2015) sekä myös ratkaisut *Authors Guild, Inc. v. HathiTrust*, 902 F. Supp. 2d 445, 460 (S.D.N.Y. 2012); *Authors Guild, Inc. v. Google, Inc.*, 954 F. Supp. 2d 282, 291 (S.D.N.Y. 2013); sekä *Authors Guild v. HathiTrust*, 755 F.3d 87 (2nd Cir. 2014).

¹⁹⁰ Google Books on Googlen tarjoama palvelu, jonka tarkoituksena on koota maailman isoin kirjasto. Palvelun toteuttaminen käsitti kirjojen ja lehtien skannaamista, muuntamista tiedostomuotoon, jota sivuston käyttäjät algoritmit pystyvät hyödyntämään muun muassa hakutoimintoja varten.

¹⁹¹ *Authors Guild v. Google, Inc.* 804 F.3d 202, kohta 217 (2d Cir. 2015). Katso myös yhdysvaltalaisesta oikeuskäytännöstä seuraavat ratkaisut: *A.V. ex rel. Vanderhye v. iParadigms, LLC*, 562 F.3d 630, kohdat 639–640 (4th Cir.2009), *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146, kohta 1165 (9th Cir.2007), sekä *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, kohta 819 (9th Cir.2003).

Näin ollen ilmenee, että tekijänoikeus ja hakukoneiden suorittama tekstin- ja tiedonlouhinta pyrkivät eri päämääriin. Lisäksi tekijänoikeuden suojan piirin ulottaminen tekstin- ja tiedonlouhintaan tuntuisi ristiriitaiselta, kun otetaan huomioon, ettei tekijänoikeuden tarkoituksena ole rajoittaa pääsyä informaatioon. Päinvastoin tekijänoikeuden kommunikatiivinen funktio on yksi sen keskeisimmistä piirteistä.¹⁹²

Strowel kyseenalaistaa teknologisia prosesseja varten tehtyjen kopioiden soveltumisen taloudellisen oikeuden alle, sillä tämä rajoittaa etenkin kommunikatiivisen funktion toteutumista. *Strowel* on aiheellisesti esittänyt kritiikkiä siitä, ettei Euroopan unionin tuomioistuin ole kapaleen valmistamista koskevassa oikeuskäytännössä ottanut huomioon tekijänoikeuden kommunikatiivista ulottuvuutta. Ratkaisuissa ei ole tutkittu sitä, onko alkuperäisistä teoksista tehdyillä teknisillä kopioilla dialogista suhdetta yleisöön. Ratkaisuiden perusteluista on jäänyt myös täysin puuttumaan arvio siitä, tulisiko hyödyntämisen olla luonteeltaan taloudellista tai voittoa tavoittelevaa.¹⁹³

Tekijänoikeuden alkuperäisenä ajatuksena ei ole ollut myöntää tekijöille tuloja teoksiensa johdannaisista käyttötarkoituksista, kuten teoksien käyttämisestä hakukonetoiminnassa taikka laajan korpuksen indeksoinnissa.¹⁹⁴ Sen sijaan perinteisesti tekijänoikeus on nähty instrumenttina, joka mahdollistaa teoksien todellisen hyödyntämisen kontrolloimisen sekä takaa oikeudenhaltijoille kohtuullisen korvauksen. On myös muistettava, että kaksi vuosisataa sitten tekijänoikeuden tarkoitukseksi nähtiin myös julkisen oppimisen edistäminen yhteiskunnassa.¹⁹⁵

Law and Technology -tutkimusmetodin toiseen osa-alueeseen voidaan vastata, että tekstin- ja tiedonlouhinnan yhteydessä otettavat kopiot muodostavat riskitekijän tekijänoikeuden haltijan taloudelliselle oikeudelle eli oikeudenhaltijan oikeudelle saada asianmukainen korvaus teostensa käyttämisestä. Taloudellisen oikeuden toteutumista ovat vaikeuttaneet kopioiden siirtyminen digitaaliseen ympäristöön, jossa kopioiden valmistaminen ja eteenpäin jakaminen on teknisesti helppoa.

¹⁹² *Strowel* 2018, s. 207.

¹⁹³ *Strowel* 2018, s. 207–215.

¹⁹⁴ *Strowel* 2018, s. 206 sekä myös yleisesti aiheesta *Poort* 2018, s. 283–332.

¹⁹⁵ Yhdysvaltalaisessa järjestelmässä tämä on edelleen vielä keskeisesti läsnä, kun puolestaan Euroopassa vallitsevassa *droit d'auteur* -järjestelmässä kyseisen periaatteen vaikutus on selkeästi heikentynyt. Katso tarkemmin tekijänoikeuden historiasta ja periaatteista *Strowel* 1993 sekä *Strowel*in uudemmat teokset *Strowel* 2014, s. 701–703 sekä *Strowel* 2018, s. 215.

Voidaankin todeta, että teknologinen kehitys on asettanut tekijänoikeuden digitaaliseen kriisiin, mikä on aiheuttanut epäselvyyttä siitä, miten laajasti kappaleen valmistamiseen liittyvää suojamekanismia voidaan soveltaa. Euroopan unionissa on kuitenkin tunnistettu tiettyjä tilanteita, joissa voidaan rajoittaa tekijänoikeutta. Seuraavassa luvussa tarkastellaan yhtä näistä poikkeuksista tarkemmin ja sovelletaan kyseistä poikkeusta hakukoneiden suorittamaan kaupalliseen tekstin- ja tiedonlouhintaan.

4 TEKIJÄNOIKEUDEN RAJOITTAMINEN

4.1 Yleisesti

Tekijänoikeuden rajoittaminen tulee harkittavaksi vaihtoehdoksi, kun tutkitaan tekijänoikeuden yhteensovittamista teknologiaan kuuluvien digitaalisten kopioiden kanssa. Jos oletetaan, että hakukoneiden suorittaman tekstin- ja tiedonlouhinnan aikana tehtävät kopiot täyttävät tietoyhteiskuntadirektiivissä määritellyn yksinoikeuden kriteerit kappaleen valmistamisesta. Seurauksena on, että tekstin- ja tiedonlouhintateknologiaa käytettäessä tulisi olla aina oikeudenhaltijan etukäteinen suostumus kappaleen valmistamiseen. Etenkin hakukoneiden toiminnan kannalta tämä olisi kestävä ajatus, sillä indeksoinnin yhteydessä indeksointirobottien tulisi hankkia etukäteisiä lupia indeksointikopioiden valmistamiseen. Käytännössä näiden lupien saaminen olisi mahdotonta ja estäisi indeksoinnin.

Näin ollen seuraavissa luvuissa 4 ja 5 keskitytään tutkimaan sekä voimassa olevaa oikeustilaa tekijänoikeuteen tehtävien poikkeuksien suhteen että selvittämään *Law and Technology* -tutkimusmetodin kolmatta osa-aluetta, eli onko olemassa tilanteita, joissa on tarpeen lain avulla kannustaa tekstin- ja tiedonlouhinnan käyttöön tiettyjen arvojen suojelemiseksi. Tässä luvussa 4 tarkastellaan, miten oikeustila ennen tekijänoikeusdirektiivin voimaantuloa hahmottaa kaupallisen tekstin- ja tiedonlouhinnan ja tekijänoikeuteen tehtävien poikkeuksien ja rajoitusten välistä suhdetta.¹⁹⁶ Luvussa 5 puolestaan tarkastellaan, mitä muutoksia tekijänoikeusdirektiivi tuo oikeustilaan. Sitä ennen kuitenkin käydään läpi yleisesti tekijänoikeuden rajoittamiseen liittyviä periaatteita, koska kuten *Hugenholtz* on todennut:

”Immateriaalioikeuden paradoksi piilee järjestelmässä, joka edistäessään tietoa yhteiskunnassa kuitenkin samalla myös rajoittaa sitä.”¹⁹⁷

*Hugenholtz*in esittämiin sanoihin kiteytyy hyvin yksi immateriaalioikeuden keskeisimmistä haasteista – se on jatkuvaa puntarointia niin julkisen kuin yksityisen edun välillä. Lainsäädännössä olevien rajoitusten ja poikkeuksien tarkoituksena on sallia tekijänoikeudella suojattujen

¹⁹⁶ Tutkielmassa on tehty tietoinen valinta sen suhteen, että ensiksi käsitellään oikeustilaa ennen tekijänoikeusdirektiiviä ja vasta sen jälkeen käsitellään, mitä muutoksia tekijänoikeusdirektiivi toisi oikeustilaan. Valintaa perustellaan sillä, että tekijänoikeusdirektiivin tarkoituksena on ollut tuoda selkeyttä oikeustilaan ja poistaa tekstin- ja tiedonlouhintaan liittyviä esteitä. Ymmärryksen saaminen siitä, mikä oikeustila on ollut ennen tekijänoikeusdirektiiviä, mahdollistaa arvioinnin siitä, onko tekijänoikeusdirektiivillä pystytty ratkaisemaan sen säätämisen taustalla olevat epäselvyystekijät.

¹⁹⁷ *Hugenholtz*in puheenvuoro: ”Owning science: Intellectual Property Rights as Impediments to Knowledge Sharing” Communia-konferenssissa Italiassa Turinissa 29.6.2001. Ks. *Hugenholtz* 2001.

teoksien käyttö ilman lisenssiä julkisen edun nimissä. Tämän ajatellaan edistävän yhteiskunnan kehittymistä sekä Euroopan unionin keskeisimpien arvojen toteutumista, joita suojellaan sanan- ja informaatiovapaudella.¹⁹⁸ Voidaan esittää, että hakukonetoiminnalla on keskeinen vaikutus sanan- ja informaatiovapauden toteutumiselle yhteiskunnassa, sillä ne mahdollistavat pääsyn informaatioon sekä antavat näin ollen myös mahdollisuuden mielipiteen muodostamiselle.

Tekijänoikeuden rajoitukset määräytyvät kansainvälisten sopimusten ja EU:n tekijänoikeutta koskevan sääntelyn perusteella. Keskeisenä instrumenttina on kansainvälinen ”*three step*” -testi, jonka perusteella arvioidaan ovatko rajoitukset ja poikkeukset perusteltuja. Testin mukaan rajoitukset ja poikkeukset ovat sallittuja, jos 1) yksinoikeuksien rajoitukset ja poikkeukset rajoittuvat tiettyihin erityistapauksiin, 2) rajoitukset ja poikkeukset eivät ole ristiriidassa teoksen tavanomaisen hyödyntämisen kanssa eivätkä 3) rajoitukset ja poikkeukset kohtuuttomasti vahingoita oikeudenhaltijan oikeutettuja etuja.¹⁹⁹

Testiä ja sen kriteereitä on oikeuskirjallisuudessa tulkittu, ja kriteerien sisällölle on annettu sitä laajentavia määritelmiä. Esimerkiksi *Hugenholtzin* ja *Okedijin* mukaan poikkeukset ja rajoitukset, 1) jotka eivät ole liian laajoja, eivätkä 2) riistä oikeudenhaltijoilta joko niiden oikeaa taikka potentiaalista merkittävää tulonlähdettä sekä 3) eivät aiheuta suhteetonta vahinkoa oikeudenhaltijoille, läpäisevät testin.²⁰⁰ Myös *Geigerin*, *Griffithsin* ja *Hiltyn* näkemyksen mukaan testiä on tulkittava siten, että kriteerit ovat tasapainossa keskenään. Monipuolinen ja kaikenkattava arviointi kykenisi yhtäältä huomioimaan ratkaisun vaikutukset muun muassa julkiselle edulle, tieteen ja kulttuurin kehitykselle sekä yhteiskunnan ja talouden kasvulle. Toisaalta arviointi mahdollistaisi myös ottamaan huomioon sen, että oikeudenhaltijat saisivat oikeudenmukaisen korvauksen henkisestä luomistyöstään.²⁰¹

Tekstin- ja tiedonlouhintateknologian tunkeutuessa tekijänoikeuden rajapintaan, on syntynyt tarve arvioinnille siitä, tulisiko poikkeuksien ja rajoitusten soveltua myös tekstin- ja tiedonlouhintaan. Edellä esitetty *Geigerin*, *Griffithsin* ja *Hiltyn* malli *three step* -testin tasapainoisesta soveltamisesta tukisi rajoitusten ja poikkeuksien soveltamista tekstin- ja tiedonlouhintaan,

¹⁹⁸ *Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: oikeudellinen analyysi* 2018, s. 8. Katso myös *Geiger* 2009, s. 27 sekä *Pila – Torremans* 2019, s. 282.

¹⁹⁹ ”*Three step*” -testi on implementoitu useaan kansainväliseen sopimukseen. Katso esimerkiksi TRIPS-sopimuksen 13 artikla. Kansainväliset sopimukset sisältävät myös yksittäisiä rajoituksia ja poikkeuksia, katso esimerkiksi Bernin yleissopimuksen 10 bis tai Rooman yleissopimuksen 15 artikla.

²⁰⁰ *Hugenholtz – Okediji* 2012, s. 3.

²⁰¹ *Geiger – Griffiths – Hilty* 2008, s. 489–496.

sillä kyseinen teknologia edistäisi julkista etua etenkin huomioiden sen vaikutukset tieteen ja talouden kehitykselle.²⁰² Seuraavaksi tarkastellaan yksityiskohtaisemmin millaisia poikkeuksia ja rajoituksia tietoyhteiskuntadirektiivi sisältää.²⁰³

Tietoyhteiskuntadirektiivi sisältää sekä pakollisia että vapaaehtoisia rajoituksia ja poikkeuksia²⁰⁴. Kyseisen direktiivin keskeisenä tavoitteena on ollut rajoitussäännösten harmonisointi Euroopan unionin sisämarkkinoilla ja esittää tyhjentävä luettelo oikeuksiin sovellettavista rajoituksista ja poikkeuksista. Direktiiviä säädettäessä tunnistettiin, että teknologisen kehityksen myötä tekijänoikeuden ja lähioikeuksien on mukauduttava vastaamaan riittävällä tasolla taloudellisia tosiseikkoja, kuten uusia hyödyntämismuotoja. Lisäksi harmonisoinnilla on pyritty eri oikeudenhaltijaryhmien ja suojatun aineiston käyttäjien välisten oikeuksien ja etujen oikeudenmukaiseen tasapainoon.²⁰⁵

Tämän tutkielman kannalta olennaisin tekijänoikeuden rajoitus kätkeytyy tietoyhteiskuntadirektiivin 5 artiklan 1 kohtaan, jonka sisältämä poikkeus on implementoitu kaikkien jäsenvaltioiden kansalliseen lainsäädäntöön.²⁰⁶ Mainitun kohdan mukaan kappaleen valmistamista koskeva yksinoikeus ei koske tilapäistä kappaleen valmistamista, jos kappaleen valmistaminen on 1) väliaikaista tai satunnaista, 2) erottomaton ja välttämätön osa teknistä prosessia, 3) sen ainoa tarkoitus on joko mahdollistaa välittäjän toimesta tapahtuva teoksen tai muun suojatun aineiston siirto verkossa kolmansien osapuolien välillä tai sen laillinen käyttö, sekä 4) sillä ei ole itsenäistä taloudellista merkitystä.

²⁰² *Quintais* 2017, s. 197–205 sekä *Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: oikeudellinen analyysi* 2018, s. 8.

²⁰³ Tietoyhteiskuntadirektiivin lisäksi tietokoneohjelmadirektiivi, vuokraus- ja lainausoikeusdirektiivi sekä tietokantadirektiivi sisältävät myös tekijänoikeuteen tehtäviä poikkeuksia ja rajoituksia. Tässä tutkielmassa keskitytään kuitenkin tutkimusongelman kannalta keskeisiin poikkeuksiin ja rajoituksiin.

²⁰⁴ Tietoyhteiskuntadirektiivin 44 resitaalin mukaan pakollisen tekijänoikeuteen tehtävän rajoituksen lisäksi tietoyhteiskuntadirektiivi mahdollistaa jäsenvaltioille, että ne voivat säätää myös muita poikkeuksia ja rajoituksia, jos ne eivät kohtuuttomasti rajoita oikeudenhaltijan laillisia etuja eivätkä ole ristiriidassa aineiston tavanomaisen hyödyntämisen kanssa. Kuitenkin nämä vapaaehtoiset poikkeukset ja rajoitukset koskevat tilanteita, joissa tehdään kopioita tekijänoikeudella suojatusta aineistosta tieteellistä, kulttuurillista, poliittista, turvallisuuden liitännäistä taikka yhteiskunnallista tarkoitusta varten. Näin ollen lähtökohtaisesti ulkopuolelle rajautuu täysin toiminta, jolla on kaupallisia päämääriä. Täten vapaaehtoisia poikkeuksia ja rajoituksia ei ole tarkoituksenmukaista käsitellä tässä tutkielmassa. Ks. lisää aiheesta tietoyhteiskuntadirektiivin 5 artiklan kohdat 2–5. Kyseisiä kohtia on myös sovellettu tuoreessa oikeuskäytännössä. Katso esimerkiksi seuraavat ratkaisut asia C-161/17 *Renckhoff*, asia C-469/17 *Funke Medien NRW*, asia C-476/17 *Pelham* sekä asia C-516/17 *Spiegel Online*.

²⁰⁵ Tietoyhteiskuntadirektiivi, resitaalit 5 ja 31.

²⁰⁶ Tietoyhteiskuntadirektiivi, resitaali 33.

Tätä tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan mukaista poikkeusta on tulkittava kuitenkin suppeasti. Kyseessä on nimittäin poikkeus yleisestä säännöstä eli siitä, että kaikenlainen kappaleen valmistaminen suojatusta teoksesta edellyttää kyseisen teoksen oikeudenhaltijan suostumusta.²⁰⁷ Seuraavaksi tarkastellaan yksityiskohtaisesti sitä, miten artiklassa asetetut eri kriteerit soveltuvat hakukoneiden suorittamaan kaupalliseen tekstin- ja tiedonlouhintaan.

4.2 Tietoyhteiskuntadirektiivin 5 artiklan 1 kohta

4.2.1 Kappaleen valmistaminen on väliaikaista tai satunnaista

Tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan poikkeussäännöksen ensimmäinen kriteeri asettaa ajallisen vaatimuksen tekijänoikeudella suojatusta teoksesta otettujen kopioiden säilyttämiselle. Kopioiden on oltava luonteeltaan tilapäisiä, mikä tarkoittaa sitä, että niiden on oltava joko väliaikaisia taikka satunnaisia. Vaatimus kopioiden tilapäisyydestä taikka satunnaisuudesta asettaa haasteita hakukonetoiminnalle. Seuraavaksi tarkastellaan kyseisiä haasteita tarkemmin relevantin oikeuskäytännön valossa.

Ratkaisussa *Infopaq I* on määritelty, että kappaleen valmistamista voidaan pitää tietoyhteiskuntadirektiivin tarkoittamalla tavalla väliaikaisena vain, jos kopion olemassaoloaika on rajoitettu siihen, mikä on tarpeen, jotta tekninen prosessi, jota varten kopio on otettu, voi toimia asianmukaisella tavalla. Tekniseltä prosessilta edellytetään kuitenkin sitä, että prosessi on siinä määrin automaattinen, että valmistettu kopio poistetaan automaattisesti ilman ihmisten toimintaa mahdollisimman pian. Kopio tulisi taten poistaa silloin, kun kopion on katsottu täyttäneen prosessin mukaisen tehtävänsä, jota varten kopio on alunperin tehty.²⁰⁸

Ongelmaksi on noussut se, ettei oikeuskäytännössä olla määritelty tarkemmin ajanmäärettä. Näin ollen on siis jäänyt edelleen epäselväksi, kuinka kauan kopiota voidaan säilyttää, jotta kopioita voitaisiin pitää väliaikaisina.²⁰⁹ Seuraavaksi sovelletaan edellä käsiteltyä ajallista vaatimusta hakukoneiden suorittamaan tekstin- ja tiedonlouhintaan.

²⁰⁷ Katso seuraavat ratkaisut Euroopan unionin tuomioistuimen ratkaisukäytännöstä: Asia C-5/08 *Infopaq I*, kohdat 56 ja 57; asia C-302/10 *Infopaq II*, kohta 27; yhdistetyt ratkaisut C-403/08 ja C-429/08 *Premier League*, kohta 162; asia C-360/13 *Public Relations Consultants Association*, kohta 23; asia C-476/01 *Kapper*, kohta 72; sekä C-36/05 *Komissio v Espanja*, kohta 31.

²⁰⁸ Asia C-360/13 *Public Relations Consultants Association*, kohta 28; asia C-5/08 *Infopaq I*, kohta 64; sekä asia C-302/10 *Infopaq II*, kohta 30.

²⁰⁹ *Triaille – de Meeûs d’Argentuil – de Francquen* 2014, s. 43.

Kuten luvussa 2.2.6 todettiin, osana hakukoneiden indeksointia tuotetaan kopioita tekijänoikeudella suojatusta materiaalista. Indeksointiprosessi on luonteeltaan automaattinen, sillä indeksointirobottien toiminta perustuu algoritmeille. Algoritmit päättävät, milloin kopio otetaan ja kuinka kauan kopiota säilytetään. Indeksoinnin yhteydessä valmistettu kopio voidaan poistaa automaattisesti ilman ihmisen väliintuloa. Näin ollen voidaan esittää, että hakukoneiden tekstin- ja tiedonlouhintaprosessin aikana valmistetut kopiot täyttävät *Infopaq I* -ratkaisussa esitetyn vaatimuksen automaattisuudesta.

Tästä huolimatta ei ole täysin varmaa, voidaanko osana hakukoneiden indeksointiprosessia valmistettuja kopioita pitää tietoyhteiskuntadirektiivin mukaisesti väliaikaisina. Indeksoinnin yhteydessä indeksointirobottien tuottamat kopiot nimittäin tallennetaan hakukoneen palvelimille ja kopioita säilytetään niin kauan, kunnes ne korvataan uusilla kopioilla, jotka luodaan kun indeksointirobotit käyvät uudestaan kyseisellä sivustolla.²¹⁰ Keskeistä arvioinnissa on se, ettei kopioita tulisi säilyttää hakukoneen palvelimilla pidempään kuin on tarpeen indeksoinnin suorittamiseksi. Kuitenkin luvussa 2.2 saavutetun ymmärryksen mukaan itse hakukoneen toiminnan kannalta kopioita tarvitaan ainoastaan indeksin laatimista varten. Kun internetsivusta on louhittu tieto indeksiin, ei internetsivusta laaditulle kopiolle ole enää tarvetta itse *indeksointiprosessissa*. Tässä tutkielmassa esitetään tulkinta siitä, että indeksoinnin yhteydessä otettujen kopioiden osalta ajallinen vaatimus kopioiden säilyttämiselle ylittyy.

Arvioinnissa on kuitenkin huomioitava, että samoja indeksointia varten otettuja kopioita hyödynnetään myös välimuistikopioina ("*cache copies*"). Kopioiden tarve ei siis rajoitu ainoastaan indeksointiprosessiin, vaan välimuistikopioita hyödynnetään muun muassa nopeuttamaan hakukoneen toimintaa²¹¹ sekä käytetään hakutuloksien esittämiseen²¹². Välimuistikopioihin, jotka

²¹⁰ *Peguera* 2012, s. 174.

²¹¹ Välimuistikopiot tehostavat hakukoneiden sekä internetin toimintaa. Välimuistikopiot mahdollistavat nopeamman pääsyn internetsivuille. Tehokkuus perustuu sille, että välimuistikopiot vapauttavat kaistalta tilaa sekä vähentävät viiveitä, joita voi ilmetä varsinkin ruuhkaisilla internetsivuilla. Välimuistikopiot internetsivuista voidaan nimittäin tallentaa paikalliselle palvelimelle, mikä puolestaan mahdollistaa sen, kun internetsivulle pyritään uudestaan, voidaan välimuistikopio ladata paikallisen verkon nopeudella. Ks. lisää osoitteesta https://www.webope-dia.com/TERM/P/proxy_cache.html.

²¹² Googlen esittäessä hakutuloksia tarjoaa se samalla myös hakutuloksen tehneelle hakijalle pääsyn välimuistikopioon. Kyseisiä välimuistikopioita voidaan hyödyntää esimerkiksi tilanteissa, joissa jostain syystä ei ole pääsyä halutulle internetsivulle. Välimuistikopio mahdollistaa pääsyn siihen samaiseen kopioon, jonka perusteella kyseinen internetsivu on aikoinaan indeksoitu. Ks. lisää aiheesta *Google Support* osoitteesta <https://support.google.com/webmasters/answer/7041154?hl=fi>.

nopeuttavat hakukoneen toimintaa, liittyvät ongelmallisuudet on ratkaisu direktiivissä sähköisestä kaupankäynnistä²¹³ ja erityisesti sen 13 artiklassa²¹⁴.

Puolestaan välimuistikopioihin, joita käytetään osana hakutuloksien esittämistä, liittyy oikeudellista epäselvyyttä. Kansallisella tasolla on tulkittu, että esimerkiksi Googlen tarjoamassa News-palvelussa tuotetut välimuistikopiot eivät täyttäneet tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan mukaista vaatimusta. Belgialainen tuomioistuin on tulkinnut, että palvelussa tarjotut välimuistikopiot ovat olleet linkkeinä tarjolla kuluttajille. Tämän on katsottu merkitsevän sekä kappaleen valmistamista että yleisölle välittämistä.²¹⁵

On toisaalta huomautettava, että käyttötarkoitus suoraan kuluttajille tarjottavien linkkien ja välimuistikopioiden kohdalla on eri kuin indeksin rakentamisen kannalta välttämättömien kopioiden. Myöskään indeksoinnin yhteydessä välimuistikopioita ei anneta yleisön saataville, vaan niitä hyödynnetään teknisen prosessin yhteydessä osana ideksointia. Indeksointiprosessin yhteydessä tuotettavia kopioita hyödynnetään siis eri käyttötarkoituksiin, ja on ongelmallista arvioida, miten eri teknisiä prosesseja varten otettuja samoja kopioita tulisi arvioida tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan näkökulmasta. Voidaanko kopion pidempiaikainen säilyttäminen oikeuttaa sillä, että samaa kopiota käytetään toisessa teknisessä prosessissa? Oikeudellisesta epäselvyydestä johtuen ei voida siis antaa yksiselitteistä vastausta siitä, miten hakukoneiden indeksointiprosessin yhteydessä tuotettavia kopioita tulisi arvioida väliaikaisuuden näkökulmasta.

Satunnaisuuden arvioiminen ei myöskään ole selkeää eikä arvioinnin tueksi ole olemassa juurikaan oikeuskäytäntöä. *Vilanka* on aiheellisesti kritisoinut sitä, että tietoyhteiskuntadirektiivin

²¹³ Euroopan parlamentin ja neuvoston direktiivi 2000/31/EY, annettu 8 päivänä kesäkuuta 2000, tietoyhteiskunnan palveluja, erityisesti sähköistä kaupankäyntiä, sisämarkkinoilla koskevista tietyistä oikeudellisista näkökohdista.

²¹⁴ Artikla mahdollistaa sen, ettei esimerkiksi internetpalvelun tarjoaja ole vastuussa välimuistikopioista, jos välimuistikopiot ovat automaattisia, tilapäisiä ja lyhytaikaisia. Lisäksi edellytetään, että välimuistikopioiden tarkoituksena on muille palvelun vastaanottajille niiden pyynnöstä myöhemmin tehtävän tiedonsiirron tehostaminen.

²¹⁵ Katso belgialaisen valitustuomioistuimen ratkaisu *Google Inc. v. Copiepresse*, Cours d' Appel (CA)/Hoven van Beroep (HvB)(Courts of Appeals) Bruxelles, 9e ch. May 26, 2011, R No. 2011/2999 No. 817. Ks esimerkiksi lisää Copiepresse-ratkaisujen vaikutuksesta hakukoneiden toimintaan *Klein* 2008. Lisäksi *Peguera* on käsitellyt kattavasti, miten välimuistikopiot ovat haasteellisia, jos niitä käytetään Cache-linkkeinä Googlen hakutuloksissa. Ks. *Peguera* 2012, s. 176 ja s. 187–201.

5 artiklan 1 kohdan termejä ”tilapäinen”, ”väliaikainen” ja ”satunnainen” ei olla tarkalleen määritelty.²¹⁶ *Triaille, de Meeûs d’Argentuil* sekä *de Francquen* ovat esittäneet, ettei ole myöskään poissuljettua, että osana tekstin- ja tiedonlouhintaa tehtävät kopiot olisivat satunnaisia.²¹⁷

Infopaq I -ratkaisussa sekä tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan mukaisen rajoitusnormin soveltamisessa asetetut kriteerit indeksointiprosessin yhteydessä tehtävien kopioiden säilytysajalle ovat monitulkintaiset eikä niistä voida suoraan tehdä päätelmää siitä, soveltuvatko kyseiset kriteerit hakukoneiden suorittamaan tekstin- ja tiedonlouhintaan. Seuraavaksi siirrytään tarkastelmaan muita tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan asettamia kriteereitä.

4.2.2 Erottomaton ja välttämätön osa teknistä prosessia

Tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan toinen kriteeri sisältää vaatimuksen siitä, että kopioiden on oltava erottomaton ja välttämätön osa teknistä prosessia. Tätä vaatimusta on tarkasteltu etenkin ratkaisussa *Infopaq II*.

Kyseisessä ratkaisussa on määritelty se, että ollakseen erottomaton ja välttämätön osa teknistä prosessia, tulee kopion valmistamisen tapahtua täysin teknistä prosessia täytöntöönpanossa ja ettei kopiota suoriteta kokonaan tai osittain tällaisen prosessin ulkopuolella. Välttämättömyysvaatimuksen näkökulmasta edellytetään myös sitä, että asianomainen tekninen prosessi ei voisi sujua asianmukaisesti ja tehokkaasti ilman kyseistä kopiota.²¹⁸

Sovellettaessa tätä vaatimusta hakukoneiden suorittamaan tekstin- ja tiedonlouhintaan on ilmeistä, ettei kyseinen teknologia voisi toimia ilman kopioiden valmistamista. Luvussa 2.1 määriteltiin tekstin- ja tiedonlouhinteknologian koostuvan lähtökohtaisesti kolmesta eri vaiheesta. Yksi näistä vaiheista käsittää kopioimisen. Hakukoneiden kohdalla tämä tarkoittaa indeksointirobottien tekemiä kopioita internetsivuista. Kyseiset kopiot ovat välttämättömiä, jotta voidaan siirtyä tekstin- ja tiedonlouhinnassa seuraavaan vaiheeseen, eli louhintaan. Louhintavaiheessa

²¹⁶ *Vilanka*, 2006, s. 613.

²¹⁷ *Triaille – de Meeûs d’Argentuil – de Francquen* 2014, s. 46.

²¹⁸ Asia C-302/10 *Infopaq II*, kohta 30 sekä myös asia C-5/08 *Infopaq I*, kohta 61.

algoritmit esikäsittelevät kopioita ja tuottavat niistä uutta informaatiota, jota kootaan indekseihin.²¹⁹ Näin ollen on myös ilmeistä, ettei hakukoneteknologian taustalla toimiva tekstin- ja tiedonlouhinta voisi toimia asianmukaisesti ja tehokkaasti ilman kyseisien kopioiden valmistamista.

Oikeuskäytännössä on määritelty, että kopiointi voinee tapahtua missä tahansa teknisen prosessin vaiheessa. Kopio voi siis esimerkiksi aloittaa taikka päättää teknisen prosessin.²²⁰ Edellä on jo viitattu siihen, että tekstin- ja tiedonlouhinnan yhteydessä kopio tehdään teknisen prosessin aikana sen jälkeen, kun on saatu pääsy materiaaliin. Myöskään ihmisen toiminnalle ei ole asetettu rajoituksia oikeuskäytännössä. Ihminen saa siis toiminnallaan osallistua tekniseen prosessiin esimerkiksi käynnistämällä manuaalisesti kopion valmistamisen.²²¹ Hakukoneiden suorittaman tekstin- ja tiedonlouhinta ja etenkin sen indeksointiprosessi ovat täysin automatisoituja. Näin ollen voidaan esittää, että tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan toinen kriteeri täyttyy hakukoneiden kohdalla.

4.2.3 Kappaleen valmistamisen tarkoitus

Kolmas kriteeri liittyy tarkoitukseen, joka on kappaleen valmistamisen taustalla. Tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan mukaan kappaleen valmistamisen on joko mahdollistettava teoksen tai muun aineiston siirto verkossa kolmansien osapuolien välillä taikka teoksen tai muun aineiston laillinen käyttö.

Kappaleen valmistamisen on ensinnäkin mahdollistettava toimet, jotka mahdollistavat selailun sekä aineiston tilapäisen tallentamisen välimuistiin. Kappaleen valmistamiseen katsotaan myös sisältyvän tallennus, joka mahdollistaa siirtojärjestelmien tehokkaan toiminnan edellyttäen, että välittäjä ei muuta tietoja eikä puutu toimialalla laajalti hyväksytyyn ja käytetyn teknologian lailliseen käyttöön saadakseen tietoja talletetun tiedon käytöstä.²²²

²¹⁹ Louhintavaihe koostuu kahdesta eri alatasosta. Ensimmäisessä alatasossa käsiteltävä materiaali jäsenellään, puhdistetaan virheistä ja epäolennaisesta tiedosta sekä mahdollisesti muunnetaan koneluettavaan muotoon. Toisessa alatasossa keskitytään enemmän materiaalin analysoimiseen ja jäsentelyyn. Katso tarkemmin aiheesta tämän tutkielman luvusta 2.1.

²²⁰ Asia C-302/10 *Infopaq II*, kohta 31.

²²¹ Asia C-302/10 *Infopaq II*, kohta 32.

²²² Tietoyhteiskuntadirektiivi, resitaali 33.

Vaatimuksen soveltuminen tekstin- ja tiedonlouhintaan on haasteellista. *Ducato* ja *Strowel* ovat esittäneet, ettei tekstin- ja tiedonlouhinnan tarkoituksena ylipäänsä ole mahdollistaa välittäjän toimesta tapahtuva teoksen tai muun aineiston siirto verkossa kolmansien osapuolien välillä.²²³ Tutkielmassa esitetään, että hakukoneiden suorittama tekstin- ja tiedonlouhinta mahdollistaa puolestaan jatkuvan tiedon etsimisen internetistä. Hakukone toimii tiedon välittäjänä, mutta se ei itsessään siirrä teoksia tai aineistoa, vaan ainoastaan tuottaa hakutuloksia ja tarjoaa pääsyn kyseisille internetsivuille.

Toisaalta tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan mukaan kappaleen valmistaminen on myös mahdollista, jos sen tarkoituksena on mahdollistaa teoksen tai muun aineiston laillinen käyttö. Käyttöä on lähtökohtaisesti pidettävä laillisena, jos oikeudenhaltija sallii sen taikka käyttöä ei ole erikseen lailla rajoitettu.²²⁴ Oikeuskäytännössä Euroopan unionin tuomioistuin on pitänyt laillisuusarvioinnissa ratkaisevana, ettei ole indikaattoreita siitä, että teknistä prosessia oltaisiin käytetty johonkin muuhun tarkoitukseen.²²⁵ Tekstin- ja tiedonlouhinnan ja hakukoneiden toiminnan tarkoituksena on kerätä tietoa uusista lähteistä ja analysoinnin perusteella tuottaa uutta tietoa. Näin ollen sitä voidaan pitää laillisena, koska tieto itsessään ei ole tekijänoikeudella suojattua. Kun otetaan huomioon hakukonetoiminnan keskeinen merkitys informaatiolle, sen kehitykselle sekä leviämisellä yhteiskunnassa, voidaan pitää ilmeisen epätodennäköisenä, että toimintaa pidettäisiin laittomana.

On kuitenkin tunnistettava myös lisensointijärjestelyihin liittyvät epäselvyydet, sillä oikeudenhaltijan rajoittaessa lisenssin ehdoilla materiaalin käyttötarkoitusta ei kopiointia voida pitää enää laillisena.²²⁶ Olemassaolevat lisensointijärjestelmät tuovat siis epäselvyyttä oikeustilaan myös sen suhteen, voidaanko niiden välityksellä estää tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan mukaisen poikkeuksen soveltuminen.

4.2.4 Ei taloudellista itsenäistä merkitystä

Viimeinen tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan mukainen kriteeri liittyy kopioimisen taloudelliseen merkitykseen. Kopioimisella ei saa olla itsenäistä taloudellista merkitystä. Ongelmaksi nousee kuitenkin se, ettei Euroopan unionin tuomioistuimessa ole juurikaan otettu

²²³ *Ducato – Strowel* 2018, s. 13.

²²⁴ Tietoyhteiskuntadirektiivi, resitaali 33 sekä yhdistetyt asiat C-403/08 ja C-429/08 *Premier League*, kohta 168.

²²⁵ Asia C-302/10 *Infopaq II*, kohdat 40–44. Keskeistä arvioinnissa oli se, ettei tanskalaisessa lainsäädännössä yhteenvedon laatimista uutisartikkeleista oltu määritelty kielletyksi toiminnaksi.

²²⁶ *Triaille – de Meeûs d’Argentuil – de Francquen* 2014, s. 47.

kantaa tämän viimeisimmän kriteerin tarkkaan sisältöön. Näin ollen keskeiseksi lähteeksi taloudellisen itsenäisyyden arvioimisessa on *Infopaq II* -ratkaisu, jossa on asetettu kaksi vaatimusta taloudellisen itsenäisen merkityksen arvioimiseen.

Tässä luvussa arvioidaan ensiksi näiden kahden vaatimuksen sisältöä, minkä jälkeen tarkastellaan yleisesti tietoyhteiskuntadirektiivin 5 artiklan 1 kohtaan liittyviä taloudellisia haasteita. Taloudellinen ulottuvuus on relevantti kriteeri, joka tulisi ottaa huomioon tekijänoikeuden poikkeuksen soveltamista arvioitaessa, koska esimerkiksi hakukoneiden toiminta on luonteeltaan taloudellisesti voittoa tavoittelevaa ja sen hyödyntämä tekstin- ja tiedonlouhinta on taloudellisesti arvokasta.

Ratkaisussa *Infopaq II* määritellään ensinnäkin, ettei tilapäisen kappaleen valmistaminen saa mahdollistaa sellaisen lisäedun saamista, joka ylittää suojatun teoksen laillisesta käytöstä saadun edun.²²⁷ Tämän vaatimuksen tulkinta on ongelmallista, sillä on vaikea erottaa taloudellinen lisäetu omaksi kokonaisuudekseen analysointiprosessista. Erottaminen edellyttäisi sitä, että pystyttäisiin tarkalleen määrittämään mikä arvo ylipäättensä koko tekstin- ja tiedonlouhintateknologialla on.²²⁸

Lisäksi tässä tutkielmassa kyseenalaistetaan, onko tekstin- ja tiedonlouhinnassa hyödynnettävillä kopioille edes arvioitavissa mitään taloudellista lisäetua. Nimittäin teknologia ei hyödynnä suojaa saavan teoksen omaperäistä ilmaisua, vaan teoksen sisältämää tietoa. Näin ollen on erittäin haastavaa soveltaa tätä vaatimusta hakukonetoimintaan. Indeksointiprosessin yhteydessä laadittaville kopioille on vaikea määritellä arvoa ja on osin kyseenalaista, onko kopioilla ylipäättensä yksinoikeuden kannalta taloudellista arvoa, jota tulisi edes suojella.

Toinen *Infopaq II* -ratkaisussa esitettävä vaatimus koskee teoksen muuntamista.²²⁹ Tilapäinen kappaleen valmistaminen ei saa johtaa alkuperäisen teoksen muuttamiseen.²³⁰ Hakukoneiden suorittama tekstin- ja tiedonlouhinta ei pyri muuntamaan alkuperäistä teosta, vaan ainoastaan

²²⁷ Asia C-302/10 *Infopaq II*, kohta 54. Sama on myös jo aikaisemmin esitetty yhdistetyissä asioissa C-403/08 ja C-429/08 *Premier League*, kohdassa 177.

²²⁸ *Triaille – de Meeûs d’Argentuil – de Francquen* 2014, s. 47.

²²⁹ Yleisenä tekijänoikeudellisena periaatteena on, että oikeudenhaltijalla on yksinoikeus teoksensa muuttamiseen. Tämä on vahvistettu esimerkiksi Bernin yleissopimuksessa. Sen 12 artiklan mukaan kirjallisten, tieteellisten ja taiteellisten teosten tekijöillä on yksinomainen oikeus sallia teostensa muunteleminen, sovittaminen tai muulla tavoin muuttaminen.

²³⁰ Asia C-302/10 *Infopaq II*, kohta 54. Myös yhdistetyissä asioissa C-403/08 ja C-429/08 *Premier League*, kohdassa 177 todetaan sama.

louhimaan siitä uutta informaatiota. Voidaan esittää, että hakukonetoiminta täyttää tämän vaatimuksen.

Näiden kahden taloudellisen vaatimuksen tarkastelu osoittaa, että niiden tulkintaan liittyy paljon epäselvyyttä eikä löydy tarpeeksi oikeuskäytäntöä selkeyttämään oikeustilaa. Esimerkiksi *Strowel* on perustellusti kritisoinut sitä, että taloudellisen arvon analysointi on ollut puutteellista Euroopan unionin tuomioistuimen ratkaisukäytännössä. On myös todettava, ettei kappaleen valmistamista koskevan yksinoikeuden laaja tulkinta ole tasapainossa tietoyhteiskuntadirektiivin sisältämien muiden poikkeuksien soveltamisen kanssa. Kyseisiä poikkeuksia ei olla onnistuneesti sovellettu taloudellisen kriteerien valossa, vaikka monet poikkeukset sisältävät maininnan siitä, että arvioinnissa tulisi ottaa huomioon taloudellista seikkoja.²³¹

Depreeuw ja *Strowel* ovat esittäneet, että Euroopan unionin oikeuskäytännössä tulisi ottaa arviointikriteeriksi se, miten teosta hyödynnettään ja onko kyseisellä teoksen omaperäisen ilmaisuuden hyödyntämisellä taloudellista merkitystä.²³² Tässä tutkielmassa nojaututaan myös tähän näkemykseen ja esitetään, että uusien teknologioiden ja niiden aikana tehtävien kopioiden analysoinnissa tulisi keskittyä siihen, *mihin* teknologian arvo perustuu. Jos arvo tulee siitä, että hyödynnetään teoksen alkuperäistä ilmaisutapaa, tulisi seuraavaksi arvioida, merkitseekö se sellaisen lisäedun saamista, joka ylittää suojatun teoksen laillisesta käytöstä saadun edun. Hakukoneiden kohdalla on kuitenkin todettava, ettei siinä käytettävän tekstin- ja tiedonlouhintateknologian arvo perustu alkuperäisen ilmaisutavan hyödyntämiselle, vaan ainoastaan analysoitavassa materiaalissa olevalle informaatiolle. Tämän vuoksi tutkielmassa esitetään, ettei ole edes tarpeen arvioida tarkemmin indeksointiprosessin yhteydessä toteuttavien kopioiden taloudellista merkitystä.

Tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan sisältämän yleisen tekijänoikeuden rajoitusnormin yksityiskohtaisempi analysointi osoittaa, ettei voida yksiselitteisesti vastata siihen, miten rajoitusnormi soveltuu hakukonetoimintaan tai ylipäänsä tekstin- ja tiedonlouhintaan. Rajoitusnormin soveltuminen on nimittäin riippuvainen monista yksittäisistä seikoista.²³³ Kyseinen rajoitusnormi on joustamaton eikä kykene huomioimaan tekstin- ja tiedonlouhintateknologian

²³¹ *Strowel* 2018, s. 216. Tämä käynee ilmi myös ratkaisuista *Infopaq II* ja *Premier League*. Analysointi on kuitattu ainoastaan maininnalla siitä, ”että teknisessä prosessissa tuotetut kappaleen valmistamista koskevat toimet mahdollistavat pääsyn suojattuihin teoksiin. Koska alkuperäisillä suojatuilla teoksilla on mitattavissa oleva taloudellinen arvo, on myös niihin pääsyllä taloudellinen merkitys.” Ks. tarkemmin asia C-302/10 *Infopaq II*, kohta 48. Myös yhdistetyissä asioissa C-403/08 ja C-429/08 *Premier League*, kohdassa 174 todetaan sama.

²³² *Deepreuw* 2014, s. 67 sekä *Strowel* 2018, s. 216.

²³³ *Triaille – de Meeûs d’Argentuil – de Francquen* 2014, s. 46.

ominaispiirteitä, ja etenkin indeksoinnin yhteydessä tuotettavien kopioiden arvioiminen rajoituskriteerien valossa on monitulkintaista. Haasteita liittyy etenkin puutteelliseen taloudelliseen arviointiin.

Näin ollen on selkeää, että digitaalinen ympäristö on herättänyt muutostarpeen tekijänoikeuslainsäädännölle. Seuraavaksi luvussa 5 käsitellään ja analysoidaan Euroopan unionissa omaksettua ratkaisua tekstin- ja tiedonlouhinnan ajamaan muutostarpeeseen.

5 DIGITALISAATIOKEHITYS JA TEKIJÄNOIKEUSDIREKTIIVI

5.1 Digitalisoituva Eurooppa ja tekijänoikeusdirektiivi

5.1.1 Taustaa

Tekijänoikeuden kappaleen valmistamista koskevan oikeuskäytännön sekä sen taustalla olevan yhteiskunnallisen ilmiön käsittely edellä ovat osoittaneet selkeästi, ettei hakukoneiden suorittamaa kaupallista tekstin- ja tiedonlouhintaa voida yksiselitteisesti sovittaa yhteen tekijänoikeuden kanssa. Myös eri maissa omaksutut ratkaisut tekstin- ja tiedonlouhintaan liittyvissä käytännöissä ovat muodostaneet riskin tekstin- ja tiedonlouhintateknologian kehittymiselle Euroopan unionissa ja heikentäneet sen kilpailullista asemaa.²³⁴ Näin ollen Euroopan unionissa on ollut tarpeen kehittää ratkaisu tekijänoikeuden ja tekstin- ja tiedonlouhinnan välisen jännitteen purkamiseksi.²³⁵

Digitalisaatiokehityksen aiheuttamiin laajoihin ja nopeisiin muutoksiin on Euroopan unionissa reagoitu *Digitaalisten sisämarkkinoiden strategialla Euroopalle* sekä myöhemmin osana tätä strategiaa tekijänoikeusdirektiivillä.²³⁶ Tekijänoikeusdirektiiviä säädettäessä on tunnistettu, että teknologisen kehityksen myötä tekijänoikeuden ja lähioikeuden on mukauduttava vastaamaan riittävällä tasolla taloudellisia tosiseikkoja, kuten uusia hyödyntämismuotoja. Tekijänoikeusdirektiivin harmonisoinnilla on pyrkimys saavuttaa eri oikeudenhaltijaryhmien ja suojatun aineiston käyttäjien välisten oikeuksien ja etujen välinen oikeudenmukainen tasapaino.²³⁷ Tasapainon saavuttamiseksi lähestymistavaksi valittiin tekijänoikeutta ja lähioikeutta rajoittavien poikkeuksien listan kasvattaminen tekijänoikeusdirektiivissä.

Tämän tutkielman kannalta erityisesti tekijänoikeusdirektiivin artikkelit 3 ja 4 ovat keskeisiä. Ne nimittäin sisältävät tekstin- ja tiedonlouhintaan sovellettavat poikkeukset, joiden tarkoituksena on selventää tekstin- ja tiedonlouhintaan liittyviä epävarmuustekijöitä. Tekijänoikeusdirektiivin

²³⁴ Tekijänoikeusdirektiivi, resitaali 10. Myös *Filippov* ja *Hofheinz* ovat esittäneet huolen siitä, että Euroopassa vallitseva oikeudellinen epävarmuus on johtanut siihen, että Eurooppa on jäämässä jälkeen tekstin- ja tiedonlouhintaan perustuvassa tieteellisessä kehityksessä. Tutkimukset osoittavat, että Eurooppa on esimerkiksi Yhdysvaltoja ja Kiinaa jäljessä tekstin- ja tiedonlouhintaan liittyvissä julkaisuissa ja patenteissa. Ks. lisää *Filippov – Hofheinz* 2016, s. 3–8 ja 14. Aiheesta myös *Samuelson* 2018.

²³⁵ Euroopan komissio on tunnistanut, että tekstin- ja tiedonlouhinnan edistämiseksi yhteiskunnassa on neljä eri lähestymistapaa. Ensinnäkin on esitetty, että tekijänoikeuden lisensointijärjestelmiä olisi muutettava selkeämmiksi. Toiseksi on esitetty, että tekijänoikeuslainsäädäntöön olisi otettava avoin tulkintanormi. Kolmas lähestymistapa jännitteeseen on ollut, että kappaleen valmistamista koskevaa oikeutta olisi tulkittava uudestaan. Neljänneksi on ehdotettu, että olisi säädettävä uusi harmonisoitava poikkeus tekijänoikeus- ja tietokantalainsäädäntöön. Ks. lisää *Euroopan komission tutkimuksen ja innovoinnin pääosaston asiantuntijalausunto* 2014, s. 5.

²³⁶ *Digitaalisten sisämarkkinoiden strategia Euroopalle*, s. 3–4.

²³⁷ Tietoyhteiskuntadirektiivi, resitaalit 5–6.

3 artikla koskee tieteellistä tekstin- ja tiedonlouhintaa, kun 4 artikla koskee puolestaan muihin tarkoituksiin harjoitettua tekstin- ja tiedonlouhintaa. Seuraavaksi tarkastellaan kyseisten artikloiden sisältöä tämän tutkielman kannalta relevantein osin sekä analysoidaan, miten niissä on ratkaistu edellä tutkielmassa esitetyt tekstin- ja tiedonlouhintaan liittyvät haasteet.

5.1.2 Tekstin- ja tiedonlouhinta tieteellistä tutkimusta varten

5.1.2.1 Tekijänoikeusdirektiivin 3 artikla

Tekijänoikeusdirektiivin 3 artiklassa määritellään ne edellytykset, joiden perusteella tieteellistä tarkoitusta varten suoritettu tekstin- ja tiedonlouhinta on sallittua. Kyseisen artiklan mukaan:

”1. Jäsenvaltioiden on säädettävä poikkeuksesta direktiivin 96/9/EY 5 artiklan a alakohdassa ja 7 artiklan 1 kohdassa, direktiivin 2001/29/EY 2 artiklassa ja tämän direktiivin 15 artiklan 1 kohdassa säädettyihin oikeuksiin, kun kyseessä on kappaleen valmistaminen ja kopiointi, joita tutkimusorganisaatiot ja kulttuuriperintölaitokset toteuttavat suorittaakseen tieteellistä tutkimusta varten tekstin- ja tiedonlouhintaa teoksista tai muusta suojatusta aineistosta, joihin niillä on laillinen pääsy.

2. Teoksista tai muusta suojatusta aineistosta 1 kohtaa noudattaen valmistetut kappaleet ja kopiot on säilytettävä asianmukaista suojaustasoa käyttäen, ja niitä voidaan säilyttää tieteellistä tutkimusta varten, mukaan lukien tutkimustulosten todentamiseen.

3. Oikeudenhaltijoiden on voitava soveltaa toimenpiteitä, joilla varmistetaan niiden verkkojen ja tietokantojen turvallisuus ja eheys, joissa teoksia tai muuta suojattua aineistoa säilytetään. Tällaiset toimenpiteet eivät saa ylittää sitä, mikä on tarpeen tämän tavoitteen saavuttamiseksi.

4. Jäsenvaltioiden on kannustettava oikeudenhaltijoita sekä tutkimusorganisaatioita ja kulttuuriperintölaitoksia määrittelemään 2 kohdassa tarkoitettua velvoitteen ja 3 kohdassa tarkoitettujen toimenpiteiden soveltamista koskevia yhteisesti sovittuja parhaita käytäntöjä.”

Artikla on luonteeltaan pakollinen, eli jäsenvaltioiden on implementoitava sen sisältö kansalliseen lainsäädäntöönsä.²³⁸ Seuraavaksi siirrytään yksityiskohtaisemmin tarkastelemaan artiklan asettamia vaatimuksia.

²³⁸ Tekijänoikeusdirektiivi, resitaalit 5, 9, 11 ja 18.

5.1.2.2 Aineellinen ja organisatorinen soveltamisala

Artiklan ensimmäisessä kohdassa määritetään niin artiklan aineellinen kuin organisatorinen soveltamisala. Tekstin- ja tiedonlouhinnan tieteellinen poikkeus on siis säädettävä koskemaan ensinnäkin tietokannan tekijän yksinoikeutta sekä johonkin jäsenvaltioon sijoittautuneen lehtijulkaisujen kustantajan yksinoikeutta.²³⁹ Lisäksi poikkeus on ulotettava koskemaan tietoyhteiskuntadirektiivin 2 artiklassa määriteltyä kappaleen valmistamista koskevaa yksinoikeutta, jonka sisältöä on tarkasteltu jo yksityiskohtaisesti tämän tutkielman 3.1 luvussa. On myös huomattava, että artikla rajaa tekstin- ja tiedonlouhinnan ainoastaan kappaleen valmistamiseen. Tässä tutkielmassa tämän tulkitaan tarkoittavan sitä, että yleisölle välittäminen on kiellettyä.²⁴⁰

Aineellista soveltamisalaa määriteltäessä on myös keskeistä tarkastella, missä tilanteissa eri organisaatioilla on oikeus suorittaa tekstin- ja tiedonlouhintaa. Artiklan ensimmäisessä kohdassa määritellään, että tekstin- ja tiedonlouhinta on sallittua tieteellistä tutkimusta varten. Tekijänoikeusdirektiivin johdanto-osan perustelukappaleessa tieteelliseen tutkimukseen sisällytetään niin luonnontieteet kuin humanistiset tieteet.²⁴¹

Artiklan soveltamisen ulkopuolelle on siis rajattu kaikki muu toiminta kuin tieteellinen tutkimus, jonka taustalla ei ainakaan suoraan saisi olla voiton tavoittelu ja muut liikeyrityksille ominaiset taloudelliset intressit.²⁴² Tämä tarkoittaa sitä, ettei artikla sovellu hakukoneiden suorittamaan kaupalliseen tekstin- ja tiedonlouhintaan. Nimittäin hakukonetoiminnan taustalla vaikuttavat selkeästi taloudelliset intressit. Hakukoneet saavat tuloja esimerkiksi esittämistään mainoksista sekä hakutuloksien suuntaamisesta. Lisäksi on keskeistä huomioda Googlen kannalta, että se saa hakukonetoiminnan avulla myös kerättyä hakukoneen käyttäjistä arvokasta big dataa, jota se voi hyödyntää muilla kaupallisilla alustoillaan.²⁴³

²³⁹ Lehtijulkaisujen kustantajan yksinoikeus on uusi tekijänoikeusdirektiivissä tunnistettu lähioikeus. Tähän yksinoikeuteen liittyvää problematiikkaa ei kuitenkaan käsitellä perusteellisemmin tässä tutkielmassa. Ks. lisää esimerkiksi Tekijänoikeusdirektiivi, resitaali 12; *Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: oikeudellinen analyysi* 2018, s. 20; sekä *Geiger – Bulayenko – Frosio* 2018, s. 28.

²⁴⁰ Tutkielmassa otetun rajauksen vuoksi tutkielmassa ei käsitellä enempää yleisölle välittämiseen liittyvää kokonaisuutta.

²⁴¹ Tekijänoikeusdirektiivi, resitaali 12.

²⁴² Tekijänoikeusdirektiivin 2 artiklan 1 a) kohta.

²⁴³ *Ezrachi* ja *Stucke* ovat kirjoittaneet ansiokkaasti alustataloudesta ja siitä, miten digitalisoituminen on muuttanut yhteiskuntaa ja yritysten välistä kilpailua. Osana tätä tutkimusta he ovat selvittäneet, miten esimerkiksi Google pystyy monipuolisesti kaupallisesti hyödyntämään keräämäänsä dataa. Oleellista on myös tunnistaa, kuinka valtavia määriä dataa Google kerää eri lähteistä. Google ei tarjoa pelkästään hakukonetoimintaa, vaan kokonaisen ekosysteemin käyttäjillensä. Ks. lisää aiheesta *Ezrachi – Stucke* 2016.

Tieteellistä tutkimusta koskeva vaatimus on herättänyt erilaisia kannanottoja lainsäädäntövaiheessa. Kulttuuri- ja koulutusvaliokunta (”CULT”) on ottanut omassa lausunnossaan tiukan kannan ja katsonut, että tekstin- ja tiedonlouhinnan organisatorisen soveltamisalan tulisi olla rajattu tiukasti tieteelliseen tutkimukseen.²⁴⁴ Teollisuus-, tutkimus- ja energiavaliokunta (”ITRE”) sekä Sisämarkkina- ja kuluttajansuojalautakunta (”IMCO”) ovat puolestaan edellyttäneet, ettei tekstin- ja tiedonlouhintaa tulisi rajata ainoastaan tieteelliseen tutkimukseen.²⁴⁵ Varsinkin IMCO on korostanut, ettei kapea tutkimusorganisatorinen määritelmä ole tarkoituksenmukainen ja on kannattanut määritelmää, jolla ei tehdä eroa käyttäjien eikä tarkoitusten välillä sekä esittänyt, että lukuoikeuden tulisi merkitä samalla myös louhintaoikeutta.²⁴⁶ Myös Oikeudellisten asioiden valiokunta (”JURI”) on esittänyt, että tekstin- ja tiedonlouhinnan edunsaajien joukon tulisi olla laaja, ja ettei tekstin- ja tiedonlouhinnan tulisi olla rajattua vain tiettyyn tarkoitukseen.²⁴⁷ Näistä kannanotoista huolimatta tekijänoikeusdirektiivin 3 artiklassa on päädytty kapeaan soveltamisalan organisatoriseen rajaukseen.

Tekijänoikeusdirektiivin ensimmäisessä kohdassa nimittäin määritetään, että artiklan mukainen poikkeus soveltuu ainoastaan tutkimusorganisaatioihin ja kulttuuriperintölaitoksiin. Tekijänoikeusdirektiivissä tutkimusorganisaatiolla tarkoitetaan yliopistoja, korkea-asteen oppilaitoksia, niiden kirjastoja, tutkimuslaitoksia, ja niiden kaltaisia yhteisöjä ja sairaaloita, jotka harjoittavat tutkimustoimintaa. Näillä orgaaneilla voi olla erilaisia oikeudellisia muotoja ja rakenteita, mutta olennaiseksi kriteeriksi on nostettu se, että niiden toiminnan tulee olla joko voittoa tavoittelematonta tai niiden toiminnan tulee samalla toteuttaa valtion tunnustamaa yleisen edun mukaista tehtävää. Yleisen edun mukainen tehtävä on kuitenkin tulkinnalle avoin määritelmä. On esitetty, että se voisi ilmetä esimerkiksi julkisessa rahoituksessa taikka kansallisten lakien säännöksissä tai hankintasopimusten määräyksissä.²⁴⁸

²⁴⁴ Kulttuuri- ja koulutusvaliokunnan lausunto oikeudellisten asioiden valiokunnalle ehdotuksesta Euroopan parlamentin ja neuvoston direktiiviksi tekijänoikeudesta digitaalisilla sisämarkkinoilla (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)). 4.9.2017. Tarkistukset 4, 42, 44 sekä 45.

²⁴⁵ Teollisuus-, tutkimus- ja energiavaliokunnan lausunto oikeudellisten asioiden valiokunnalle ehdotuksesta Euroopan parlamentin ja neuvoston direktiiviksi tekijänoikeudesta digitaalisilla sisämarkkinoilla (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)). 1.8.2017. Tarkistukset 5 ja 35.

²⁴⁶ Sisämarkkina- ja kuluttajansuojavaliokunnan lausunto oikeudellisten asioiden valiokunnalle ehdotuksesta Euroopan parlamentin ja neuvoston direktiiviksi tekijänoikeudesta digitaalisilla sisämarkkinoilla (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)). 14.6.2017. Tarkistukset 6 ja 7.

²⁴⁷ Mietintö ehdotuksesta Euroopan parlamentin ja neuvoston direktiiviksi tekijänoikeudesta digitaalisilla sisämarkkinoilla (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)) Oikeudellisten asioiden valiokunta, 29.6.2018. Tarkistukset 8, 9, 55, 64 sekä 65.

²⁴⁸ Tekijänoikeusdirektiivin resitaali 12 sekä *Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: oikeudellinen analyysi* 2018, s. 20.

Tekijänoikeusdirektiivin tutkimusorganisaation määritelmään eivät näin ollen sisälly organisaatiot, joissa liikeyrityksillä on ratkaiseva vaikutusvalta, jonka perusteella liikeyritykset voisivat käyttää määräysvaltaa tutkimusorganisaatioissa joko osakkeenomistajina tai osakkaina tai vastaavien rakenteiden avulla, ja syntyisi epäily liikeyrityksen vaikuttamisesta tutkimustuloksiin.²⁴⁹ Soveltamisalaa tulkittaessa on selkeää, että puhtaasti kaupalliset tarkoitukset rajautuvat pois. On kuitenkin tunnistettava, että tieteellistä tutkimusta tehdään myös osana yksityisen sektorin kanssa tehtäviä hankkeita ja projekteja.

Tiettyissä määrin tekijänoikeusdirektiivissä on onnistuneesti tunnistettu yksityisen sektorin läsnäolo tieteellisissä tutkimuksissa, esimerkiksi yliopistojen ja tutkimuslaitoksien yhteistyöt yksityisen sektorin kanssa. Tutkimusorganisaatiot nimittäin hyötyvät 3 artiklan poikkeuksesta myös tilanteissa, kun ne harjoittavat tutkimustoimintaa julkisen ja yksityisen sektorin kumppanuuden puitteissa. Tämä mahdollistaa sen, että tutkimusorganisaatiot ja kulttuuriperintölaitokset saavat käyttää yksityisiä kumppaneitaan tekstin- ja tiedonlouhinnan suorittamisessa myös niiden teknisiä välineitä käyttämällä.²⁵⁰ Tutkimusorganisaatioiden kumppanuus muiden toimijoiden kanssa vaatii kuitenkin sisältönsä ja laajuutensa puolesta täsmennystä.

Kulttuuriperintölaitoksen määritelmän alle sisällytetään tekijänoikeusdirektiivissä yleisölle avoimet kirjastot, kansalliskirjastot, museot, arkistot, kansallisarkistot sekä elokuva- tai äänitearkistot. Kulttuuriperintölaitoksiin katsotaan myös kuuluvan, siltä osin kuin kyse on niiden arkistoista ja yleisistä kirjastoista, oppilaitokset, tutkimusorganisaatiot sekä julkisen sektorin

²⁴⁹ Tekijänoikeusdirektiivi, resitaali 12.

²⁵⁰ Tekijänoikeusdirektiivi, resitaali 11.

yleisradio-organisaatiot.²⁵¹ Organisatoriseen soveltamisalaan sisällytetään myös tutkimusorganisaatioihin ja kulttuuriperintölaitoksiin yhteydessä olevat henkilöt.²⁵² Organisatorinen kuvaus kattaa suhteellisen laajasti erilaisia tieteellisiä toimijoita. Rajaus herättää kuitenkin kritiikkiä.

Tekijänoikeusdirektiivin 3 artiklassa otettu rajaus organisatoriseen sekä aineelliseen soveltamisalaan on ongelmallinen useasta syystä. Ensinnäkin se rajaa täysin sekä muiden toimijoiden suorittaman että kaupallisen tekstin- ja tiedonlouhinnan sen ulkopuolelle. Tässä tutkielmassa asetetaan IMCO:n kanssa samalle kannalle siitä, ettei tekstin- ja tiedonlouhintaa määriteltäessä olisi saanut tehdä erottelua eri käyttäjien ja tarkoitusten välillä. Organisatorinen rajaus voi heikentää start up -toimijoiden sekä yksittäisten tutkijoiden mahdollisuuksia suorittaa tieteellistä tutkimusta. Määritelmä jättää myös täysin avoimeksi sen, saavatko esimerkiksi journalistit ja muut media-alan toimijat harjoittaa tekstin- ja tiedonlouhintaa esimerkiksi vähentääkseen valeutisia.²⁵³

Toiseksi organisatorisen rajauksen tarkempi tarkastelu edellä on paljastanut, ettei aina voida yksiselitteisesti määritellä täyttääkö toimija 3 artiklan vaatimukset tutkimusorganisaatiosta taikka kulttuuriperintölaitoksesta. Artikla jättää siis edelleen oikeustilan epäselväksi sen suhteen, mitkä toimijat rajanvetotapauksissa hyötyvät artiklasta ja voivat harjoittaa tekstin- ja tiedonlouhintaa artiklan perusteella. Avoimet ilmaisut johtavat siihen, että oikeustilan selkeyttämiseksi tarvitaan oikeuskäytäntöä Euroopan unionin tuomioistuimelta. Lisäksi vaarana on myös tekijänoikeudellisen normiston pirstaloituminen Euroopan unionissa, kun jäsenvaltiot implementoivissaan direktiiviä tulkitsevat eri tavoin direktiivin avoimia ilmaisuja.

²⁵¹ Tekijänoikeusdirektiivi, resitaali 13.

²⁵² Tekijänoikeusdirektiivi, resitaali 14. Lisäksi katso *European Copyright Society* 2017, s. 3. Aikaisemmassa tekijänoikeusdirektiiviehdotuksessa (ehdotuksessa Euroopan parlamentin ja neuvoston direktiivi tekijänoikeudesta digitaalisilla sisämarkkinoilla COM(2016) 593 final 2016/0280 (COD)) ei oltu huomioitu tutkimusorganisaatioissa ja kulttuuriperintölaitoksien yhteydessä toimivia henkilöitä. *European Copyright Society* on kritisoinut perustellusti sitä, että kyseisessä ehdotuksessa kaupallisen tekstin- ja tiedonlouhinnan täysi rajaaminen poikkeuksen ulkopuolelle aiheuttaisi käytännössä sen, etteivät esimerkiksi journalistit ja opettajat voisi suorittaa tekstin- ja tiedonlouhintaa ilman tekijänoikeuden haltijan lupaa. Nykyisessä muodossaan olevassa tekijänoikeusdirektiivi ja erityisesti sen artikla 4, mahdollistavat tietyin edellytyksin myös edellä mainittujen henkilöiden suorittaman kaupallisen tekstin- ja tiedonlouhinnan. Jäljempänä esitetyllä tavalla ilmenee kuitenkin, että kaupalliseen tekstin- ja tiedonlouhintaan sisältyy edelleen merkittäviä rajoituksia, jotka voivat vaikuttaa kielteisesti toimittajien ja opettajien suorittamaan tekstin- ja tiedonlouhintaa, mikäli henkilöt eivät sisälly 3 artiklan tutkimusorganisaatioiden tai kulttuuriperintölaitoksien henkilömäärittelyyn taikka heidän harjoittama tekstin- ja tiedonlouhinta on luonteeltaan kaupallista.

²⁵³ *Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: oikeudellinen analyysi* 2018, s. 20.

5.1.2.3 Laillinen pääsy

Tekijänoikeusdirektiivin 3 artiklan sisältämän aineellisen ja organisatorisen soveltamisalan määrittelyn jälkeen voidaan siirtyä tulkitsemaan, millaisia vaatimuksia hyödynnettävälle materiaalille asetaan. Tekijänoikeusdirektiivin 3 artiklan 1 kohdassa mainitaan, että tieteellistä tutkimusta varten tekstin- ja tiedonlouhintaa voidaan harjoittaa materiaalista, johon toimijalla on *laillinen pääsy*. Kyseinen vaatimus ottaa suoraan kantaa aikaisemmin luvussa 2.1 esitettyyn tekstin- ja tiedonlouhintaprosessin ensimmäiseen vaiheeseen, eli materiaaliin pääsyyn.

Tekijänoikeusdirektiivin johdanto-osan perustelukappaleessa määritetään, että laillisella pääsyllä tarkoitetaan sitä, että toimijoilla on pääsy sisältöön avointa saatavuutta koskevien toimintaperiaatteiden (*”open access policy”*) nojalla taikka oikeudenhaltijoiden ja tutkimusorganisaatioiden ja/tai kulttuuriperintölaitosten välisien sopimusjärjestelyjen²⁵⁴ kautta taikka muilla laillisilla keinoilla.²⁵⁵ Lisäksi laillinen pääsy katsotaan olevan myös materiaaliin, johon verkossa on vapaasti pääsy.²⁵⁶

Oikeuskirjallisuudessa on kritisoitu sitä, ettei laillisen pääsyn -vaatimusta ole tarkemmin määritelty tekijänoikeusdirektiivissä.²⁵⁷ Näin ollen on myös epäselvää, voidaanko kyseisen vaatimuksen tulkintaan hyödyntää tietoyhteiskuntadirektiivin 5 artiklan 1 kohdassa mainittua ilmaisua *laillinen käyttö*, jota on jo tämän tutkielman luvussa 4.2.3 analysoitu. On huomattava, että kyseisen luvun analysoinnissa on esitetty lopputulos, ettei ilmaisun tarkkaa sisältöä tiedetä ja sitä käsittelevä oikeuskäytäntö on tulkinnalle altista.²⁵⁸ Vaikkakin laillinen pääsy -vaatimusta ei olla selkeästi määritelty, sen soveltamiseen tekstin- ja tiedonlouhintaan liittyy jo nyt ongelmia. Seuraavaksi siirrytään tarkemmin tarkastelemaan, mihin ongelmat liittyvät ja niiden mahdollisia seurauksia.

²⁵⁴ Tekijänoikeusdirektiivin resitaalissa todetaan, että sopimusjärjestelytilanteet kattavat myös tutkimusorganisaatioiden ja kulttuuriperintölaitosten alla toimivat henkilötahot. Ks. lisää tekijänoikeusdirektiivi, resitaali 14.

²⁵⁵ Tekijänoikeusdirektiivin resitaalin mukaan tällaisia sopimusjärjestelyitä ovat esimerkiksi erilaiset tilaukset. Ks. lisää tekijänoikeusdirektiivi, resitaali 14.

²⁵⁶ Tekijänoikeusdirektiivi, resitaali 14.

²⁵⁷ Aiheesta lisää *Geiger – Bulayenko – Frosio* 2018, s. 32.

²⁵⁸ Tietoyhteiskuntadirektiivin 33 resitaalissa mainitaan, että käyttöä on pidettävä laillisena, jos se on oikeudenhaltijan sallimaa taikka sitä ei ole rajoitettu lailla. Ks. myös yhdistetyt asiat C-403/08 ja C-429/08 *Premier League*, kohta 168 sekä asia C-302/10 *Infopaq II*, kohta 42.

Vaatus laillisesta pääsystä materiaaliin on tieteellisen tekstin- ja tiedonlouhinnan näkökulmasta kahdesta syystä ongelmallinen. Ensinnäkin se tekee tekstin- ja tiedonlouhinnan ensimmäisen vaiheen, pääsyn materiaaliin, yksityistilauksesta (*”private ordering”*) riippuvaiseksi. Seurauksena on, että tekstin- ja tiedonlouhinnassa hyödynnettävä materiaali voidaan rajata tehokkaasti tietyiltä käyttäjiltä suoraan oikeudenhaltijoiden toimesta, jos oikeudenhaltijat kieltäytyvät myöntämästä pääsystä materiaaliin taikka myöntävät pääsyn, mutta rajoitetuin taikka kohtuuttomin ehdoin.²⁵⁹ Näin ollen lisensoinnille altistaminen ei selkeytä oikeustilaa. Vaikutus on päinvastainen: lisensointi lisää oikeustilan epäselvyyttä tuomalla siihen mahdollisesti monimutkaisia ja laajoja sopimusehtoja. Materiaaliin pääsyyn vaikuttaa näin ollen myös merkittävästi toimijan neuvotteluvoima markkinoilla.²⁶⁰

Toiseksi yksityistilaukseen siirtymisen seurauksena voi olla, että oikeudenhaltijat sisällyttävät tilausmaksuihinsa tekstin- ja tiedonlouhinnalle arvioidut kustannukset.²⁶¹ Käytännössä kustannuksien lisääminen tilausmaksuihin voi johtaa tekstin- ja tiedonlouhintahankkeiden kustannuksien nousuun sekä rajoittaa tiettyjen toimijoiden mahdollisuuksia hankkia lisenssejä, vaikka lisenssien avulla saatava materiaali olisikin tarpeen tutkimusta varten. Vaarana on, ettei tutkimuksessa voida hyödyntää riittävällä laajuudella materiaalia kustannussyistä. Luvussa 2.1 on esitetty, että pääsyn rajoittamisen seurauksena voi myös olla, että toimijat joutuvat etenevissä määrin turvautumaan ilmaisiin lähteisiin, joissa tieto voi kuitenkin olla viallista. Näin ollen tutkimuksen laatu voi heiketä.

Tekstin- ja tiedonlouhinnassa käytettävän materiaalin laajuuden lisäksi kustannuksien nousu vaikuttaa tekstin- ja tiedonlouhintaa harjoittaviin toimijoihin. Kustannuksien nouseminen rajaa pois ne toimijat, joilla ei ole riittävästi markkinavoimaa. Vain harvat tutkimusorganisaatiot voivat hankkia lisenssejä kaikille tekstin- ja tiedonlouhintateknologian kannalta merkitykselliseen aineistoon. Ongelma koskee Euroopan unionissa etenkin sellaisia jäsenvaltioita, joilla on rajoitetummat mahdollisuudet saada rahoitusta. Tämä vuorostaan kasvattaisi entisestään kuilua erivaraisten tutkimuslaitosten välillä ja lisäksi eroja jäsenvaltioiden välillä tieteen ja innovaation kehityksessä.²⁶² Pääsyn rajoittamisesta seurauksena voi siis olla, että toiminta keskittyy erityisesti toimijoille, joilla on paljon markkinavoimaa.

²⁵⁹ Katso lisää *European Copyright Society* 2017, s. 4.

²⁶⁰ Katso lisää *Zeybek*, 2018.

²⁶¹ Katso lisää *European Copyright Society* 2017, s. 4.

²⁶² *Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: oikeudellinen analyysi* 2018, s. 22 sekä *Geiger – Bulayenko – Frosio* 2018, s. 29.

Tekstin- ja tiedonlouhintateknologian altistaminen lisensoinnille rajoittaa tiedettä tarpeettomasti eikä sen kuuluisi olla oikeudenhaltijoiden kontrollin, toimijoiden markkinavoiman taikka epäselkeiden lisensointijärjestelyiden rajoittamaa. Seurauksena on tekstin- ja tiedonlouhinnan keskittyminen tietyille toimijoille, tieteellisen tutkimuksen laadun heikentyminen sekä myös negatiiviset vaikutukset jakamistalouden kehittymiselle²⁶³. Pääsy materiaaliin on tekstin- ja tiedonlouhinnan kannalta välttämätön vaihe. On kuitenkin selkeää, ettei direktiivissä otettu lähestymistapa poista lisensointiin liittyviä haasteita, vaan päinvastoin, se altistaa harmonisoitavan poikkeuksen nojalla tekstin- ja tiedonlouhinnan rajoitettaville lisensointijärjestelyille ja epäselville sopimusehdoille. Tekstin- ja tiedonlouhintaan vaikuttavat myös tietyt turvallisuustoimenpiteet. Seuraavaksi siirrytään tarkastelemaan, miten tekijänoikeusdirektiivin artikla suhtautuu kopioiden säilyttämiseen ja teknologisiin rajoituksiin.

5.1.2.4 Turvallisuustoimenpiteet

Luvussa 4.1 todettiin, että *three step* -testin yhtenä vaatimuksena on, etteivät rajoitukset ja poikkeukset saa kohtuuttomasti vahingoittaa oikeudenhaltijan oikeutettuja etuja. Näin ollen on perusteltua, että tekijänoikeusdirektiivin 3 artiklan toinen kohta sisältää myös vaatimuksia tekstin- ja tiedonlouhintaprosessiin liittyvistä turvallisuustoimenpiteistä. Tässä tutkielmassa tarkoitetaan turvallisuustoimenpiteillä sekä kopioiden turvalliseen säilyttämiseen liittyviä seikkoja että teknologisia rajoituksia.

Turvallisuustoimenpiteitä on perusteltu ensinnäkin sillä, että oikeudenhaltijat saavat suuria määriä pyyntöjä suojattuun aineistoon. Toiseksi turvallisuustoimenpiteitä perustellaan sillä, että suojattua aineistoa saatetaan ladata määrällisesti useita kertoja. Seurauksena on todellinen riski siitä, että järjestelmien ja tietokantojen turvallisuus ja eheys voivat vaarantua.²⁶⁴ Turvallisuustoimenpiteillä voidaan siis varmistaa, että tekijänoikeudella suojattua materiaalia käytetään ainoastaan siinä laajuudessa kuin on tarpeen sekä ainoastaan poikkeuksen taustalla olevaa tarkoitusta varten.

²⁶³ *European Copyright Society* on esittänyt perustellun huolensa siitä, että jakamistalouden kehittymisen heikentymisellä on vaikutuksia niin kulttuurin, tiedon kuin myös teknologian kehittymiselle. Ks. lisää *European Copyright Society* 2017, s. 4–5.

²⁶⁴ Tekijänoikeusdirektiivi, resitaali 16 sekä esimerkiksi IMCO on esittänyt direktiivin säätämisvaiheessa kantansa siitä, että teknisten suojatoimien käyttö tulisi olla tiukasti rajattua. Ks. lisää aiheesta Sisämarkkina- ja kuluttajan-suojavaliokunnan lausunto oikeudellisten asioiden valiokunnalle ehdotuksesta Euroopan parlamentin ja neuvoston direktiiviksi tekijänoikeudesta digitaalisilla sisämarkkinoilla (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)). 14.6.2017, s. 3.

Ensin käsitellään, mitä vaatimuksia kopioiden säilyttämiseen liittyy. Tämän jälkeen siirrytään tutkimaan tarkemmin, mitä tarkoitetaan teknologisilla rajoituksilla tekijänoikeusdirektiivissä ja miksi ne ovat ongelmallisia tekstin- ja tiedonlouhinnan kannalta.

Tekijänoikeusdirektiivin 3 artiklan toisen kohdan mukaan tieteellistä tekstin- ja tiedonlouhintaa varten valmistetut kappaleet ja kopiot on säilytettävä asianmukaista suojaustasoa käyttäen. Tiettyissä tilanteissa, esimerkiksi tieteellisen tutkimuksen todentamiseksi, saattaa syntyä tarve myöhemmässä vaiheessa palata takaisin tekstin- ja tiedonlouhinnassa hyödynnettyyn materiaaliin. Täten on tärkeää, että tekstin- ja tiedonlouhintaa harjoittavat toimijat voivat säilyttää analysointiprosessin aikana otetut kopiot.²⁶⁵ Tekijänoikeusdirektiivin johdanto-osan perustelukappaleessa asetetaan kuitenkin säilyttämistä koskien sekä laadullisia että ajallisia lisävaatimuksia.

Laadullinen vaatimus koskee tapaa, jolla kopioita säilytetään. Jos kopioiden säilyttäminen on tarpeen, on ne tallennettava suojattuun ympäristöön. Kopioiden säilyttämistä koskevien vaatimusten on kuitenkin oltava oikeasuhteisia ja rajoituttava siihen, mikä on tarpeen kopioiden säilyttämiseksi turvallisesti ja luvattoman käytön estämiseksi.²⁶⁶ Ajallisen vaatimuksen näkökulmasta puolestaan kopioita tulisi siis saada säilyttää niin kauan kuin on tekstin- ja tiedonlouhintaa varten tarpeen. Vaatimus koskee niin lisensoinnin takana olevaa materiaalia kuin verkossa yleisön vapaasti saataviin saatettua materiaalia, johon oikeudenhaltijat eivät olet pidättäneet pääsyä.²⁶⁷ Näiden kopioiden säilyttämistä koskevien vaatimuksien lisäksi voivat oikeudenhaltijat asettaa teknologisia rajoituksia.

Kuten edellä luvussa 2.1 esitettiin, ovat oikeudenhaltijat voineet rajoittaa toimijoiden pääsyä eri medioihin ja tietokantoihin teknologisilla rajoituksilla, eli niin kutsutuilla TPM-rajoituksilla. Tekijänoikeusdirektiivin 3 artiklan 3 kohdan mukaan oikeudenhaltijoiden on mahdollista asettaa edelleen tiettyjä teknologisia rajoituksia (TPM-rajoituksia).²⁶⁸ Rajoituksilla pyritään siihen, että suojattuun aineistoon olisi pääsy ainoastaan sellaisilla henkilöillä, joilla on suojattuun aineistoon myös laillinen pääsy²⁶⁹. Tekijänoikeusdirektiivin johdanto-osan perustelukappaleen

²⁶⁵ Tekijänoikeusdirektiivi, resitaali 15.

²⁶⁶ Tekijänoikeusdirektiivi, resitaali 15.

²⁶⁷ Tekijänoikeusdirektiivi, resitaali 18.

²⁶⁸ Tarkoituksena on varmistaa, että tekijöille ja muille oikeudenhaltijoille Euroopan unionin oikeudessa myönnettyjen oikeuksien suojan toteutuu sekä oikeuksia voidaan tehokkaasti käyttää. Ks. lisää tekijänoikeusdirektiivi, resitaali 7.

²⁶⁹ Laillisella pääsillä tarkoitetaan sitä, että toimijalla on pääsy aineistoon joko sillä perusteella, että materiaali on vapaasti saatavilla taikka sopimusjärjestelyiden perusteella taikka muilla laillisilla keinoilla.

mukaan tällaisia teknologia rajoituksia ovat esimerkiksi IP-osoitteen varmennus tai käyttäjän tunnistauminen.²⁷⁰

Teknologisilta rajoituksilta vaaditaan, että niiden tulisi olla oikeasuhteisia asiaan liittyviin riskeihin nähden eivätkä ne saisi ylittää sitä, mikä on tarpeen järjestelmän turvallisuuden ja eheyden varmistamiseksi, eikä niillä saisi heikentää tekijänoikeusdirektiivin sallimaa tekstin- ja tiedonlouhintaa.²⁷¹ Lähtökohtaisesti vaatimus oikeasuhteisuudesta tuo tietynlaisen kohtuullisuusvaatimuksen turvallisuustoimenpiteisiin. On kuitenkin ongelmallista, että oikeudenhaltijoilla on edelleen oikeus vapaasti valita, mitä asianmukaisia turvallisuustoimenpiteitä ne soveltavat suojaamaan omia oikeuksiaan.²⁷² Seurauksena voi siis olla, että oikeudenhaltijoiden turvallisuustoimenpidekäytännöissä voi olla huomattaviakin eroavaisuuksia, jotka voivat johtaa edelleen oikeustilan pirstaloitumiseen.

Tekijänoikeusdirektiivissä on pyritty tunnistamaan sopimusoikeudelliseen näkökulmaan liittyvät ongelmat. Tekijänoikeusdirektiivin 7 artiklan 1 kohdassa määritellään nimenomaisesti, että oikeudenhaltijat eivät saa sopimusehdoilla tehdä tieteellistä tekstin- ja tiedonlouhintaa koskevan artiklan sisältöä täytöntöönpanokelvottomaksi.²⁷³ On kuitenkin ongelmallista, että sopimusoikeudellinen suoja ei kata kaupallista tekstin- ja tiedonlouhintaa.

Edellä käsitellystä käy ilmi, että tekijänoikeusdirektiivissä on onnistuttu tunnistamaan niitä epäselvyyksiä, joita on liittynyt muun muassa materiaaliin pääsyä rajoittaviin toimenpiteisiin, eli lisensointijärjestelmiin ja TPM-rajoituksiin. Kuitenkin tarkempi tarkastelu osoittaa, että materiaaliin pääsyyn liittyy edelleen epäselvyyttä, joka voi asettaa eri tekstin- ja tiedonlouhintaa harjoittavat toimijat eriarvoiseen asemaan. Myöskään tekijänoikeusdirektiivin omaksuma organisatorinen määritelmä ei anna yksiselitteistä vastausta kaikista tilanteista, joihin tekijänoikeusdirektiivin 3 artikla soveltuu. Keskeisintä on, että tekijänoikeusdirektiivin 3 artikla sulkee kokonaan pois kaupallisen tekstin- ja tiedonlouhinnan. Tekijänoikeusdirektiivin 4 artikla kattaa

²⁷⁰ Tekijänoikeusdirektiivi, resitaali 16.

²⁷¹ Tekijänoikeusdirektiivi, resitaali 16 sekä Kulttuuri- ja koulutusvaliokunnan lausunto oikeudellisten asioiden valiokunnalle ehdotuksesta Euroopan parlamentin ja neuvoston direktiiviksi tekijänoikeudesta digitaalisilla sisämarkkinoilla (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)). 4.9.2017, s. 2.

²⁷² Tekijänoikeusdirektiivi, resitaali 7.

²⁷³ Tekijänoikeusdirektiivin 7 artiklan 1 kohdassa todetaan, että: ”tekijänoikeusdirektiivin 3, 5 ja 6 artiklassa säädettyjen poikkeusten vastaiset sopimusehdot ovat täytöntöönpanokelvottomia. Kaupallisesta tekstin- ja tiedonlouhinnasta säädetään tekijänoikeusdirektiivin 4 artiklassa.”

poikkeuksen, jota sovelletaan, kun tekstin- ja tiedonlouhintaa harjoitetaan muuta kuin tieteellistä tarkoitusta varten. Seuraavassa luvussa käsitellään relevantein osin tarkemmin kyseistä artiklaa.

5.1.3 Tekstin- ja tiedonlouhinta muuta tarkoitusta varten

5.1.3.1 Tekijänoikeusdirektiivin 4 artikla

Tutkielmassa on jo aikaisemmin tunnistettu, ettei tekstin- ja tiedonlouhintateknologioiden hyödyntäminen rajaudu pelkästään tieteelliseen tutkimukseen, vaan kyseistä teknologiaa voidaan hyödyntää laajalti muilla elämän osa-alueilla sekä moniin muihin tarkoituksiin, kuten julkisella sektorilla palveluita tuottaessa, yhtiöiden liiketoiminnassa, uusissa sovelluksissa sekä esimerkiksi koneoppimisen kehittämisessä ja hakukoneissa. Tekstin- ja tiedonlouhinnan rajoittaminen pelkästään tieteelliseen toimintaan olisi yhteiskunnan, tieteen ja innovaation kehityksen vastaista.²⁷⁴ Oikeusvarmuuden lisäämiseksi esimerkiksi kaupallisella sektorilla tekijänoikeusdirektiiviin otettiin myös 4 artikla, jota puolestaan voidaan soveltaa tilanteisiin, joita 3 artikla ei kata. Kyseisen artiklan mukaan:

”1. Jäsenvaltioiden on säädettävä poikkeuksesta tai rajoituksesta direktiivin 96/9/EY 5 artiklan a alakohdassa ja 7 artiklan 1 kohdassa, direktiivin 2001/29/EY 2 artiklassa, direktiivin 2009/24/EY 4 artiklan 1 kohdan a ja b alakohdassa sekä tämän direktiivin 15 artiklan 1 kohdassa säädettyihin oikeuksiin, kun kyseessä on kappaleen valmistaminen ja kopiointi tekstin- ja tiedonlouhintaa varten teoksista ja muusta suojatusta aineistosta, joihin on laillinen pääsy.

2. Edellä olevan 1 kohdan mukaisesti valmistettuja kappaleita ja kopioita voidaan säilyttää niin kauan, kuin tekstin- ja tiedonlouhintaa varten on tarpeen.

3. Edellä 1 kohdassa säädettyä poikkeusta tai rajoitusta sovelletaan edellyttäen, että oikeudenhaltijat eivät ole nimenomaisesti pidättäneet oikeuksiaan kyseisessä kohdassa tarkoitettuun teosten ja muun suojatun aineiston käyttöön asianmukaisella tavalla, kuten koneluettavilla keinoilla, kun kyseessä on verkossa yleisön saataviin saatettu sisältö.

4. Tämä artikla ei vaikuta tämän direktiivin 3 artiklan soveltamiseen.”

Seuraavaksi tarkastellaan artiklan tarkempaa sisältöä yksityiskohtaisesti. Tarkastelussa otetaan erityisesti huomioon hakukoneiden suorittama tekstin- ja tiedonlouhinta. Artikla sisältää osin

²⁷⁴ Tekijänoikeusdirektiivi, resitaali 18.

samoja vaatimuksia kuin edellä käsitelty tekijänoikeusdirektiivin 3 artikla. Kopioiden säilyttämisestä ja laillista pääsyä koskevat vaatimukset ovat samat. Toiston välttämiseksi näitä vaatimuksia ei käsitellä alla. Osin myös artikloiden soveltamisalassa on yhteneväisyyksiä. Tässä tutkielman 5.1.3 luvussa käsitellään ainoastaan mahdollisia eroavaisuuksia sekä kaupallisesta kontekstista johtuvia erityisvaatimuksia soveltamisalaan liittyen.

5.1.3.2 Aineellinen ja organisatorinen soveltamisala

Samoin kuin yllä käsitellyssä tekijänoikeusdirektiivin 3 artiklassa, myös 4 artiklan ensimmäisessä kohdassa käsitellään tilanteet, joissa jäsenvaltioiden tulee säätää tietyistä poikkeuksista.²⁷⁵ Nämä tilanteet ovat muutoin samat kuin 3 artiklassa, mutta 4 artiklan mukaisesti poikkeuksia koskettavat myös tietokoneohjelmien oikeudenhaltijoiden yksinoikeuksia.²⁷⁶ Onkin mielenkiintoista, ettei tietokoneohjelmadirektiivin mukaisia yksinoikeuksia ole sisällytetty tieteellistä tekstin- ja tiedonlouhintaa koskevan artiklan alle.²⁷⁷

Myös muutoin artiklan ensimmäinen kohta vastaa sisällöllisesti hyvin paljon 3 artiklan vastaavaa kohtaa. Eroavaisuutena on kuitenkin se, ettei 4 artiklassa olla tehty organisatorista rajoitusta eikä määritelty sitä, että kappaleen valmistaminen tekstin- ja tiedonlouhintaa varten tulisi tapahtua tieteellistä tarkoitusta varten. Näin ollen organisatorisesti neutraali artikla mahdollistaa tekstin- ja tiedonlouhinnan, jos toimijalla on laillinen pääsy materiaaliin. Kuten luvussa 5.1.2.3 todettiin, on laillisen pääsyn vaatimus useasta syystä ongelmallinen. Lisäksi luvussa 5.1.2.2

²⁷⁵ Euroopan parlamentin tarkistukset 12. syyskuuta 2018 ehdotukseen Euroopan parlamentin ja neuvoston direktiiviksi tekijänoikeudesta digitaalisilla sisämarkkinoilla (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)). P8_TA(2018)0337 Tekijänoikeus digitaalisilla sisämarkkinoilla ***I (*”Euroopan parlamentin muutosehdotus tekijänoikeusdirektiiviin 2018”*). Kyseisessä muutosehdotuksessa ehdotettiin, että kansallisella tasolla kaupallinen tekstin- ja tiedonlouhinta olisi perustunut vapaaehtoisuuteen. Tämä olisi perustunut ehdotuksen 3a artiklaan, jolla annettiin jäsenvaltioille mahdollisuus säätää poikkeuksesta tai rajoituksesta, jota oltaisiin sovellettu sellaisista laillisesti käytettävissä olevista teoksista ja muusta aineistosta valmistettuihin kappaleisiin ja niiden kopioihin, jotka muodostavat osan tekstin- ja tiedonlouhintaprosessia, edellyttäen, että niiden oikeudenhaltijat eivät ole nimenomaisesti pidättäneet oikeutta sen mukaiseen teosten ja muun aineiston käyttöön, myös koneellisesti luettavissa olevin keinoin. Jäsenvaltiot olisivat siis itse saaneet päättää olisivatko ne implementoineet kaupalliseen tarkoitukseen perustuvan tekstin- ja tiedonlouhinnan ja missä laajuudessa. Katso aiheesta tarkemmin *Euroopan parlamentin muutosehdotus tekijänoikeusdirektiiviin 2018*, muutosehdotus 65. Kyseisen artiklan perustuminen vapaaehtoisuuteen olisi ollut oikeusvarmuuteen ja tehokkuuteen harmonisoitavan oikeusjärjestelmän kannalta kestävä ajatus.

²⁷⁶ Ks. tarkemmin tietokoneohjelmadirektiivin 4 artiklan 1 kohdan alakohdat a ja b.

²⁷⁷ Näyttää siltä, että tietokoneohjelmat ovat saaneet erityisen suojan tekijänoikeusdirektiivissä. Tietokoneohjelmista ei voida ottaa kopioita tekijänoikeusdirektiivin 3 artiklan nojalla, vaan ainoastaan tekijänoikeusdirektiivin 4 artiklan asettamissa puitteissa. Käytännössä tämä tarkoittaa sitä, että tieteellisen tutkimuksen kohdalla kopioita voidaan edelleen ottaa tekstin- ja tiedonlouhinnan yhteydessä tietokoneohjelmista, mutta 4 artiklan asettamien tiukempien kriteerien nojalla. Tekijänoikeusdirektiivin johdanto-osan perustelukappaleista eikä lainvalmisteluaineistosta löydy kuitenkaan suoraan perusteita tälle ratkaisulle. On mahdollista, että esimerkiksi lobbauksella on vaikutettu tähän ratkaisuun. Aiheeseen ei kuitenkaan tätä enempää perehdytä tässä tutkielmassa. Tekijänoikeusdirektiivin säätämisaikana vallinneesta voimakkaasta lobbauksesta on hyvin kirjoitettu esimerkiksi lähteessä *Corporate Europe Observatory* 2018.

kyseenalaistettiin, onko ylipäänsä tarpeen tehdä erottelua tieteellisten ja kaupallisten toimijoiden välillä.

5.1.3.3 Oikeuksien pidättäminen asianmukaisella tavalla

Tekijänoikeusdirektiivin 4 artiklan organisatorisen ja aineellisen soveltamisalan käsittely paljasti, että tekijänoikeusdirektiivin 4 artikla kattaa enemmän toimijoita sekä tarkoitusperiä. Kuitenkin tekijänoikeusdirektiivin 4 artiklan 3 kohta rajaa huomattavasti kaupalliseen tekstin- ja tiedonlouhintaan soveltuvan poikkeuksen soveltamista. Kyseisen kohdan mukaan kaupallista tekstin- ja tiedonlouhintaa voidaan nimittäin harjoittaa ainoastaan, jos oikeudenhaltijat eivät ole nimenomaisesti pidättäneet oikeuksiaan kappaleen valmistamiseen ja kopiointiin *asianmukaisella tavalla*.

Keskeiseksi nousee tulkinta siitä, mitä tarkoitetaan oikeuksien pidättämisellä asianmukaisella tavalla. Ensinnäkin tekijänoikeusdirektiivin 4 artiklan 3 kohta toteaa, että verkossa yleisön saataviin saatettu sisältö voidaan pidättää asianmukaisesti koneluettavilla keinoilla. Tätä on edelleen täsmennetty tekijänoikeusdirektiivin johdanto-osan perustelukappaleessa 18, jonka mukaan koneluettaviin keinoihin kuuluu metatiedot ja verkkosivuston tai palvelun ehdot.²⁷⁸

Oikeuksien asianmukaiselle pidättämiselle on myös asetettu vaatimuksia. Tekijänoikeusdirektiivin johdanto-osan perustelukappaleessa 18 todetaan, että oikeuksien pidättäminen tekstin- ja tiedonlouhintaa varten ei saisi vaikuttaa muuhun käyttöön.²⁷⁹ Jää kuitenkin epäselväksi se, kuinka tehokkaasti voidaan toteuttaa koneluettavia keinoja, jotka eivät rajoita yleisön saataviin saatetun sisällön muuta käyttöä. Lisäksi tulisi tarkentaa sitä, mitä muulla käytöllä tarkoitetaan.

Tekijänoikeusdirektiivin johdanto-osan perustelukappaleessa otetaan myös kantaa tilanteisiin, joissa sisältöä ei ole saatettu yleisön saataviin. Myös tällöin oikeudenhaltija voi pidättää asianmukaisesti oikeutensa, jotta oikeudenhaltija voi varmistaa yksinoikeutensa kunnioittamisen. Esimerkeiksi on esitetty sopimusjärjestelyitä sekä yksipuolista ilmoitusta.²⁸⁰ Tästä seuraa, että tilanteissa, joissa sisältöä ei ole saatettu yleisön saataviin, ei ole selvyttä siitä, miten tarkalleen määritellään ne kriteerit, joiden perusteella voidaan katsoa oikeuksien tulleen asianmukaisesti

²⁷⁸ Tekijänoikeusdirektiivi, resitaali 18.

²⁷⁹ Tekijänoikeusdirektiivi, resitaali 18.

²⁸⁰ Tekijänoikeusdirektiivi, resitaali 18.

pidätetyiksi. Tekstin- ja tiedonlouhintaa harjoittava toimija voi joutua tilanteeseen, jossa toimijan tulee itse arvioida, voiko sisällöstä valmistaa kopioita.

Avoin ilmaisu oikeuksien asianmukaisesta pidättämisestä ilman arviointikriteerejä voi johtaa myös siihen, että oikeudenhaltijoiden keinot pidättää oikeudet voivat olla kohtuuttomia tai oikeudenhaltijat voivat suosia tiettyjä toimijoita ilman hyväksyttävää perustetta. Kun otetaan huomioon, että ei-tieteellinen tekstin- ja tiedonlouhintateknologia edesauttaa myös tieteellisen tekstin- ja tiedonlouhintateknologian kehittymistä sekä ylipäättään tuottaa merkittävää arvoa tieteen ja yhteiskunnan kehittymiselle, tulisi teknologian rajoittamisperusteet olla rajattuja, tasapuolisia sekä läpinäkyviä.

Tekijänoikeusdirektiivin 4 artiklan tarkastelu osoittaa, että sekä termien ”asianmukaisella tavalla” että myös ”laillinen pääsy” sisällöt on jätetty avoimiksi ja tulkinnalle alttiiksi. Kuitenkin termin ”asianmukaisella tavalla” merkitysisältö ratkaisee sen voivatko esimerkiksi kaupalliset toimijat kopioida materiaalia, joka on lähtökohtaisesti suojattu Euroopan unionin oikeudessa tekijänoikeudella. Edellä esitetyn perusteella todetaan, että tekijänoikeusdirektiivissä esitetyt vaatimukset laillisesta pääsystä materiaaliin sekä asianmukaisesta oikeuksien pidättämisestä jättävät oikeustilan epäselväksi ja voivat johtaa kohtuuttomiin tekstin- ja tiedonlouhintaan kohdistuviin rajoituksiin. Lisäksi tulkinnalle avoimet vaatimukset voivat johtaa siihen, että jäsenvaltioissa omaksutaan erilaisia lähestymistapoja tekstin- ja tiedonlouhintaan.

On huomioitava, että tekijänoikeusdirektiivin yhtenä keskeisimpänä tavoitteena on ollut selvittää Euroopan unionissa vallinnutta oikeustilaa koskien tekijänoikeutta ja sen lähioikeuksia. Artikloiden tarkempi tarkastelu kuitenkin osoittaa, että ne sisältävät avoimia ja tulkinnanvaraisia ilmaisuja, jotka pitävät oikeustilan epävarmana ja mahdollistavat sen, että jäsenvaltioissa omaksutaan erilaisia lähestymistapoja tekstin- ja tiedonlouhintaan. Seuraavassa luvussa analysoidaan, miten voitaisiin sovittaa yhteen tekijänoikeuden taloudellinen ulottuvuus tekstin- ja tiedonlouhinnan kanssa tavalla, joka edistäisi kokonaisvaltaisesti yhteiskunnan kehittymistä.

6 KANNANOTTO MUUTOSTARPEESTA

6.1 Kannanoton rakenteesta

Edellä tässä tutkielmassa on systematisoitu ja tulkittu Euroopan unionissa voimassa olevan lain-säädännön ja oikeuskäytännön valossa, mikä on tällä hetkellä oikeustila hakukoneiden suorittaman kaupallisen tekstin- ja tiedonlouhinnan suhteen. On selvää, että tämän hetken kehitys luoda jatkuvasti uusia direktiivejä ja poikkeuksia kattamaan uusia teknologisia innovaatioita, pirstaloittaa juridista pelikenttää entisestään luoden oikeudellista epävarmuutta.²⁸¹ Lisäksi näyttää siltä, että tämän hetkinen oikeustila asettaa rajoitteita etenkin kaupalliselle tekstin- ja tiedonlouhinnalle. Tässä luvussa 6 vastataan *Law and Technology* -tutkimusmetodin neljänteen osa-alueeseen esittämällä kannanottoja siitä, miten hakukoneiden suorittamaa kaupallista tekstin- ja tiedonlouhintaa voitaisiin kehittää oikeudellisella tasolla Euroopan unionissa.

Kannanoton esittäminen jakautuu tässä luvussa kolmeen teemaan. Ensimmäiseen teemaan sisältyy yleisellä tasolla perehtyminen tekijänoikeuden ja tekstin- ja tiedonlouhinnan yhteentörmäyksessä huomioitaviin lähtökohtiin. Tämän jälkeen analyysin toinen teema käsittää arvioinnin siitä, miten tekijänoikeuden taloudellista ulottuvuutta voitaisiin kehittää digitaaliseen ympäristöön sopivaksi. Kolmannessa teemassa keskitytään siihen, miten markkinavaikutusanalyysia sekä erityisesti yhdysvaltalaisista fair use -doktriinia voitaisiin hyödyntää tekijänoikeuden ja kaupallisen tekstin- ja tiedonlouhinnan yhteensovittamisessa.

6.2 Kannanotolle asetettavat lähtökohdat

Ensinnäkin tutkittaessa tekijänoikeuden ja tekstin- ja tiedonlouhinnan yhteensovittamista on löydettävä oikeudenmukainen tasapaino²⁸². Tasapainon löytäminen tarkoittaa punnintaa, jossa

²⁸¹ *European Copyright Society* 2017, s. 4.

²⁸² Ratkaisuissa *Renckhoff* ja *Pelham* Euroopan unionin tuomioistuin on vahvistanut, että kansallisella tasolla on huomioitava digitaalisessa ympäristössä oikeudenmukainen tasapaino, jonka on yhtäältä säilytettävä Euroopan unionin perusoikeuskirjan 17 artiklan 2 kohdassa taattu tekijänoikeuden ja lähioikeuksien haltijoiden henkisen omaisuuden suoja. Mutta toisaalta on huomioitava aineiston käyttäjien intressit ja perusoikeudet sekä yleinen etu, esimerkiksi sananvapauden ja tiedonvälityksen vapauden muodossa. Lisäksi tuoreissa ratkaisuissa *Spiegel Online* ja *Funke Medien NRW* Euroopan unionin tuomioistuin on nimenomaisesti ottanut kantaa kappaleen valmistamista koskevan yksinoikeuden suhteesta siihen säädettävään poikkeukseen. Ratkaisuissa on edelleen vahvistettu näkemys siitä, että tulkinnan on oltava sopusoinnussa perusoikeuskirjassa taattujen perusoikeuksien kanssa. Ks. Asia C-161/17 *Renckhoff*, kohta 41; asia C-476/17 *Pelham*, kohta 32; asia C-516/17 *Spiegel Online*, kohta 59; sekä asia C-469/17 *Funke Medien NRW*, kohta 43. Nämä ratkaisut ovat keskittyneet tarkastelemaan tietoyhteiskuntadirektiivin 5 artiklan sisältämien vapaaehtoisten poikkeuksien soveltamistilanteisiin liittynyttä punnintaa eri oikeuksien ja intressien välillä. Koska hakukoneiden suorittaman kaupallisen tekstin- ja tiedonlouhinnan ja kappaleen valmistamista koskevan yksinoikeuden yhteensovittamisen taustalla ovat oikeudet ja intressit ovat samoja, voidaan näitä ratkaisuja soveltaa analogisesti tässä tutkielmassa esitettävään kannanottoon.

otetaan huomioon niin keskenään kilpailevia peruseriaatteita ja -oikeuksia kuin myös eri tahojen oikeudet ja intressit. Tärkeitä periaatteita ovat Euroopan unionin perusoikeuskirjan²⁸³ 11 artiklassa vahvistetut sananvapaus ja tiedonvälityksen vapaus, 13 artiklassa vahvistettu taiteen ja tutkimuksen vapaus, 14 artiklan mukainen oikeus koulutukseen sekä 17 artiklassa turvattavat teollis- ja tekijänoikeudet.²⁸⁴

Kohtuullista tasapainoa punnittaessa on yhtäältä tunnistettava hakukoneiden keskeinen asema takaamassa tiedon jatkuvaa virtaa internetissä sekä miten hakukoneet turvaavat yksilöiden pääsyn tietoon ympäri maailmaa. Hakukoneilla on ilmeinen merkitys sananvapaudelle, tiedonvälityksen vapaudelle, tieteen kehittymiselle sekä yleiselle edulle.

Toisaalta on tunnistettava myös oikeudenhaltijan oikeus saada yksin määrätä henkisen luomistyönsä hyödyntämisestä. Analysoinnissa on näin ollen huomioitava, että Euroopan unionissa on taattava mahdollisimman korkea tekijänoikeuden suojan taso.²⁸⁵ Korkea suojan taso vaatii sitä, että oikeudenhaltijoilla on niin moraalinen kuin taloudellinen oikeus henkiseen luomistyöhönsä eikä tätä voida ilman hyväksyttävää syytä rajoittaa.²⁸⁶

Oikeudenmukaisen tasapainon lisäksi analysoinnissa on pyrittävä lopputulokseen, joka edistää sisämarkkinoiden toimintaa. Kilpailukyky onkin keskeinen tekijä ja *Digitaalisten sisämarkkinoiden strategiassa Euroopalle* asetetaan merkittävä painoarvo markkinoiden kansainväliselle ulottuvuudelle. Euroopan unionin kilpailuaseman vahvistuminen ja entistä houkuttelevamman sijoittautumispaikan muodostuminen maailmanlaajuisesti toimiville yrityksille edellyttää, että Euroopan markkinoiden avoimuutta on ylläpidettävä ja kehitettävä edelleen digitaalisella.²⁸⁷ Näin ollen tulisi siis pyrkiä analysoinnissa myös ratkaisuun, joka huomioi kilpailulliset vaikutukset. Innovaatiotoimintaan tulisi kannustaa eikä asettaa kohtuuttomia rajoituksia uusille teknologioille, kuten tekstin- ja tiedonlouhinnalle.

Seuraavaksi luvuissa 6.3 ja 6.4 sovelletaan tässä luvussa esitettyjä lähtökohtia konkreettisemmallalla tasolla.

²⁸³ Euroopan unionin perusoikeuskirja 2010/C 326/02.

²⁸⁴ Ks. Tietoyhteiskuntadirektiivi, resitaali 31 sekä Euroopan unionin tuomioistuimen oikeuskäytännöstä ratkaisut asia C-467/08 *Padawan*, asia C-145/10 *Painer*, asia C-201/13 *Deckmyn*, asia C-469/17 *Funke Medien NRW*, asia C-476/17 *Pelham* sekä asia C-516/17 *Spiegel Online*.

²⁸⁵ Ks. Tietoyhteiskuntadirektiivi, resitaali 22 sekä tekijänoikeusdirektiivi, resitaalit 2, 3 ja 62.

²⁸⁶ Ratkaisussa *Infopaq I* todetaan, että korkean suojan tason tarkoituksena on taata oikeudenhaltija asianmukainen korvaus näiden teoksien käyttämisestä, jotta oikeudenhaltijat voivat jatkaa luovaa ja taiteellista työtään. Ks. lisää C-5/08 *Infopaq I*, kohta 40.

²⁸⁷ *Digitaalisten sisämarkkinoiden strategia Euroopalle*, s. 19.

6.3 Paluu tekijänoikeuden peruseriaatteille

Kuten edellä luvussa 3.2 on todettu, ovat hakukoneiden suorittamaan kaupalliseen tekstin- ja tiedonlouhintaan liittyvät oikeudelliset ongelmat osa laajempaa ilmiötä – tekijänoikeuden digitaalista kriisiä, jolle on ollut ominaisena piirteenä tekijänoikeuden peruseriaatteista irtaantuminen. Oikeudenhaltijalle kuuluvaa kappaleen valmistamista koskevaa suojaamekanismia on entisestään laajennettu uudessa tekijänoikeusdirektiivissä.

Tekijänoikeusdirektiivi perustuu nimittäin tulkinnalle siitä, että tekstin- ja tiedonlouhinnassa valmistettavat kopiot kuuluvat oikeudenhaltijan kappaleen valmistamista koskevan yksinoikeuden alaan. Näyttääkin siltä, että tekstin- ja tiedonlouhinnan ja tekijänoikeuden yhteensovittamisessa erkaannutaan entisestään idea-ilmaisu -jaottelusta, joka on yksi keskeisimmistä tekijänoikeuden peruseriaatteista. Idea-ilmaisu -jaottelu ei anna oikeudenhaltijalle oikeutta ideoihin, faktoihin, konsepteihin taikka mihinkään muihin teoksen ei-ilmaisua ilmaiseviin osiin. Idea-ilmaisu -jaottelun tarkoituksena on suojata alkuperäisen teoksen ilmaisevia elementtejä.²⁸⁸ Tutkielmassa esitetään, että hakukoneiden suorittaman tekstin- ja tiedonlouhinnan yhteydessä laadittava indeksi, joka perustuu indeksointirobottien valmistamiin kopioihin, ei ole oikeudenhaltijan yksinoikeuden loukkaus, sillä indeksit hyödyntävät suojaamattomia faktoja ja ideoita.

Lisäksi Euroopan unionissa vallitsevaa oikeustilaa kriittisesti tarkasteltaessa on huomioitava, että tekijänoikeuden ytimessä on teoksien kommunikatiivinen funktio. Kun teoksesta otetaan puhtaasti tekninen kopio osana teknologista prosessia²⁸⁹, teosta ei käytetä itse teoksena. Toisin sanoen kappaleen valmistamiseen ei sisälly samanlaista kommunikatiivista vaihetta kuin mikä alkuperäisen teoksen oikeudenhaltijan ja sen yleisön välillä on. Kommunikatiivisen vaiheen puuttuminen sekä se seikka, ettei kyseessä ole teoksen ilmaisevien elementtien käyttö, tukevat entisestään näkemystä siitä, ettei tekstin- ja tiedonlouhinnassa tapahdu sellaista kappaleen valmistamista, joka loukkaisi tekijänoikeutta. Kappaleen valmistamista koskeva oikeus olisikin

²⁸⁸ *Sag* 2019, s. 11 ja s. 59. Lisäksi *European Copyright Society* on esittänyt, että idea-ilmaisu -jaottelua tulisi selkeämmin soveltaa tekijänoikeuteen. Ks. *European Copyright Society* 2017, s. 5–6.

²⁸⁹ Myöskään itse hakukonetoiminnan lopputulos, jossa esitetään listauksena hakutuloksia ja linkkejä internetsivuille, ei hyödynnä alkuperäisen teoksen kommunikatiivista vaihetta. Sillä hakukonetulokset poikkeavat merkittävästi alkuperäisen teoksen ilmaisusta. *Sag* on perusteellisesti analysoinut sitä, miksi metadataa ei voida pitää teoksen alkuperäisen ilmaisun hyödyntämisenä. Katso lisää *Sag* 2019, s. 60.

muotoiltava uudelleen siten, että se ottaisi huomioon tekijänoikeuden kommunikatiivisen näkökulman ja rajoittaisi täten kappaleen valmistamista koskevan yksinoikeuden tilanteisiin, joissa teoksia käytetään teoksina ja myös ilmaisuvoimaisesti.²⁹⁰

Ratkaisuksi tekijänoikeuden peruseriaatteista erkaantumiseen esitetään tässä tutkielmassa, että myös Euroopan unionissa tulisi omaksua avoin tulkintanormi, joka yhdysvaltalaisen fair use -doktriinin tapaan nojautuu oikeudenmukaisuusarviointiin.²⁹¹ Myös eurooppalaisesta tekijänoikeuskeskustelusta löytyy tätä tukevia kannanottoja.

Esimerkiksi *Ohly* on kritisoinut eurooppalaista lähestymistapaa, jonka mukaan tekijänoikeuden yksinoikeuden piiriin kuuluvat toimet määritellään melko matalalla kynnyksellä tekijänoikeuden loukkauksiksi. Hänen mukaansa arvioinnin pitäisi perustua enemmän markkinaherkkään oikeudenmukaisuusarvioon, joka kykenee paremmin punnitsemaan eri oikeuksien ja etujen välillä.²⁹² Myös *Rognstad* ja *Poort* ovat esittäneet ratkaisumallia, jossa kolmannen oikeus hyödyntää teosta tulisi kytkeä kohtuullisuusarviointiin eikä kappaleen valmistamiseen.²⁹³

Lisäksi *Euroopan parlamentin sisäasioiden pääosaston politiikkayksikkö* on esittänyt näemyksen siitä, että eurooppalaiseen järjestelmään tulisi ottaa avoin tulkintanormi, joka soveltuisi tekstin- ja tiedonlouhintaan, mutta sitä voitaisiin soveltaa myös muihin innovatiivisiin teknologioihin. Avoimen tulkintanormin käyttöön ottaminen mahdollistaisi joustavan mekanismin,

²⁹⁰ *Strowel* 2018, s. 213. Myös *Drassinower* on tutkinut teknisten kopioiden kommunikatiivista funktiota. Hän on esittänyt, että koska teos on kommunikatiivinen teko, on myös sen uudelleentuottaminen myös uudelleenkommunikointia. Kopio on puolestaan tekijän puheen toisintamista. Tilanteissa, joissa teosta käytetään ikäänkuin vain ”musteena” ilman, että on käsillä uudelleenkommunikointia yleisölle, ei teosta käytetä myöskään teoksena. Ks. lisää aiheesta *Drassinower* 2015, s. 13 ja 87 sekä *Cocoru – Boehm* 2016, s. 7.

²⁹¹ Yhdysvalloissa tekijänoikeuteen tehtävien rajoitusten ja poikkeuksien tarkastelu perustuu fair use -doktriiniin, joka on avoin ja joustava rajoitussäännös. Yhdysvaltalaisen tekijänoikeuslain (the US Copyright Act) 107 artikla luo puitteet fair use -doktriinille. Kyseinen tekijänoikeuslain artikla määrittelee, minkä edellytyksien täytyessä oikeudenhaltijoiden yksinoikeutta voidaan rajoittaa. Yhdysvaltalaisen tekijänoikeuslain 107 artiklan mukaan tekijänoikeudella suojattujen teoksien oikeudenmukainen käyttö (”fair use”) kattaa muun muassa kopioiden ja äänitallenteiden käytön, jos tarkoituksena on esimerkiksi kritisointi, kommentointi, uutisointi, opetus, oppiminen tai tutkimus. Näin ollen kyseisiä tarkoituseriä varten toteutettua kappaleen valmistamista ei luokitella tekijänoikeuden loukkaukseksi. Lista ei kuitenkaan ole luonteeltaan tyhjentävä, sillä oikeudenmukaisuusarviointiin perustuva rajoitussäännös artikkelissa mahdollistaa myös muihin tarkoituseriin toteutetun toiminnan sallimisen. Oikeudenmukaisuusarvioinnissa huomioidaan seuraavat seikat 1) käytön tarkoitus ja luonne, onko käyttö esimerkiksi luonteeltaan kaupallista taikka liittyykö se voittoa tavoittelemattomaan koulutukseen; 2) kopioitavan teoksen luonne 3) missä määrin ja kuinka olennaisilta osin tekijänoikeudella suojattua teosta on käytetty tarkasteltaessa kokonaisuutena; sekä 4) mikä vaikutus käytöllä on tekijänoikeudella suojatun teoksen potentiaalsiin markkinoihin tai kaupalliseen arvoon. Varsinkin kohdat 1) ja 4) tuovat oikeudenmukaisuusarviointiin mukaan taloudellista perspektiiviä. Kyseinen oikeudenmukaisuusarvio voi oikeuttaa yksinoikeudella suojatun toiminnan tilanteissa, joissa se estää markkinoiden vääristymisen. Ks. lisää esimerkiksi *Ohly* 2018, s. 19 sekä yleisesti aiheesta *Sag* 2019.

²⁹² *Ohly* 2018, s. 2.

²⁹³ *Rognstad – Poort* 2018, s. 121–160.

jonka avulla uusia teknologisia innovaatioita voidaan käyttää Euroopan unionissa, vaikeivät nämä täyttäisi olemassa olevia rajoituksia ja poikkeuksia. Sen sijaan arvioitaisiin kyseisen innovaation merkitystä yleiselle edulle sekä yhteensovittamisen taustalla vaikuttavia perusoikeuksia, kuten sananvapautta sekä oikeutta tietoon.²⁹⁴

Avoimen tulkintanormin käyttöönottoa voidaan myös perustella sillä, että tulkintanormi parantaisi Euroopan unionin kilpailuasemaa esimerkiksi tekstin- ja tiedonlouhintateknologiassa ja koneoppimisessa.²⁹⁵ Yhdysvalloissa, Kanadassa ja Israelissa on jo käytössä avoin tulkintanormi, joka mahdollistaa kaupallisen tekstin- ja tiedonlouhinnan. Tämän lisäksi myös Kiinassa, Japanissa, Australiassa ja Singaporessa on lainsäädäntöä muokattu vastaamaan paremmin kaupallisen tekstin- ja tiedonlouhinnan tarpeita.²⁹⁶

Avoimeen tulkintanormiin on myös yhdistetty riskejä. *Euroopan komission tutkimuksen ja innovoinnin pääosasto* on esittänyt, että avoin tulkintanormi, joka perustuu ex post -näkökulmaan, voisi heikentää tietyssä tilanteissa oikeudellista ennustettavuutta.²⁹⁷ Avoimella tulkintanormilla voisi olla myös negatiivisia vaikutuksia transaktiokustannuksiin ja avoin tulkintanormi voisi kohtuuttomasti suojella taloudellisesti vahvoja toimijoita.²⁹⁸

Avoimen tulkintanormin lisäksi digitalisoitunut tekijänoikeus on mahdollista kytkeä paremmin sen peruseräiteisiin ottamalla käyttöön tulkinta, jossa analysoidaan hyödynnetäänkö teosta teoksena. Tämä arviointimekanismi on Euroopan unionissa käytössä jo esimerkiksi tavaramerkkiloukkauksissa.²⁹⁹ Lisäksi ranskalainen tuomioistuin on analysoinut kappaleen valmist-

²⁹⁴ *Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: oikeudellinen analyysi* 2018, s. 25. Myös *Euroopan komission tutkimuksen ja innovoinnin pääosasto* on kannattanut avointa tulkintanormia ja perustellut kantaansa sillä, että tuomarin suorittama ex post -arviointi sopii joustavuutensa ansiosta paremmin koko ajan kehittyvän teknologisen ympäristön tarpeisiin. Ks. *Euroopan komission tutkimuksen ja innovoinnin pääosaston asiantuntijalausunto* 2014, s. 52–53.

²⁹⁵ Esimerkiksi *Sag* ja *The European Alliance* ovat esittäneet perustellun huolensa siitä, että Eurooppa on jäämässä jälkeen koneoppimisen ja tekoälyn kehittämisessä. Ks. *Sag* 2019, s. 43 sekä *European Alliance for Research Excellence's statement* 2018.

²⁹⁶ *European Alliance for Research Excellence I, European Alliance for Research Excellence II* 2018 sekä *European Alliance for Research Excellence's statement* 2018.

²⁹⁷ *Euroopan komission tutkimuksen ja innovoinnin pääosaston asiantuntijalausunto* 2014, s. 52–53.

²⁹⁸ *Euroopan parlamentti, sisäasioiden pääosasto, politiikkayksikkö – Kansalaisten oikeudet ja perustuslakiasiat: oikeudellinen analyysi* 2018, s. 16 sekä *Hugenholtz* 2017, s. 257–291.

²⁹⁹ *Strowel* on kirjoittanut aiheesta kattavasti. Ks. lisää *Strowel* 2018, s. 207 ja s. 219–226. Ks. myös yhdistetyt asiat C-236/08 ja C-238/08 *Google France ja Google*, kohdat 55,56 ja 75, joissa Euroopan unionin tuomioistuin on arvioinut sitä, mitä vaaditaan, että tavaramerkkiä oltaisiin käytetty tavaramerkin mukaisella tavalla taloudellisessa merkityksessä.

mista koskevaa yksinoikeutta siitä näkökulmasta, onko teosta hyödynnetty teoksena. Ranskalainen tuomioistuin on katsonut, että tekijänoikeudella suojattuja uutisartikkeleita on voitu kopioida ja indeksoida ilman, että kyseessä olisi luvaton kappaleen valmistaminen tekijänoikeudella suojatusta teoksesta.³⁰⁰ Ratkaisussa omaksuttua tulkintaa on perusteltu sillä, että toiminta on käsittänyt puhtaasti artikkeleiden indeksoimisen. Artikkeleita ei siis tällöin oltu luettu, kopioitu taikka käytetty teoksena, vaan niistä oltiin ainoastaan louhittu informaatiota.³⁰¹ Myös hakukoneiden toiminnassa on kyse siitä, ettei teoksia lueta, kopioida taikka käytetä teoksina, vaan kyseisiä teoksia käytetään informaation louhintaa varten.

Edellä esitettyyn perustuen tutkielmassa esitetään siirtymää pois teknologianeutraalista kappaleen valmistamista koskevasta määritelmästä avoimen tulkintanormin soveltamiseen tuomioistuimissa. Lisäksi tutkielmassa esitetään *De lege sententia* -suositus siitä, että Euroopan tuomioistumessa olisi lisättävä ymmärrystä osana teknologista prosessia otettavista kopioista, jotta voidaan arvioida, hyödyntävätkö kopiot tosiasiasa alkuperäisen teoksen omaperäistä ilmaisua. Tämän toteutumiseksi tuomioistuimissa tulisi hyödyntää enemmän teknologiasektorin asiantuntijoita ja asiantuntijalausuntoja.³⁰² Näin voitaisiin paremmin turvata se, ettei innovaatiotoimintaa rajoitettaisi kohtuuttomasti Euroopan unionissa.

Edellä luvussa 4.2.4 tutkittaessa tietoyhteiskuntadirektiivin 5 artiklan 1 kohtaa on tutkielmassa esitetty, että kappaleen valmistamisen taloudellisen merkityksen arviointi on ollut puutteellista Euroopan unionin oikeuskäytännössä. Seuraavassa luvussa keskitytään niihin taloudellisiin seikkoihin, joihin tulisi kiinnittää huomioida analysoitaessa tekijänoikeuden ja kaupallisen tekstin- ja tiedonlouhinnan välistä jännitettä.

6.4 Todellinen vaikutus markkinoihin

Euroopan unionin tuomioistuimen oikeuskäytännön tarkastelu luvussa 4.2.4 antaa merkkejä siitä, että Euroopan unionissa ei ole tällä hetkellä riittäviä työkaluja hahmottamaan tekijänoikeuden ja tekstin- ja tiedonlouhinnan välistä jännitettä taloudellisesta näkökulmasta. *Ohly* on

³⁰⁰ Ranskan kassaatioistuimen (Court de cassation) ratkaisu 86–11918 vuodelta 1987. Ks. myös *Strowel* 2018, s. 223–224.

³⁰¹ Ranskan kassaatioistuimen (Court de cassation) ratkaisu 86–11918 vuodelta 1987. Ks. myös *Strowel* 2018, s. 223–224.

³⁰² *Strowelin* mukaan varsinkin taloudellisia seikkoja arvioitaessa tulisi hyödyntää enemmän asiantuntijoita. Ks. lisää *Strowel* 2018, s. 235.

nostanut esille, että eurooppalainen järjestelmä ei kykene tunnistamaan yksittäisten tekojen taloudellisia vaikutuksia, kun arvioidaan tekijänoikeuden rajoittamista. Muodollisesta teknologianeutraalista määrittelystä tulisi päästä eroon ja tulisikin omaksua järjestelmä, joka mahdollistaisi sen, että tekijänoikeussuoja vastaisi paremmin markkinatodellisuutta.³⁰³ Tästä johtuen tutkielmassa haetaan argumentatiivista tukea yhdysvaltalaisesta fair use -doktriinista, joka on oikeudenmukaisuusarvioinnin lisäksi taloudellisuusarviointiin nojautuva punnintanormi.

Fair use -doktriini korostaa, että teoksen hyödyntäminen kaupallisessa tarkoituksessa on vain yksi huomioon otettava tekijä, kun arvioidaan, voidaanko teoksen hyödyntäminen oikeuttaa oikeudenmukaisena käyttönä. Näin ollen tilanteessa, jossa on tunnistettavissa kaupallisia vaikuttimia, ei tule suoraan tehdä johtopäätöstä siitä, että teoksen hyödyntäminen olisi tekijänoikeuden loukkaus.³⁰⁴ Tunnistettaessa, että teoksen hyödyntämisen ja tekijänoikeuden rajoittamisen yhteensovittamisen taustalla saattaa olla kaupallisia vaikuttimia nojaututaan markkinavaikutusarviointiin.

Yhdysvaltalainen markkinavaikutusarviointi keskittyy siihen, vahingoittaako, ja missä määrin, lisensoimaton käyttö alkuperäisen teoksen nykyisiä taikka tulevia markkinoita. Markkinavaikutuksia tutkittaessa nykyisillä markkinoilla kiinnitetään huomiota muun muassa siihen, syrjäyttääkö teoksen hyödyntäminen alkuperäisen teoksen markkinoilta, aiheutuuko oikeudenhaltijalle riski tulojen menettämisestä tai aiheutuuko kopion leviämisestä muutoin huomattavaa vahinkoa oikeudenhaltijalle.³⁰⁵ Oikeudenmukaisuusarviointi kykenee huomioimaan kokonaisuuden ja reagoimaan teknologisen ympäristön tarpeisiin.

Yhdysvaltalaisen markkinavaikutusanalyysin vertaaminen Euroopan unionin tekijänoikeusjärjestelmään on mielenkiintoista. Esimerkiksi *Infopaq I* -ratkaisussa ei lainkaan oltu otettu huomioon sitä, mikä vaikutus tekstiotteilla olisi ollut alkuperäisen teoksen markkinoille. Ratkaisussa ei ole myöskään tarkasteltu sitä, olisivatko tekstiotteet voineet muodostaa alkuperäisestä teoksesta korvikkeen samoille markkinoille.

Lisäksi kuten tutkielmassa on aikaisemmin mainittu luvussa 4.2.4, tulisi hakukoneiden tekstin- ja tiedonlouhinnan yhteydessä valmistettavien kopioiden analysoinnissa huomioida myös se,

³⁰³ *Ohly* 2018, s. 2.

³⁰⁴ *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, kohdat 584–585 (1994).

³⁰⁵ *Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, kohta 566 (1985) sekä *Authors Guild v. Google, Inc.*, 804 F.3d 202, kohta 219 (2d Cir. 2015). Ks myös *Sag* 2019, s. 59 ja *The United States Copyright Office* 2019.

mihin teknologian arvo perustuu. Teknologian arvon määrittämiseen voidaan hyödyntää markkinavaikutusten arvioimista. Yhdysvaltalaisen oikeuskäytännön markkina-analyysissä esitetyt tulkinnat ja kannanotot huomioiden voidaan tulkita, että hakukoneiden suorittaman kaupallisen tekstin- ja tiedonlouhinnan puitteissa suoritettavat kopiot eivät muodosta korvikkeita alkuperäisestä teoksesta sen markkinoille. Hakukoneiden tuottama tieto internetsivuista hakutuloksien muodossa ei nimittäin tuota korviketta sille tiedolle, joka on internetsivuilla, eikä myöskään alkuperäiselle ilmaisulle teoksessa.³⁰⁶

Markkinavaikutuksen huomioon ottaminen on relevanttia, sillä arvioitaessa voidaanko teknologista innovaatiota ja tieteen kehittymistä rajoittaa tekijänoikeuden kustannuksella olisi keskeistä myös arvioida, mitkä ovat tekijänoikeuden rajoittamisen todelliset taloudelliset vaikutukset. Markkinavaikutusarviointi tukisi myös kappaleen valmistamista koskevan yksinoikeuden palauttamista takaisin tekijänoikeuden peruseriaatteille.

³⁰⁶ Ainoa tapa esittää, että tekstin- ja tiedonlouhinta-teknologialla olisi markkinavaikutus edellyttäisi täysin ideaalimaisuus -jaottelun sivuuttamista. Ks. lisää *Sag* 2019, s. 31.

7 YHTEENVETO

Tämän tutkielman tarkoituksena on ollut selvittää, mikä on Euroopan unionissa tämän hetkinen oikeustila hakukoneiden suorittaman kaupallisen tekstin- ja tiedonlouhinnan suhteen. Tähän tutkimusongelmaan vastaaminen on rakentunut erityisesti tutkielman lukuihin 2–5, joissa on perusteellisesti tutkittu, miten hakukoneet toimivat ja hyödyntävät tekstin- ja tiedonlouhinta-tekniologiaa, miten tekstin- ja tiedonlouhinta linkittyy tekijänoikeuteen, sekä millaisia rajoituksia voimassa oleva oikeus mahdollistaa tekijänoikeuteen tehtävän. Analyysin perusteella voidaan esittää yhteenvetona seuraavaa.

Tietotekniikan avulla hakukonetoiminta voidaan määritellä informaation, eli tiettyjen arvojen, etsinnäksi avaimien avulla laajasta indeksistä. Indeksi on puolestaan laaja tietokanta, johon on tiivistetty valtava määrä dataa internetistä. Indeksien rakentaminen perustuu tekstin- ja tiedonlouhinta-tekniologiaan, joka on pitkälle automatisoitua koneoppivien algoritmien johdosta.

Indeksointiprosessi alkaa siitä, kun algoritmeihin pohjautuvat indeksointirobotit lukevat internetiä. Indeksointirobottien vierailu internetsivuilla linkittyy heti tekstin- ja tiedonlouhinnan ensimmäiseen vaiheeseen, eli materiaaliin pääsyyn. Indeksointirobottien saadessa pääsyn internetsivulle siirrytään tekstin- ja tiedonlouhinnan toiseen vaiheeseen, eli kopiointiin. Tässä vaiheessa hakukoneen indeksointirobotit valmistavat kopioita löytämiensä internetsivujen HTML-koodeista. Kopioinnista seuraava vaihe käsittää tekstin- ja tiedonlouhinnan kolmannen vaiheen, eli louhinnan. Hakukoneiden kohdalla louhinta ilmenee siinä, että kopioiduista HTML-koodeista kerätään analysointityökaluilla dataa tiivistetyssä muodossa indeksiin.

Kuvattu indeksointiprosessi linkittyy olennaisesti tekijänoikeuteen. Indeksointia varten otettavat kopiot internetsivujen HTML-koodeista saattavat nimittäin sisältää tekijänoikeudella suojattua materiaalia, johon tämän hetken oikeuskäsityksen mukaan voidaan mahdollisesti ulottaa tekijänoikeuden taloudelliseen ytimeen kuuluva kappaleen valmistamista koskeva yksinoikeus. Jännite tekijänoikeuden ja tekstin- ja tiedonlouhinta-tekniologian välillä perustuu siihen, että oikeudenhaltijalla on lähtökohtaisesti yksinoikeus valmistaa kopioita teoksestaan. Näin ollen muut tahot tarvitsevat suostumuksen kopioiden valmistamiseen.

Oikeustilan laadullisesta analysoinnista *ennen* tekijänoikeusdirektiiviä kuitenkin ilmeni, ettei ole täyttä varmuutta sen suhteen, lukeutuvatko tekstin- ja tiedonlouhinta-tekniologian yhteydessä

otettavat kopiot yksinoikeuden alle eikä myöskään siitä, milloin tätä yksinoikeutta voidaan rajoittaa. Haasteet ilmenevät neljällä eri tasolla. *Ensimmäinen* haaste linkittyy tietoyhteiskuntadirektiivin 2 artiklaan, jossa kappaleen valmistamista koskeva yksinoikeus on sidottu teknologianeutraaliin määritelmään. Teknologianeutraalin määritelmän johdosta muodostuu vahva oletus siitä, että kappaleen valmistamiseksi määritellään myös osana teknologista prosessia valmistettava tekninen ja näkymätön kopio, jolla ei ole yhteyttä alkuperäisen teoksen ilmaisuun.

Teknologianeutraali määritelmä on myös olennaisesti yhteydessä *toiseen* tekstin- ja tiedonlouhinnan sekä tekijänoikeuden yhteensovittamisessa olevaan haasteeseen, nimittäin tekijänoikeuden digitaaliseen kriisiin. Lainsäädäntömalleja ja suojamekanismeja, jotka ovat suunniteltuja analogiseen maailmaan, sovelletaan digitaalisessa ympäristössä oikeudenhaltijan intressejä korostaen. Seurauksena on ollut tekijänoikeuden suhteen heikkeneminen reaali maailmaan. Tekijänoikeus ei suojaa enää teoksia kappaleen valmistamiselta, jonka tarkoituksena on hyödyntää teoksen alkuperäistä ilmaisua, vaan suoja on ulotettu koskemaan tilanteita, joita ei voida enää tekijänoikeuden peruseriaatteilla, kuten tekijänoikeuden idea-ilmaisu -jaottelulla taikka kommunikatiivisella funktiolla, perustella.

Kolmas haaste liittyy siihen, että tulkittaessa tekstin- ja tiedonlouhinnan käsittävän toimia, jotka luetaan kappaleen valmistamiseksi, tulee tekstin- ja tiedonlouhintaa harjottavien toimijoiden saada oikeudenhaltijalta etukäteinen suostumus kopiointiin. Tämän vuoksi oikeudenhaltijat voivat merkittäväällä tavalla vaikuttaa tekstin- ja tiedonlouhinnan ensimmäiseen vaiheeseen, eli materiaaliin pääsyyn. Oikeudenhaltijat voivat rajoittaa pääsyä erilaisilla lisensointijärjestelyillä ja TPM-rajoituksilla.

Viimeinen, eli *neljäs*, haaste on yhteydessä tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan mukaiseen pakolliseen tekijänoikeuden rajoitusnormiin ja sen soveltamiseen liittyviin vaatimuksiin. Yksityiskohtainen tarkastelu tutkielmassa on osoittanut, että artiklan esittämät vaatimukset ovat tulkinnalle alttiita, eikä suoraa päätelmää rajoitusnormin soveltumisesta hakukoneiden suorittamaan kaupalliseen tekstin- ja tiedonlouhintaan voida tehdä. Erityisesti vaatimus, joka koskee taloudellista itsenäisyyttä, tuottaa tulkintaongelmia. Oikeuskäytännössä *Infopaq I* -ratkaisussa on määritelty, ettei tilapäisen kappaleen valmistaminen saa mahdollistaa sellaisen lisäedun saamista, joka ylittää teoksen laillisesta käytöstä saatavan edun. Lisäedun määrittäminen on kuitenkin ongelmallista, sillä se edellyttäisi, että tekstin- ja tiedonlouhinta teknologialle tulisi voida määrittää jokin tietty arvo, sekä lisäksi tästä arvosta tulisi voida erottaa, mikä mahdollinen taloudellinen lisäetu kopiolla on ollut.

Edellä esitetyt haasteet ilmentävät, ettei tietoyhteiskuntadirektiivin aikana ollut riittävää oikeudellista varmuutta siitä, miten esimerkiksi hakukonetoiminta asettuu tekijänoikeudellisella rajapinnalla. Oikeustilan tarkastelu ennen tekijänoikeusdirektiiviä antoi ymmärryksen siitä, mihin tekijänoikeusdirektiivin säätämistarve on perustunut ja mahdollisti tässä tutkielmassa laadullisen analyysin tuottamisen siitä, selkeyttääkö tekijänoikeusdirektiivi epävarmuuden varjostamaa oikeustilaa. Kuitenkin tekijänoikeusdirektiivin 3 ja 4 artikloiden tarkastelu osoittaa, ettei tekijänoikeusdirektiivi onnistunut riittävällä tasolla poistamaan epäselvyyttä, joka on liittynyt etenkin lisensointijärjestelmiin ja TPM-rajoituksiin. Voidaan esittää, että tekijänoikeusdirektiivillä voi olla jopa päinvastainen vaikutus, sillä se sisältää tulkinnalle alttiita avoimia ilmauksia, kuten laillinen pääsy ja asianmukainen pidättäminen.

Tekijänoikeusdirektiivin tekemä jaottelu tekstin- ja tiedonlouhinnan tieteellisten ja muiden tarkoituserien välillä on kritiikille altis. *Oikeuden lukea* tulisi tarkoittaa myös *oikeutta louhia*. Jaottelu aiheuttaa tarpeettomia rajoituksia muita tarkoituseriä varten harjoitetulle tekstin- ja tiedonlouhinnalle. Tämä on harmillista, sillä esimerkiksi kaupallisella sektorilla harjoitettavalla tekstin- ja tiedonlouhinnalla on huomattavia vaikutuksia tekstin- ja tiedonlouhintechnologian kehittymiseen. Lisäksi kaupallisella sektorilla tehdään tieteellistä tutkimusta, sekä monilla kaupallisiin tarkoituseriin noujatuilla tekstin- ja tiedonlouhintechnologioilla on suoraan yhteiskuntaa edistävä vaikutus, esimerkiksi juuri hakukoneilla. Tutkielmassa kyseenalaistetaan ylipäätään tekijänoikeusdirektiivissä omaksuttu lähtökohta, jonka mukaan tekstin- ja tiedonlouhintechnologian yhteydessä tuotettavat kopiot kattaisivat kappaleen valmistamista koskevan määritelmän.

Tutkielmassa esitetään yhteenvedona tutkimusongelmaan, että hakukoneiden suorittamaan kaupalliseen tekstin- ja tiedonlouhintaa koskevaan oikeustilaan liittyy oikeudellista epävarmuutta, joka voi pahimmassa tapauksessa rajoittaa yhteiskunnassa pääsyä informaatioon sekä jarruttaa innovaatiota ja yhteiskunnan kehitystä.

Tämän vuoksi tutkielman luvussa 6 on esitetty kolmiosainen kannanotto kaupallisen tekstin- ja tiedonlouhinnan oikeudellisesta muutostarpeesta. *Ensinnäkin* muutostarpeen kehittämiseksi on huomioitava perusteellisesti hakukoneiden hyödyntämisen analysointiteknologian ja tekijänoikeuden välisen jännitteen taustalla vaikuttavat perusoikeudet ja intressit.

Immateriaalioikeudet ovat kasvaneet yhteiskunnan muutosten ja keksintöjen myötä olennaiseksi osaksi yhteiskuntaamme. Se, millaisen juridisen konstruktion ympärille immateriaalioikeudet rakentuvat, vaikuttaa niin markkinoihin, kilpailuun, politiikkaan, kulttuurin ja tieteen kehitykseen, sekä ylipäätään siihen millainen yhteiskuntamme tulevaisuudessa on. Juridisten konstruktioiden tulisi mahdollistaa digitalisoituneessa yhteiskunnassa mitä moninaisimmilla tasolla vapaus – vapaus sananvapauteen ja tiedonvälitykseen, vapaus taiteeseen ja tutkimukseen, vapaus koulutukseen, vapaus valitsemaansa elinkeinon sekä vapaus omistaa. Nämä tulisi sovittaa yhteen tavalla, joka veisi yhteiskuntaa parhaiten eteenpäin. Tämä edellyttää oikeudenmukaisen tasapainon löytämistä, mikä samalla myös vahvistaisi Euroopan unionin asemaa digitaalisessa kilpailussa.

Toinen osa kannanotosta käsittää tulkinnan siitä, että tekijänoikeuden laaja ja teknologia-neutraali suojafunktio on ristiriidassa tekijänoikeuden peruseriaatteiden kanssa. Voidaan todeta, että kappaleen valmistamista koskevan yksinoikeuden ulottaminen hakukoneiden suorittamaan tekstin- ja tiedonlouhintaa merkitsisi erkaantumista sekä tekijänoikeuden kommunikaatiivisesta funktiosta sekä idea-ilmaisu -jaottelusta. Taloudellisen oikeuden tulkinta tulisi palauttaa sen peruseriaatteille ja lähtökohtana tulisi olla joustava tulkintanormi, jossa tarkastelee sitä, hyödynnetäänkö tekijänoikeudella suojattua teosta teoksena, eli hyödynnetäänkö sen omaperäistä ilmaisua kaupallisessa mielessä. Hakukoneiden käyttämä tekstin- ja tiedonlouhintateknologia perustuu kopioiden ottamiseen internetsivujen HTML-koodeista ja kerätyn datan indeksoimiseen. Hakukonetoiminnan tarkoituksena on välittää hakukonetta käyttäville hakutuloksia, eli informaatiota – eikä hyödyntää alkuperäisen teoksen omaperäistä esittämistapaa kaupallisessa mielessä.

Lisäksi kannanotossa on esitetty *de lege sententia* -suositus siitä, että Euroopan unionin tuomioistuimen ratkaisukäytännössä tulisi tukeutua enemmän asiantuntijoihin ja asiantuntijoiden lausuntoihin.

Kolmas kannanoton osa-alue puolestaan korostaa sitä, että arvioitaessa tekijänoikeuden ja uusien teknologioiden yhteensovittamista tulisi arvioinnin perustua myös markkinavaikutusanalyysiin, jossa tutkittaisiin sitä, vahingoittaako ja missä määrin kopiointi alkuperäisen teoksen nykyisiä tai tulevia markkinoita, muodostamalla esimerkiksi korvikkeen alkuperäiselle teokselle. Keskeisenä osana markkinavaikutusanalyysia on tunnistaa, mihin kopioita hyödyntävän

teknologian arvo perustuu. Hakukoneiden kohdalla on kuitenkin todettava, ettei siinä käytettävän tekstin- ja tiedonlouhintateknologian arvo perustu alkuperäisen ilmaisutavan hyödyntämiselle, vaan ainoastaan analysoitavassa materiaalissa olevalle informaatiolle.

Tutkielman johdannossa on esitetty useita näkökulmia, joiden tarkastelu on jouduttu jättämään tämän tutkielman ulkopuolelle. Kuitenkin myös näillä rajauksilla on keskeisesti vaikutusta tutkimusongelmaan sekä sen taustalla oleviin rakenteisiin ja ilmiöihin. Tarve jatkotutkimukselle on ilmeinen. Kun kiinnitetään huomio varsinkin tekstin- ja tiedonlouhintaan kätkeytyvään potentiaaliin ja nopeuteen, jolla teknologia kehittyy ympäri maailmaa, voivat uudet oikeudelliset haasteet olla yhden klikkauksen päässä.

Punnittaessa eri vaihtoehtoja tekijänoikeuden ja kaupallisen tekstin- ja yhteensovittamiselle on muistettava,

“Any restriction that unreasonably delays innovation undermines the very goal of intellectual property – to bring society forward.”³⁰⁷

³⁰⁷ Ks. lainaus Husovec 2017, s. 138.