Complex nr. 1/2008

Ingvild Jørgensen

INTERNET SEARCH ENGINES' COLLECTING AND PROCESSING OF WEB PAGES – FROM THE PERSPECTIVE OF COPYRIGHT LAW

Senter for rettsinformatikk/Avdeling for forvaltningsinformatikk
Postboks 6706 St Olavs plass
0130 Oslo

Henvendelser om denne bok kan gjøres til:

Senter for rettsinformatikk
Postboks 6706 St. Olavs plass
0130 Oslo
Tlf. 22 85 01 01
www. jus.uio.no/iri/

ISBN 978-82-7226-108-4
ISSN 0806-1912

Utgitt i samarbeid med Unipub AS
Denne boken går inn i universitets- og høyskolerådets skriftserie
Trykk: e-dit AiT AS
Omslagsdesign Kitty Ensby

Senter for rettsinformatikks utgivelser i skriftserien Complex er støttet av:
Advokatfirmaet Selmer DA
Wikborg Rein & Co
Lovdata

PREFACE

Early in my studies I became certain that I wanted to write my master thesis on subjects concerning information technology, and my interest was eventually focused on topics relating to the Internet and copyright law. The final topic became clear after a conversation with my supervisor dr. juris Jon Bing, determining topics that were of great interest to me as well as having topicality. The field of search engines is interesting in every respect, most of all because of the massive importance they have for the Internet community, and the challenges legislators face when trying to stay updated in a rapidly progressing technology. Early on I found there existed little written material concerning this, and the topic proved to be a massive and challenging one. I tried to include the most relevant and important problems, and concentrated on being thorough in my analysis rather than discussing as many problems as possible.

Understanding exactly *how* search engines work proved to be a difficult task. A lot of time had to be spent acquiring this knowledge. During these studies I had good help from my father Hans Petter Jørgensen, who works with computer systems, and my two brothers Eivind and Vegard Jørgensen, who hold master degrees in communication technology and computer science respectively. They kindly took of their time and were available for me to discuss the more general aspects of the technology.

Special thanks go to my supervisor Professor Bing for helping me through the process of writing this thesis, by offering helpful advice and valuable input. I am very grateful for his support, and for the opportunity to work with the thesis at his workplace, the *Norwegian Research Center for Computers and Law,* for the periods of time that was necessary. I had many fruitful conversations with the people that work and study there, as they were very helpful and supportive.

I also wish to thank my fellow students for being available for discussions, and especially Maria Korneliussen and Vegard Emaus for reading drafts and offering their thoughts on the subject. Furthermore, I had good help from my friend Joy-Loi Chepkoech who, despite finding the subject of law exceedingly boring, proof-read the thesis at the end.

Finally I would like to thank my cohabitant Frode Stock for being supportive and understanding while I spent several months devoting most of my attention to this thesis.

Tromsø, 24.06.2007
Ingvild Jørgensen

CONTENTS

I – PRELIMINARY MATERIAL

ABBREVIATIONS

Literature:

- (Sterling, 2003): J.A.L Sterling, *World Copyright Law*, London, 2nd edition, 2003 (Reprinted 2005).
- (Fielden, et al., 2002): Ned L. Fielden and Lucy Kuntz, *Search Engines Handbook*, Jefferson (North Carolina), 2002.
- (Andersen, 2005): Mads Bryde Andersen, *IT-retten*, København, 2nd edition, 2005.
- (Rognstad, 2004): Ole Andreas Rognstad, *Fragmenter til en lærebok i opphavsrett*, Oslo, 2004.
- (Koktvedgaard, et al., 2005): Mogens Koktvedgaard, *Lærebog i Immaterialret*, 7th edition by Jens Schovsbo, København, 2005.
- (Tanenbaum, 2003): Andrew S. Tanenbaum, *Computer Networks,* Amsterdam, 4th edition, 2003.

Legislation:

- [The Norwegian Copyright Act]: Lov 12th of May 1961 no. 2 *om opphavsrett til åndsverk m.v.*
- [The Norwegian E-commerce Act]: Lov 23rd of May 2003 no. 35 *om visse sider av elektronisk handel og andre informasjonssamfunnstjenester.*
- [The Copyright-directive]: EC-directive: Directive 2001/29/EC *on the harmonisation of certain aspects of copyright and related rights in the information society.*
- [The E-commerce Directive]: EC-directive: Directive 2000/31/EC *on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce).*

II – THESIS

1 PREAMBLE

1.1 Presentation of the subject

This thesis will consider the challenges faced by copyright law in relation to Internet based search engines and their "collecting" of web pages. Examples of such search engines are Google[1], Yahoo![2] and Sesam[3]. Most modern search engines are built on the same basic technical fundaments of operation, like for example the index, which I will explain later in chapter 2.2. The distinguishing techniques are mostly "of a higher level" than the index, and these will not be focused on in this thesis. However, apart from the index I will also focus on the "cache", which is a function where a web page is copied and presented in its entirety. The cache is offered by for instance Google, Yahoo! and Windows Live Search[4], but not by other search engines like Sesam or AltaVista[5].

There is a lot of focus in the media on modern technology and copyright breaches, namely the piracy of music, movies and games. The generic term for this is "file sharing", though it is entirely possible to conduct file sharing without committing piracy; it all depends on the origin of the file. The question used to be what is lawful and what is not. Now it has evolved into being a question of which methods the rightholders, such as music companies or movie corporations, can apply in order to expose people who shares files illegally. They are not only targeting the people sharing and downloading the files, but also the people that do nothing more but to show others where they might find such files, cf. "The Napster decision"[6]. However, little attention is paid to the search engines' mode of operation, perhaps because the search engines are vital to the Internet as we know it today. Nonetheless, some take notice, but that is for the most part not to the basic operation, but more to the specialized services that for instance Google provides. Google News[7] has been sued by news corporations in Belgium for infringing on their rights; stories from the newspapers were presented as news ingresses on Google News, with

1 http://www.google.com [05.03.07].
2 http://www.yahoo.com [05.03.07].
3 http://www.sesam.no [05.03.07].
4 http://www.live.com/?searchonly=true&mkt=en-US [10.04.07].
5 http://www.altavista.com [05.03.07].
6 Rt. 2005 p. 41.
7 http://news.google.com/ [05.03.07].

a link to the complete story on the newspapers own site. Google lost the case[8], but will appeal. [9]

This is so far the most important case against Google[10], but it is not targeted at Google's main function; the search engine itself. The question of whether or not this is legal activity still stands unanswered. Moreover, even if the search engines' activities are found to be an infringement to copyright legislation, they might not be liable because of certain rules in the electronic commerce legislation. These are the question this thesis will try to answer.

I have decided to write in English because the field of copyright and e-commerce law, as well as the technology in question, is highly International and English is the working language of most who deal with these issues.

This means that I will have to translate the Norwegian legal texts to English, which might be perceived as unfortunate. However, if I were to write in Norwegian, then I would have to translate the international sources into Norwegian, which is equally unfortunate.

1.1.1 Scenario

Five friends decide to each create a web page. There is no commercial value to any of the pages; they contain pictures of friends and family, blogs etc. All of the friends link to each other. Then one of them decides he wants a search engine to be able to locate his page, so he registers his page's URL with it. When the search engine indexes his site it will also, via the links, find the four other sites. Therefore it automatically indexes them too.

The active consent given by one has now, in reality, been extended to apply for the four others as well. The question that arises is this: Can the search engine legally do this? Does it need permission to do this, and if so, does it have the necessary permission? Have the rightholders of those pages, by allowing others to link to them, given consent for anyone to register and, to a certain extent, copy their site? If some of them now decide to remove their web sites, it might very well be found again through the cache function that some search engines offer. Contents that they had wanted to remove from the Internet are therefore still available.

One aspect of this is the private web sites, but what of the newspapers that offers articles free of charge for a limited time online, and after awhile

8 http://blog.searchenginewatch.com/blog/060918-082301 [05.03.07], and http://news.zdnet.com/2100-9595_22-6159108.html [05.03.07].

9 http://www.theregister.co.uk/2006/09/19/google_versus_belgium/ [05.03.07].

10 And any other major search engine for that matter.

demands that users pay to access the articles? Many of these articles can also be found again through the cache-function; free of charge.

1.1.2 History of copyright law and problems in regard to modern technology

Copyright law primarily got its importance with the invention of the printing press. Before this there was not much need for the protection of authors against for instance unauthorized copies. This was because each copy had to be handwritten, and thus the process was time-consuming and not very practical.[11] After the art of printing was discovered however, the process was much easier and the copying of books became more and more interesting from an economic perspective. The legislators began to understand that this was an area that had to be regulated, both to protect the authors' interest in order to stimulate to further innovation, and to protect the society's interests in order to fully exercise its ideal interest in exploiting new innovations. If for instance a rightholder of an invention should have complete control over this, in every aspect, it would be impossible for others to look to the invention for inspiration and in turn invent something that's even better. It is in society's interest that there is progress, and this is set against the rightholder's interest in full control over the work that is created.

The copyright law is built up with this ponderation in mind. Whichever side's interest that is preferred will differ from country to country, although with the increased international legislation, these differences will diminish. Moreover, copyright law is also based on the old type of technology. By this I mean the non-digital way. For instance the Berne Convention protects "(…) the rights of authors in their literary and artistic works", cf. Art. 1. The Convention, along with other legislation, seems at first glance to assume a material work in order to judge it worthy of protection. This, even if there for some will exist a rather philosophical view of the fact that it is the idea that is protected and not the actual manifestation of it. [12]

When modern technology advanced, with both tangible inventions like the processor, or software like the operating system Microsoft Windows, it raised issues on how to protect it, and if there existed acceptable protection as it was. For example there was a discussion regarding whether information technology were to be protected after the laws regarding patents, copyright or if there was need for a whole new legal classification. The result was in practice that copyright-laws were used when software first started to be developed, and now most aspects within information technology are protected after the rules

11 (Koktvedgaard, et al., 2005), p. 18.
12 (Rognstad, 2004), p. 16.

of copyright. [13] New treaties, directives and legislation(s) have also been developed to try managing modern technology directly. But modern technology advances rather fast, and the legislative process is not always able to keep up. So one must be prepared to use what might seem like outdated laws on issues related to new technology, and the question that arises then is if the laws are adequate for solving the issues. This is a problem that will not disappear as long as the technology advances as fast as it does, combined with the fact that very many of the jurists and others do not know enough about the technology to develop adequate laws.

1.1.3 Definitions

There are some words and expressions that are used interchangeably in daily life; for instance "web page" and "web site". I will define how I will use them in this thesis in order to better distinguish the different meanings and technologies I describe.

"Web site": an entire domain[14], for instance: http://www.dinside.no (and all its sub-pages)

"Web page": one page on its own, for instance: http://www.dinside.no/php/art.php?id=364198

"Software": Written programs or procedures or rules and associated documentation pertaining to the operation of a computer system and that are stored in read/write memory[15], for instance Microsoft Word or Apple's iTunes.

"Hardware": The mechanical, magnetic, electronic, and electrical components making up a computer system[16], for instance a graphics card or a hard drive.

"Computer": an electronic device that stores, retrieves, and processes data, and can be programmed with instructions. A computer is composed of hardware and software, and can exist in a variety of sizes and configurations.[17]

"Desktop computer": A computer designed small enough to fit on a desktop, hence the name. It's commonly used as a description of personal computers (PCs), Macintoshes and UNIX workstations.[18]

---

13 (Andersen, 2005), p. 277.

14 A domain is an address which identifies the base of what could be a long name-hierarchy. For example New York Times' domain is http://www.nytimes.com, and all its subpages will start with this address. For instance its sport section: http://www.nytimes.com/pages/sports/index.html [22.03.07].

15 http://wordnet.princeton.edu/perl/webwn?s=software [18.02.07].

16 http://wordnet.princeton.edu/perl/webwn?s=hardware [18.02.07].

17 http://nces.ed.gov/pubs98/tech/glossary.asp [18.02.07].

18 http://www.cit.nih.gov/dnst/handbook/Main/glossary.htm [18.02.07].

1.2 Sources and method

In this section I will account for the sources I will base my thesis on, both legal and otherwise. I will explain why I use them and how they relate to each other.

1.2.1 Legal sources

The legislative sources I will base my thesis on will in most respects be Norwegian law, the Berne Convention and EC-legislation. The reason as to why I rely so heavily on International legislation is that the main development of copyright and e-commerce law takes place in these forums. Even pure national copyright-law is not suited for being developed by a country alone, away from any International influence. Rightholders of different works; be it writers, painters, composers etc., will always be interested in spreading their work to as many people and places as possible; not restricted to any one country. The legislation must therefore be ready to handle this. For instance, the Nordic countries have an almost similar legislation in copyright-law. Already in the 1930ies a formalistic cooperation was initialized between the Nordic countries.[19] This is a thesis that primarily focuses on the relations between the search engines' operation and copyright law. The reason for the somewhat extensive focus on electronic commerce in chapter three is because e-commerce legislation might prove to hinder copyright infringements from having any consequences.

The Berne Convention has set the minimum standards for what is protected, and which rights the author on the one hand and the society other hand, has for exploiting the works. With its 163 contracting parties[20], it has had a significant impact on the world, and thus is very important. The EC-directives will to some extent set the standard for the legislation in Europe, and can therefore be an important source for at least indicating how the legal status in many European countries is. Finally I have chosen to look at the Norwegian legislation, in order to get at national point of view on the issues at hand.

Even if this is a Norwegian legal thesis, the international sources shall not be seen from Norwegian point of view. When it comes to interpreting international rules and directives, the Norwegian Supreme Court has, on several occasions, stated that they shall not be understood in accordance to the Norwegian way of interpretation. The legislation shall be understood in the same way as the international community would interpret it; autonomous interpretation. For instance, an EC-directive shall be interpreted the same way by a

19 Ot.prp.nr.46 (2004-2005) page 10, section 2.5 "Nordisk lovsamarbeid".
20 http://www.wipo.int/treaties/en/ShowResults.jsp?lang=en&treaty_id=15 [05.03.07].

Norwegian court as the EC-court itself would interpret it.[21] Thus this is the method I will use throughout this thesis as well.

1.2.1.1 The Berne Convention

The Berne Convention for the Protection of Literary and Artistic Works was first established on September 9, 1886. It has been amended several times; the last major revision of the Treaty was the Paris Act of July 24, 1971, as amended on September 28, 1979.

A revision of a convention is not changing a few parts, but an alteration of the complete convention. Thus the countries that are bound by the previous text are not bound by the new version. They have to sign and ratify the new convention in order to be bound by it. Because of this different countries might be bound by different versions of the Berne Convention.[22] This might create problems relating to which versions of the Convention shall apply in between countries, but this is now more or less solved as almost all of the "Berne-countries" have agreed to the Paris Act of 1971.[23]

When considering the Berne Convention in the later chapters of this thesis I am referring to the Paris Act.

The three most important factors of the Berne Convention are that it entitles authors to national treatment and the rights granted by the Convention[24], as well as it do not contain any formal requirement to obtain protection.[25]

The principle of national treatment is stated in Article 5 (1), and guarantees an author of one union country the same protection that a union country extends to its own nationals. However, this rule does not apply for the country of origin[26], cf. the wording *"authors shall enjoy (…) in countries of the Union other than the country of origin (…)"* in Art. 5 (1).

Furthermore, it follows of the text that also *"(…) the rights specially granted by this Convention"* shall be enjoyed by the authors. This means that if the union countries give their own nationals protection that does not meet the minimum requirements of the Convention, then members of other union countries will get better protection than that country's own nationals. From this it

21 Rt. 2000 p. 996 (The Bøhler-verdict) and Rt. 2001 p. 1811 (The Finanger-case)
22 (Rognstad, 2004), p. 1.
23 (Rognstad, 2004), p. 2.
24 (Sterling, 2003), p. 605, ch. 18.03.
25 (Sterling, 2003), p. 612, ch. 18.07.
26 The country of origin can either be the authors' own country, or the country where the work was first published. The latter might enable authors from non-union countries to have the right of protection from union countries.

can be concluded that a union country's own nationals are not protected by the Convention. However, most countries will not be willing to give their own nationals less protection than citizens of other countries, so the legislation will for the most part fulfill the minimum requirements of the Convention.

The requirement of national treatment is not absolute, it does have some exceptions, for instance in relations to the length of protection, cf. Art. 7 (8).[27] However, I will not consider this further as it has no direct relevance to this thesis.

Article 5 (2) states that the union countries must grant these rights without any reservations, there cannot be any requirement of form; *"The enjoyment and the exercise of these rights shall not be subject to any formality (…)"*. For instance, USA[28] cannot deny a Norwegian author copyright protection because he or she has not labeled the copies of their work with a copyright notice. USA can however make this requirement from its own citizens, cf. the section above.

Article 9 (2) allows for the Union countries to permit reproduction of works if certain conditions are fulfilled.[29] It must be a special case, the reproduction must not conflict with the normal exploitation of the work, and finally it must not unreasonably prejudice the legitimate interests of the author. This is called the *"three step test"*, and is vital when considering if a limitation to the otherwise exclusive right of reproduction that an author has, is legal. It has been carried on by recent legislation; both in the international forums[30] and in the European Union.[31]

The United Nations agency WIPO (World Intellectual Property Organization) administers the Berne Convention, along with 23 other international treaties.[32] WIPO's main task is to create a durable Intellectual Property-system for the protection of rightholders as well as the public, hereunder including copyright.[33]

The Berne Convention's impact on later legislation

The Berne Convention allows the union members to enter into special agreements among themselves, cf. Art. 20. On 20th of December, 1996 two such

27 (Rognstad, 2004), p. 4.

28 The United States of America is a newcomer to the Berne Convention. The Convention did not enter into force there before the 1th of March 1989 (http://www.wipo.int/treaties/en/Remarks.jsp?cnty_id=1045C [28.04.07]).

29 The main rule is follows of Art. 9 (1) and states that the authors have the exclusive right of authorizing reproduction.

30 WIPO Copyright Treaty, Art. 10 cf. Art. 1 (4), and the Agreement on Trade-related aspects of intellectual property rights, Art. 13. See below for a more detailed reference of these agreements.

31 The Copyright Directive, Art. 5 (5).

32 http://www.wipo.int/about-wipo/en/core_tasks.html [02.03.07]

33 http://www.wipo.int/about-wipo/en/what_is_wipo.html [02.03.07].

agreements were created, namely the WIPO Copyright Treaty (WCT)[34] and the WIPO Performances and Phonograms Treaty (WPPT)[35]. It is the WCT that is interesting in regard to this thesis and it states in Article 1 (4) that it's contracting parties are bound by Art. 1-21 of the Berne Convention, in addition to the latter's appendix.

The WCT has carried on most of the basic rights given to authors by the Berne Convention, as well as expanded some of them. However, the most important factor of the WCT, in relation to the Berne Convention, is that it also directly regulates issues connected to the digital environment; modern technology. Thus it could be argued that this is a far more relevant source for this thesis than the Berne Convention. However, the WCT are not ratified by the same amount of countries as the Berne Convention[36], which means that it has less impact on the legal status of copyright around the world compared to the Berne Convention. Furthermore, the WCT may also be called a basis for the Copyright directive considered in this thesis. I will come back to the relation between the WCT and the Directive below.

These two factors are the reason for why I have chosen to include the Berne Convention in this thesis rather than the WIPO Copyright Treaty.

Another agreement that includes parts of the Berne Convention is the *Agreement on Trade related Aspects of Intellectual Property Rights* ("TRIPS"). The TRIPS Agreement is a part of the *World Trade Organisation Agreement* (WTO). It follows of TRIPS, Art. 9 that members of the WTO shall comply with Articles 1 through 21 of the Berne Convention, with the exception of Article *6bis* (concerning moral rights). Consequently, all WTO member states are bound by these provisions of the Berne Convention; allowing the Convention to have an even greater impact. TRIPS also have a clear regulation for dispute settlement which may be used by members who have a dispute concerning the parts of the Berne Convention which is covered by TRIPS, cf. Art. 64 cf. Art. 9. The TRIPS Agreement will not be considered further in this thesis.

Interpretation of the Convention

The primary source of understanding the Convention is the text itself, but for assistance in interpreting the text there are documentation issued from the Diplomatic Conferences where the texts were adopted.[37] These publications con-

34 http://www.wipo.int/treaties/en/ip/wct/trtdocs_wo033.html [02.03.07].

35 http://wipo.int/treaties/en/ip/wppt/trtdocs_wo034.html [02.03.07].

36 It has only 62 contracting parties, which is only a little over a third of what the Berne Convention has. http://www.wipo.int/treaties/en/ShowResults.jsp?lang=en&treaty_id=16 [02.03.07].

37 (Sterling, 2003), p. 608.

tain "*(…) minutes of the various meetings held at the Diplomatic Conference, summaries of the discussions, and the "General Report,", drawn up by the Raporteur General of the Conference"*[38]. Other official sources[39] are the discussions of the Berne Union Assembly[40] and the Executive Committee[41].[42]

The first two sources are not particularly different from the supporting sources one has for interpreting Norwegian laws, although the legal history will be much more important when it comes to Norwegian law than treaties. The next two are however, as there is no tradition of allowing supplementary work to have much weight in the Norwegian legal system. In fact, it is disputed if these types of statements shall have any impact at all. The most important argument against giving these statements any weight is that they might allow the legislator to change the way a law is to be understood. The legal position is then changed without going through the formality of passing a new law, and in worst case a "new" law might be retroactive effect, in conflict with the Constitutions Section 97.[43] The legal theorists seem to agree on that the appraisement of this must be dependent on the situation. For instance, the Supreme Court has refused to give any weight to the statements made by the Norwegian Parliament when it considered the text itself, combined with the legal history, as sufficient to give a clear picture of what was the valid legal position, cf. Rt. 1983 p. 127.

However, while the Norwegian Parliament can with some ease pass new laws when its necessity is discovered, the process of changing a convention is quite a daunting task. In order to keep the convention current it is almost a necessity to be able to come with later statements. Another important difference is that while the Norwegian Parliament passes laws that govern the Norwegian people, the Berne Union Assembly consists of countries agreeing to statements that might change the meaning of a text governing themselves. The aspect of statements made after a law and/or convention was passed cannot really be compared. The consideration made behind the principle of not giving legislation retroactive effect is to protect the citizens from the states reducing their rights. A convention will govern states on a certain area, and even if the meaning of the text is changed after it has been ratified, the previous stated concerns regarding supplementary statements do not apply. The citizens will only notice the changes after the national laws have been changed, creating a situation where a new law has been passed, and not a situation where an old law is given retroactive effect.

38 (Sterling, 2003), p. 608-609.

39 (Sterling, 2003), p. 609.

40 160 contracting parties, http://www.wipo.int/treaties/en/ShowResults.jsp?lang=en&search_what=B&bo_id=7 [02.03.07].

41 39 contracting parties, http://www.wipo.int/treaties/en/ShowResults.jsp?lang=en&search_what=B&bo_id=8 [02.03.07].

42 In this thesis however these assisting sources for interpretation might not be very useful; primarily because of the age of the Convention and the rapid development of modern technology. One is therefore left with primarily the wording and the meaning of the Convention.

43 Nygaard, Nils, *Rettsgrunnlag og standpunkt*, p. 202.

The use of the Berne Convention in relation to this thesis has the prospect of being problematic. This is because it was last amended by the Paris Act of 1971, thus when it was created no one could have foreseen the technological development that started only a decade later. This raises two problems, namely concerning how to interpret the Convention in relation to modern technology, and if it at all can be applied to modern technology. I will discuss these two problems together.

The problem of interpreting the Convention in light of modern technology raises two possible solutions. Should it be interpreted in light of what existed in 1971, and what the countries authoring the Convention probably meant back then? Or should the wording be interpreted in light of the development that has happened from 1971 to 2007? The answer to this lies in the Vienna Convention.

The Vienna Convention of 1969 regulates the Law of Treaties. This can also be used to regulate conventions because legally a convention is the same as a treaty.[44]

The Vienna Convention has not been ratified by all states, for example Norway has not yet signed or ratified it. But the rules of interpretation which is set in the Convention, is considered to merely be a declaration of already existing law.[45] The International Court of Justice has actively used these principles, and referred to them as expressed in the Vienna Convention. For example in the case from 2004 about the legal consequences of the construction of a wall in occupied Palestinian territory: "*The Court would recall that, according to customary international law as expressed in Article 31 of the Vienna Convention on the Law of Treaties of 23 May 1969, a treaty must be interpreted in good faith in accordance with the ordinary meaning to be given to its terms in their context and in the light of its object and purpose*".[46] From this it follows that that the rules of interpretation can be applied, also in relation to countries that haven't actively ratified or signed the Vienna Convention. The Berne Convention must therefore be interpreted in accordance with the rules of interpretation following the Vienna Convention.

The object and purpose of the Berne Convention is mainly to protect the rights of the rightholders. This would mean that it must be interpreted in light

44 This also follows of the Vienna Convention, art. 2 where 'treaty' is defined:" (…) "treaty" means an international agreement concluded between States in written form and governed by international law, whether embodied in a single instrument or in two or more related instruments and whatever its particular designation; (…)".

45 Harris, DJ. Cases and Materials on International Law, sixth edition, p. 786-787.

46 http://www.icj-cij.org/icjwww/idocket/imwp/imwpframe.htm, section 94 [22.02.07].

of this, enabling it to follow the changing of the times, and therefore also be applicable onto the legal issues raised by modern technology.

Another source for interpretation of the Berne Convention is the above mentioned WIPO Copyright Treaty. The WCT states in Article 1 (2) that "nothing in this Treaty shall derogate from existing obligations that Contracting Parties have to each other under the Berne Convention (…)". This means that the provisions of the WCT will all be within the scope of the Berne Convention, and may be used as a clarifying factor in relation to the Berne Convention. While I will not consider the subjects of this thesis in light of the WCT by itself, it may be used as a factor in the interpretation of the Berne Convention.

1.2.1.2 EC-directives

The European Union is made up of 27 countries[47], and because the directives are mandatory[48] for every member state, they make a vast impact on the legal situation in Europe. Furthermore, several of the EC-directives bind countries outside the Union as well, through the agreement creating the European Economic Area (EEA). The EEA-agreement establishes a formal connection between the European Union and the European Free Trade Association (EFTA), giving the EFTA countries[49] access to the internal market in the EU. The agreement requires the participating EFTA states the right to the EU's internal market, like the freedoms of goods, services, capital and persons. On the other side it requires them to adopt many of the EU policies[50], allowing many directives to have an impact on even more European states.

While the EC-directives often are described as EU-law there is an important difference. The European Union is more of a political union with its base in the *Treaty on European Union*. The European Community stems mainly from the two unions, *European Coal and Steel Community* and the *European Atomic Energy Community*, and is founded in the *Treaty establishing the European Community.*[51] The legal institutions' that stands behind the legislation and governing, such as the Council, the Commission,

47 Per 01. January 2007. http://www.europakommisjonen.no/eu_guide/index.htm [22.02.2007]

48 However, the directives are not mandatory in the sense that they must be implemented in the exact from as they are given. The Member States must achieve the results the directives wish to accompish, and are bound to take the necessary national measures in order to achieve this. (Mathijsen, 2004) p. 27.

49 EFTA is made up by Norway, Switzerland, Lichtenstein and Iceland. However, Switzerland does not take part in the EEA-agreement. http://www.efta.int/ [22.02.2007].

50 (Mathijsen, 2004) p. 448.

51 Both treaties can be found in consolidating version here: http://eur-lex.europa.eu/LexUriServ/site/en/oj/2006/ce321/ce32120061229en00010331.pdf [18.02.07].

the Parliament and the Court are all institutions of the European Community[52], thus naming directives as EC-directives, instead of EU-directives.[53] The EU and the EC are connected by Article 1 (3) of the *Treaty on European Union*; "The Union shall be founded on the European Communities (…)".

In this thesis I will focus mainly on two directives; The Copyright Directive and the Electronic Commerce Directive.

- The Copyright Directive

The copyright directive[54] is also called the *Information Society Directive* (infosoc.).[55] The directive is one of seven directives in the field of copyright and related rights.[56]

The background for the copyright directive, among other directives, is that The European Union saw the need for harmonization of rules concerning the Information Society, and explicitly stated this in the *European Commission Green Paper of 27 July 1995 on Copyright and Related Rights in the Information Society.*[57] Even before this "Green Paper", EU had taken some steps to clarify that also new technology was subject to protection, for instance through the *Council Directive of May 14, 1991, on the legal protection of computer programs*. The copyright directive is also one of the directives given in order to fulfill the intentions of the "Green Paper", and deal with the three main areas reproduction rights, the right of communication and distribution rights.[58] Although there are many directives concerning modern technology, the Copyright Directive is one of the most relevant directives for this thesis, as it deals explicitly with the reproduction rights in connection with technology. The "Green Paper" can be seen to state some of the reasons why for instance the Copyright Directive was deemed necessary.[59]

- E-commerce Directive

The E-commerce Directive is another important directive concerning modern technology in regard to for instance copyright law. The Directive was created to ensure the free movement of so-called *information society services*[60] between

52 (Sejersted, et al., 2004) p. 46.
53 (Mathijsen, 2004), p. 3-11.
54 2001/29/EC
55 (Sterling, 2003) p. 862.
56 (Sterling, 2003) p. 763.
57 http://europa.eu/scadplus/leg/en/lvb/l24152.htm [04.03.07].
58 http://europa.eu/scadplus/leg/en/lvb/l26053.htm [04.03.07].
59 (Mathijsen, 2004) p. 30.
60 View 3.2.1 below for a more thorough explanation of the term.

the Member States, cf. Art. 1 (1) of the directive. It is stated in Recital 5 of the Directive that the development of information society services is hampered by a number of legal obstacles concerning the freedom of establishment and the freedom to provide services. Moreover there exists legal uncertainty in regard to "*(…) the extent to which Member States may control services originating from another Member State*". The E-commerce Directive takes a number of measures to overcome these obstacles, for instance by providing an exemption from liability for certain service providers, cf. Art. 12-14. Even if these exemptions may provide a form of immunity from criminal liability, it is not the intent of the Directive to harmonize the field of criminal law, cf. Recital 8. The Directive should therefore not reach any further than the field it was explicitly set to regulate.

However, the exemptions from liability must also be considered to apply for possible copyright infringements. This is supported by Recital 50 to the E-commerce Directive which states that the Directive and the Copyright Directive must come into force at approximately the same time, because they are co-dependent for establishing a clear framework of relevant rules.

Interpretation of the directives

In order to interpret EC-directives, one must start with the directive itself, and the understanding of the text. The text is however not always to be looked at in the same way as one is used to from the national way of interpretation. The EUs legal sources are all to be found in the different languages of the member states and every language is of equal importance. No one language takes precedence over another, as is the case in for example the Berne Convention.[61] This means that the legal sources in most cases are not interpreted strictly from the wording, but more from the meaning behind the regulations. The EC-court pointed this out in the 1982 CILFIT-case[62], cf. "grounds" no. 18-20[63].

61 If there is difference of opinion in the interpretation of the text, then the French version take precedence over the English version, cf. Art. 37 (1) (c).

62 Srl CILFIT and Lanificio di Gavardo SpA v Ministry of Health, Case 283/81.

63 "18. To begin with, it must be borne in mind that community legislation is drafted in several languages and that the different language versions are all equally authentic. An interpretation of community law thus involves a comparison of the different language versions. 19. It must also be borne in mind, even where the different language versions are entirely in accord with one another, that community law uses terminology which is peculiar to it. Furthermore, it must be emphasized that legal concepts do not necessarily have the same meaning in community law and in the law of the various member states. 20. Finally, every provision of community law must be placed in its context and interpreted in the light of the provisions of community law as a whole, regard being had to the objectives thereof and to its state of evolution at the date on which the provision in question is to be applied."

If the text therefore is looked at in context with the preamble that introduces every directive, one will get a better and perhaps more correct interpretation of the meaning of the Articles. The preamble to the directive will for the most part be highly important. This preamble consists of several Recitals. For example the Copyright directive has 15 Articles and 61 Recitals. The preamble will probably for any one directive states something about the background and reasons for the directive, as well as more specific details regarding the understanding of the articles.

However, this is not always enough and there are other sources to consider. The legal histories of EC-directives are not very important in order to understand the text. The European Court of Justice does not apply the legal history when interpreting a directive. While there might be some access to use the legal history of Conventions in order to interpret these, this access is almost non-existent in regard to EC-directives. A source that is important for interpretation of directives and other regulations however, is *the court of Justice of the European Communities*[64]. Yet, concerning the issues in this thesis there might not be much in terms of court-decisions. This is because both the issues at hand, and the legislation attempting to regulate them, are of relatively new age.

The copyright directive has one more source that can be look at as a factor in the interpretation of it; even if this source probably cannot be directly applicable.[65] This source is the WIPO Copyright Treaty. In 2000 a Council Decision stated that the European Community approved of the WIPO Copyright Treaty, cf. 2000/278/EC Article 1 (1).[66] The copyright directive was in part a follow-up to this decision[67], as it implements many of the international obligations brought on by the WCT, cf. Recital 15. Two examples of obligations following the WCT are concerning the *protection of technological measures* and *rights management information,* cf. the WCT Art. 11-12 cf. the Directive Art. 6-7.

1.2.1.1 Norwegian Legislation

The two main sources of Norwegian legislation that I will consider in this thesis are the *Copyright Act* and the *E-commerce Act*.

64 http://www.curia.europa.eu/en/instit/presentationfr/index_cje.htm [04.03.07]

65 Even if the WIPO-treaty are older, and a form of basis for the Directive, it cannot at all be viewed as form of legal history to the Directive.

66 2000/278/EC: Council Decision of 16 March 2000 on the approval, on behalf of the European Community, of the WIPO Copyright Treaty and the WIPO Performances and Phonograms Treaty.

67 In addition to the work taking place in order to implement the treaties into the Communtiy.

- The Copyright Act

The Norwegian Copyright Act is called *åndsverkloven* and is originally from 1961. It has been changed several times, the last major change being in 2005 when it had to be adapted in order to fulfill the international obligations Norway has through the EEA.[68]

Although there is not complete unity among the Nordic countries on the field of copyright law, the legislation is highly similar. Already in the 1930ies a formalistic cooperation was initialized between the Nordic countries.[69] The similarities can be illustrated by referring to the curriculum for the elective subject "Copyright law" at the Faculty of Law, University of Oslo. Here they use a Danish book as one of the textbooks in the curriculum.[70] The choice of including Norwegian law might also serve as an illustration on how the Nordic countries have adapted to the challenges copyright law is faced with as the technology progresses. However, since all the Nordic countries to some extent are bound by the same EC-legislation, the difference between the Nordic solutions and in the rest of Europe are lesser than before.

The Norwegian Copyright Act had to be adapted in order to be in accordance with the Copyright Directive, cf. the *Decision of the EEA Joint Committee No 110/2004 of 9 July 2004*, Article 1 (2).[71]

It follows of the legislative history from the last major revision of the Norwegian Copyright Act that the main purpose is to bring the law to be in accordance with the Copyright Directive.[72] Included in this revision was the section concerning the right of reproduction, cf. Ot.prp.nr. 46, section 3.1, page 14. It would be a natural assumption then that if a conclusion is made in accordance with the Copyright Directive, then one must come to the same conclusion in accordance with The Norwegian Copyright Act. That might be true in most, or even all cases, but cannot be stated as a fact before the law is interpreted more closely. It has happened before that EC directives have been implemented into Norwegian law in a way that wasn't in complete accordance with the directive. By this I mean that the intention was to bring Norwegian legislation into accordance with the directives, but the Norwegian legislators had gotten the meaning of the directive wrong. The Supreme Court in Norway ruled in the Finanger-case[73] that Norwegian law shall be interpreted in a presumptive way; it is an understanding that the legislators meant to be

68 Ot.prp.nr. 46 (2004-2005).
69 Ot.prp.nr.46 (2004-2005) p. 10, section 2.5 "Nordisk lovsamarbeid".
70 http://www.uio.no/studier/emner/jus/jus/JUR5810/h06/pensumliste.xml [04.03.07].
71 http://eur-lex.europa.eu/LexUriServ/site/en/oj/2004/l_376/l_37620041223en00450046.pdf [04.03.07].
72 Ot.prp.nr.46 (2004-2005), chapter 2.1, p. 5.
73 Rt. 2000 s. 1811.

in accordance with the international obligations Norway has undertaken.[74] However when it is obvious that the current legislation cannot be understood in the same way as the directive, then the law will take precedence. This is because it is the legislators' responsibility and exclusive task to bring the legislation in correspondence with the international obligations undertaken by the country.[75]

In addition to this, the Directive does not require the states to implement it exactly as it is. For instance the exceptions and limitations to the exclusive right of reproduction are listed in the Copyright Directive. This is an exclusive list over the exceptions and limitations that the states are allowed to grant, cf. Recital 32 of the preamble to this Directive. However, the states are not required to implement all of them. This means that something might be deemed legal after the Directive, but not after the national law of one of the states. The Norwegian law is an example of one way of implementing the Directive.

This is why I deem it necessary to consider Norwegian law as well as the Directive.

Furthermore it is worth mentioning that in regard to the Norwegian Copyright Act the limitations to the author's otherwise exclusive right of reproduction is called just that, *limitations,* instead of *exceptions*, as they are called in for instance the Copyright Directive. This is because the society's interest in the works created is weighted in the author's interest in having complete control over them. These are equal considerations, and the term "limitations" is therefore appropriate.

With this in mind it must also be mentioned that one of the main considerations behind the extensive protection rightholders get is based on the need to protect their economic interest. For most people this will be the main reason for creating works. If an author's work was not protected against others getting economic gain from it without his consent, this would severely stagger the creative process in a society by leaving many to opt out of taking the artistic path.

- The E-commerce Act

The Electronic commerce Act entered into force in 2003. Originally the exemptions from liability that the Directive provided were not included in the law, but this was planned and quickly remedied, cf. the legal history; Ot.prp.no.31 (2002-2003) page 12. The adaption of the E-commerce Act is a direct consequence of the E-commerce Directive, and hence the objectives and scope of the Act is the same as the Directive, cf. the Act, Sec. 1 cf. the Directive Art. 1.

74 Rt. 2000 s. 1811, p. 1826.
75 Rt. 2000 s. 1811, p. 1832

1.2.1 Other sources

The Internet is used widely as a source for technical information in this thesis. The main reason for this is that technology will upgrade all the time and the sources on the Internet are often the ones that are updated most recently. Books will fast get outdated. However, the Internet is also notoriously unreliable. Everyone can publish information without anyone ever checking the accuracy of it. One indication of the seriousness of a web page is how the information is presented. While many might lay much work into creating a believable web site for scams, there are fewer that will bother to do so in order to misguide people as to how for instance servers or clients work.[76] This combined with the assessing of what actually is written there, and who has written it will give some clue as to how reliable the information is. Who has linked to the site, and how one came by it will give further indications to its reliability.

1.2.1.1 Wikipedia

Wikipedia.org is an online encyclopaedia. Its contributors are all the users of the Internet that want to participate. It is available in at least 34 different languages with the English version being the largest with 1,741,164 articles.[77] As a comparison it is worth mentioning that the Norwegian version has about 109,705 articles.[78] The fact that every article is not checked for accuracy by someone trusted as a "completely reliable source", is reason to be careful with the information one finds here. To compare, one can in most cases be completely sure that for example Encyclopædia Britannica[79] has checked its sources, and can be trusted. However it is proven that Wikipedia›s methods with the users securing that the information is correct has worked. At least it will probably be correct in regard to areas that are interesting and relevant for more than a highly limited circle of people. Mistakes do occur, but they are normally quickly discovered by others and corrected. This is part of the reason why I use it as a source. Moreover, other authors also refer to Wikipedia; Mads Bryde Andersen refers to wikipedia.org as a source for technical terms and information in his book "IT-retten".[80] Furthermore, it has happened on occasion that American courts have used Wikipedia as a source for information.[81] It is because of these reasons that I find Wikipedia trustworthy enough

76 For a closer explanation of these two terms, se chapter 2.2.1 below.
77 http://en.wikipedia.org/wiki/Main_Page [16.04.07].
78 http://no.wikipedia.org/wiki/Hovedside [27.04.07].
79 http://www.britannica.com [18.02.07].
80 (Andersen, 2005), p. 25.
81 http://www.bt.no/utenriks/article336207.ece [05.03.07].

to use as a source for some of the technical information that I need, although I will use it with caution.

1.3 Challenges faced in this thesis, and the road ahead

This subject presumes the knowledge of the technology behind the search engines operation, and thus I will have to devote some parts of the thesis to the explanation of this technology.

This is because without an understanding of how the technology *really* works, it will be nearly impossible to fully comprehend the problems faced in relation to it. This is not only a problem in regard to copyright law, but concerns all legal practice regarding information technology.[82]

The first major question treated in this thesis is whether or not the search engines actually reproduce web pages, and if so, if they violate the presumptive exclusive right a rightholder has of reproduction. If this reproduction is found to be a violation then I will analyze if any of the limitations to the otherwise exclusive right of reproduction apply for the search engines. The last discussion in this respect will be whether the rightholders have already consented to some forms of reproduction, because they have published the web pages on the Internet.

The other major question of this thesis is in regard to the cache function that some search engines' offer. Is the cache a reproduction, and if so, do the search engines have the necessary legal grounds operate these caches? If they do not have the legal grounds, can they be held responsible, or will the exemptions from liability passed with the electronic commerce legislation apply for the search engines' caches as well?

Finally I will look at possible alternative solutions for ensuring the continued and unhindered operation of the search engines; especially some aspects of American Law will be interesting in this respect.

82 (Andersen, 2005), p. 75.

2 SEARCH ENGINES' INDEXING OF WEB PAGES

2.1 Introduction

Before I discuss if the search engine has the permission to index web pages,

I must establish if it needs this permission, i.e. that indexing actually is a reproduction in relation to the legislation.

One of the most important rights that the copyright legislation intends to protect is the rightholder's exclusive right of reproduction. Whether one studies the Berne Convention, EC-directives concerning copyright or Norwegian copyright law, one finds that this is at its centre. By this I mean that on the one side there is the exclusive right of reproduction which is held by the rightholder, and on the other side there are the exceptions and limits to this otherwise exclusive right. The rightholders right to reproduction is the *"exclusive right to authorise or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part"*, as provided for in Article 2 of the Copyright Directive.[83]

The exclusive right a rightholder has to control his own work is divided in two, cf. the Norwegian Copyright Act, Sect. 2. These two parts are the right of reproduction and the right to make the work available to the public. One could have an interesting discussion on both of these rights in relation to the methods of Internet-based search engines, but in this thesis I will only concentrate on the right of reproduction. This will enable me to make a more thorough analysis of the subject.

The exclusive right of reproduction has a "negative" and a "positive" side. The negative side would be the right to prohibit others from doing anything with the work in question without the consent of the rightholder. The positive side is the right for the rightholder to do what he wants with his or her work.[84] These rights have been implemented in Norwegian law mainly through the Norwegian Copyright Act, Sect. 2 and Sect. 3. One could say that while Sect. 2 provides for the rightholder's economic rights, Sect. 3 provides for the moral sides of the exclusive rights, for instance the right to be named on a reproduction.[85] The

83 http://eur-lex.europa.eu/LexUriServ/site/en/oj/2001/l_167/l_16720010622en00100019.pdf [18.02.07].

84 (Rognstad, 2004), p. 76.

85 (Rognstad, 2004), p. 76.

perspective that copyright law is divided into economic and moral rights is also well established from an international point of view.[86]

There are claims that indexing does not constitute a reproduction. These claims often come from people with little or no legal education, but with vast technological experience. One such man is for instance Danny Sullivan. He is well reputed when it comes to search engine technology.

For example, at the "Search Engine Strategic Conference" there was a session called "Conversation with Eric Schmidt hosted by Danny Sullivan", which featured Eric Schmidt, the CEO and chairman of Google Inc.[87] A session which is also transcripted at Google's own press page.[88] Here Sullivan both asked questions as well as communicate questions from others.

Sullivan can be understood to say that if you mean indexing is a reproduction, then you have not really grasped what indexing really is. He says: "Nothing in the index is anything you as a human being could read".[89] By this he means that an index breaks down a site in different parts. Sentences are taken apart and stored word by word, with different IDs identifying if they for example were part of a hyperlink, a title and so on. The question that arises from this is if there are any demands for accessibility when making the decision if something is reproduced. Is it so that if a search engine copies all of a site, but breaks in up in pieces so it can decide how much and what of the contents a user can see, then it isn't a reproduction? Is the defining point of a reproduction how much information is spread out amongst the users of the service, or how much is gathered by the company that provides the service?

Before this can be analyzed one must understand how the search engines works; more specifically one must understand the index and the complex process of creating it.

2.2 How the search engines' operate; the gathering and indexing of web pages

Search engines operate by different techniques and with different technological solutions; how they locate information, how they process this information and

86 WIPO Worldwide Academy, Course "DL-001 Primer on Intellectual Property", Module 2, p. 3.
87 http://en.wikipedia.org/wiki/Eric_E._Schmidt. [18.02.07].
88 http://www.google.com/press/podium/ses2006.html [18.02.07].
89 http://blog.searchenginewatch.com/blog/051021-113341 [18.02.07].

finally how they store it. Some things are similar however, and I will mention these aspects first. In order to do this I must introduce certain terms.

2.2.1 The collection of data

The index is vital to any search engine because it is what makes the collected data searchable in an efficient way. The search engine gathers data to be indexed with the use of web crawlers ("crawlers"). Before I further explain the functions of the index I will explain some technical expressions:

2.2.1.1 URL and hyperlinks

An URL (Uniform Resource Locator)[90] is a document's address on the Internet.[91] In order to find these addresses one needs directions. The most common way for giving these directions is hyperlinks. These are also called just "links". A link can *"(...) refer to a document, e.g. a web page, or other resource, or to a position in a web page"*.[92] There are different types of links. You have the ones pointing you to the main page of a web site, for instance http://www.nytimes.com. A link can point you to a certain point in a web page; for instance a certain paragraph.

2.2.1.2 Client

As mentioned above, the search engine gathers data from the Internet through web crawlers.[93] The "home" of this crawler is the computer where it originates; *the client*.[94] A client is a computer where a request is sent from, and it is not something exclusively used by search engines. It is common in every aspect of daily use of computer technology, be it private or business. When one sits at home using the Internet, one does so through a web browser. Typical web browsers are Internet Explorer, Mozilla Firefox or Opera. When a person types in an URL in the address field, the browser sends out a request to the server that the web site resides on, and will get a response which it stores in a temporary internet file on your computer and then shows the site requested on the screen. That person's computer then acts as a client, using the web browser to send out requests much in the same way that a web crawler does. The differ-

90 http://en.wikipedia.org/wiki/Uniform_Resource_Locator [18.02.07]
91 (Andersen, 2005), p. 66.
92 http://en.wikipedia.org/wiki/Hyperlink [18.02.07]
93 I will explain the web crawler's function more thoroughly below in section 2.1.1.4.
94 (Fielden, et al., 2002), p. 15.

ence is primarily that the web crawler is programmed to send out automated requests on a very large scale, while the browser will wait for a human user to input a specific request and will process only that request.

2.2.1.3 Server

A server is a computer which has been configured to share its resources and run applications for other computers.[95] A server can fulfill a wide aspect of tasks, for example it can function as a printer server, a mail server or an application server. For instance, if someone has several computers at home and wants to be able to use the printer for all of them, it would be rather impractical having to physically connect each computer to the printer. Instead a good solution would be to set up a home network and configure one of the computers to share the printer with the rest of the computers on this network. This example illustrates how a regular desktop computer also can function as a server.

Although this is widely used, it must be said that for the most part servers are so called dedicated servers. This means that they do not fulfill any other function than the server part. Larger businesses will for example set up large servers which contain all the software the employees have access to. This can be done in several ways; I will mention one of these possibilities. In a business the employees might have their own desktop computer and can access the servers in order to either install or update a program. The following example might illustrate the practice: While users operate with at local hard drive, which the software is installed on, they won't have the direct access to the installation files. That is because these are stored on a server which might not even be in the same building or city, depending on the size or preferences of the business.

This is one of many possibilities of how to use a server, and it is not possible or necessary to go through more of them in this regard. Although it is important to note that search engines also will use servers as storage areas for the web pages they have crawled.

2.2.1.4 Web crawlers

The web crawler is a program that collects data from web pages on the Internet, and it finds the pages through their respective URLs. The crawler does not actually go around the Internet gathering data, but instead it sends out requests to a server asking to return a specified web page, and then scans

95 http://www.webopedia.com/DidYouKnow/Hardware_Software/2005/servers.asp [18.02.07].

that page for data and further URLs.[96] Then it follows the URLs found, and the process has started.

The crawlers will identify which URLs to visit with the use of a URL server.[97] Each search engine will typically use several crawlers. Google estimates that they used about three crawlers simultaneously in its early beginning.[98] This allows the search engine to cover a lot of ground in a short time period. Different crawlers might also be given different task. For instance one crawler will be programmed to check URLs that already have been indexed, in order to check for updates, while another might be focused on crawling newly discovered web pages.

When a crawler brings back a new URL, the corresponding web page will be given a specified document ID so that it may be retrieved again later.[99]

2.2.1.5 The index

The index is vital to any search engine. It is the index that enables the search engine software efficiently and reliably to search every web page the web crawler has "brought back" to the search engines servers. The index itself can to a certain extent be compared to an index file in a library, which will show what a book contains in keyword form, and where the actual book can be located. However, the indexing process is more than just the index itself.

It is not possible to give an exact description on how each search engine operates, even if one just looks at the major and well known engines. This is because the modus operandi, including storage, process of information, and finally how the search engine ranks results, are well-protected business secrets.[100]

However, it is possible to say something about their ground-structure. There are some things, like a database, that every search engine must possess, in order to expediently and safely process the information it handles. The index cannot necessarily be called a database in itself, but is built on top of one or more databases. There are also other aspects of the indexing-process that

96 http://www.google.com/librariancenter/articles/0512_01.html [18.02.07]

97 Some crawlers will exclusively look for new URLs and store these in the URL server. Other crawlers are dedicated to search the actual information of the page, and will be told where to go next by the URL server.

98 http://infolab.stanford.edu/~backrub/google.html ,- part 4.3 "crawling the web" [18.02.07]

99 http://infolab.stanford.edu/~backrub/google.html ,- part 4.1 " Google Architecture Overview" [18.02.07].

100 How to retrieve the most accurate hits to every search in as little time as possible is the most important goal for search engines. And they continuously develop new techniques in order to obtain this goal.

are necessary to explain, but an understanding of how a database works is paramount in order to understand how an index relates to the complete data stored by the search engine.

Database

A database is a collection of tables containing columns and rows. The different data is input in the different columns and rows, much like a spreadsheet in a calculation program like Microsoft Excel. Each column will define a value and each row will contain single entries corresponding to that value. To illustrate we can picture a database that contains two tables. Table one contains personal information, and has the columns "Personal Identification number" (PIN), "last name", "first name", "address" and "zip code". A row in this table would look like this:

PIN	Last name	First name	Address	Zip Code
01020312312	Doe	John	Happyroad 26	9400

(Figure 1)

Table two would contain zip codes and the corresponding postal addresses. A row in this table would look like this:

Zip Code	Postal address
9400	Harstad

(Figure 2)

Now, this model of a basic database could easily help us search and present information. Let us say that we have 100 different zip codes with corresponding addresses, and a 1000 people listed in table one, evenly distributed on the different zip codes. The database would then enable us to make a list over for example every one that lived in the area with the zip code '9400'. However, updating this database manually would be a daunting task, especially if it becomes extensive.

This is why the most common type of databases is the relational database. This database structure lets us connect different tables to each other[101], provided that they have at least one similar value. If we use the example above,

101 http://php.about.com/od/learnmysql/ss/mysql_3.htm [18.02.07].

we see there that the value "zip code" would be the primary key in table two, and the foreign key in table one.

A primary key is an entry that can be used to uniquely identify a row in a table.[102] In table one in the example above, every one of the values listed (in the columns), apart from the Personal Identification number, would be a bad idea as a primary key because the risk of similar input is high. For instance two or more people might have the same name. Therefore it is normal to add a key value that one could be sure is unique. If table one for example listed employees at a firm then one could be sure that the Personal Identification number will be unique for each employee. Therefore it would only correspond to that employee's personal information. A foreign key is a value that also exists in another table, often as the primary key in that table.[103] The function of the foreign key is to link the tables together. The foreign key also prevents wrongful input in the table where it is located. The value it represents in the table it "originates" from must exist as an entry there.

When the zip code is entered into table one, the corresponding postal address will automatically be inserted into the row because the tables are linked. An extract of the table would then look like this (the border style and arrows are for illustrative purposes only).

PIN	Last name	First name	Address	Zip Code	Postal address
01020312312	Doe	John	Happyroad 26	9400	Harstad

Zip Code	Postal address
9400	Harstad

(Figure 3)

This saves time because once the postal address is registered in table two it will not have to be manually registered in any table in the whole database again. Any changes to be made will be done in table two, and the rest of the tables that are linked to this table would automatically be updated. If for instance Harstad changed its zip code from '9400' to '9500', the database administrator would only have to input the change once in table two, and table one would automatically update. If there for example are 10,000 employees listed in the table one, and 8 000 of them lived in the part of Harstad which had the zip code 9400, it is easy to understand how the relational model saves time.

102 http://en.wikipedia.org/wiki/Primary_key [18.02.07].
103 http://en.wikipedia.org/wiki/Foreign_key [18.02.07].

The model also prevents double-posting of data, since you would not have to repeat any input. By this I mean that one would have to do the original input in the table where the value originates. For instance; if I try to input the zip code 5555 in table one and that entry does not exist in table two, the database would not let me do that. I would first have to register it in table two.

This was a rather basic illustration on how a relational database works. In a larger database, like the ones search engines make use of, the workings of the database would be much more complex. The basics remain the same, but the database is expanded to such a size that on first sight it might be hard to relate it completely to the example above. But to have the workings of such a database in mind will help better understand how the index works.

The process of creating the index

When the crawlers return with the data from web pages, they will send the web pages that are crawled to a store server[104], which in turn will compress and store this data in full in a repository[105]. In regard to most search engines the repository will most likely be a database, and thus the data is stored in tables. But to search all of the data in the repository when a user inputs a query would be extremely ineffective and would probably take hours, if not days, depending on the size of the search engine.

The database therefore needs an index. The index will be a copy of a part of a table.[106] This means that the data is still stored in the table where it was registered, but the index will contain the data necessary to make the database searchable, and it will manage the data in a way that makes the search effective and accurate.

The process of indexing a crawled and stored web page is the product of a number of functions, performed by an *"indexer"* and a *"sorter"*.[107] The indexer will read the data in the repository before it will uncompress the files and analyzes them.[108] The indexer will then record the word and other information about it, like its position in the document and approximate font size before it distributes this data into storage. In this storage the words will be connected to the DocID, creating what is called a forward index.

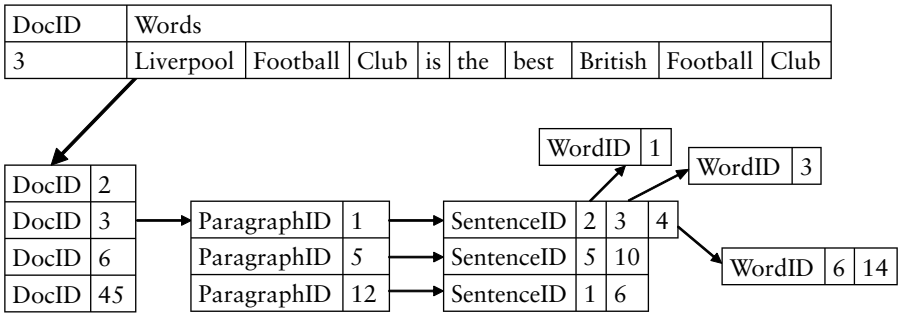
104 http://infolab.stanford.edu/~backrub/google.html - part 4.1 "Google Architecture Overview" [18.02.07].
105 http://infolab.stanford.edu/~backrub/google.html - part 4.2.2 "Repository" [18.02.07].
106 http://en.wikipedia.org/wiki/Index_(database) [18.02.07].
107 http://infolab.stanford.edu/~backrub/google.html - part 4.1 Google Architecture overview [18.02.07]
108 See previous footnote (Google Architecture overview).

And extract from the forward index will in simplified form look like this: [109]

DocumentID	Words								
3	Liverpool	Football	Club	is	the	best	British	football	club
10	Football	clubs	are	big	businesses	in	the	world	today

(Figure 4)

There will be several tables keeping track of these IDs. One will perhaps contain only the DocID, and every word that corresponds to that DocID. Then there is the tables keeping track of the words. One might contain the data as to in which paragraph the word was located. Another might be connected to this table, and keeping track on which sentence the word appeared in, under the actual paragraphs. And in the end of this chain of tables there might be a table keeping track on the word itself; where it appeared in the sentence. There might also be tables storing data concerning if the word was in capital letters, italicized, in bold letters and so on. This all depending on how complex and accurate a database is. A search engine's database must be quite accurate in order to get as good and relevant results as possible back to the user. Below is an illustration on how the tables might be linked.



(Figure 5)

This shows that the word "Liverpool" can be found in Documents 2, 3, 6 and 45. In document 3 the word appears in paragraphs 1, 5 and 12. In paragraph 1 the word is found in sentence 2, 3 and 4. In sentence 2 it is the 1st word. In sentence 3 it is the 3rd word, and in sentence 4 it is the 6th and 14th word.

109 http://en.wikipedia.org/wiki/Search_engine_indexing [18.02.07]

The next step in the indexing-process is to invert the forward index.[110] This will lead to the search engine being able to effectively search for web pages that matches key words input by users of the search engine software.

To illustrate a search in the now inverted index we can use the phrase *Liverpool Football Club*. The index will for example show that the word *Liverpool* has appeared the entries which have the DocID 2, 3, 6 and 45, while word *football* appeared in entries 2, 3, 10 and 45, and *club* appeared in entries 1, 3, 10 and 45.

If a search now is executed with the above phrase, the search engine software will look through the inverted index looking for matches. It will first look at entries that match each keyword, and then look for which entries (and thus web pages) that contain all of the words.

Written out like a table, the process would in a simplified form look like this:

Keyword	**Document IDs**			
Liverpool	2	3	6	45
Football	2	3	10	45
Club	1	3	10	45
All keywords		3		45

(Figure 6)

It is then most likely that the most relevant web sites in accordance to the search will be the ones with the IDs 3 and 45. The search engine will now know, as illustrated with the example in figure 5, that in document 3 the word "Liverpool" is the first word in the second sentence in paragraph one. If the word "Football" is the second word in the second sentence in paragraph one, while "Club" is the third word, it is very likely that this document is relevant to the query.

The different search engines have different technologies and methods that they use to rank the pages that seem to be an equal good hit. For instance, one of the methods used by Google is to see who has linked to the pages in question.[111] If, for example, the Football Association of England[112] has linked to the web page which has DocID 3, the search engine will have a good clue

110 http://www.google.com/librariancenter/articles/0512_01.html - section three under "Crawling and Indexing". [10.03.07].

111 http://www.google.com/librariancenter/articles/0512_01.html - "Ranking results" [10.03.07]. This method is called "Citation Ranking", originally developed by Eugene Garfield in 1960, http://en.wikipedia.org/wiki/Citation_index [10.03.07].

112 http://www.thefa.com/default.aspx [18.02.07]

as to which of the two sites in question is the official, and thus probably most relevant, web site for Liverpool Football Club.

The index will in most cases not be stored on just one computer. For instance Google uses several, so that each computer can simultaneously look through their part of the index. Because of this the index as a whole will be searched a lot more efficiently than if only one, albeit powerful, computer should do all the work. Google estimates that over 500 computers will be in action for each search made by a user, and this leads to a search result being brought back to the user in less than half a second.[113]

2.2.2 Is indexing a reproduction?

Introduction

The question is if it's both the indexing process and the index that creates the reproduction, or if a reproduction already must have happened for the index to be created. If neither is a reproduction then there is no issue, because then the search engines' would not have breached the copyright legislation.

Even though the index is only a part of the process when a search engine makes web pages searchable, it is the most crucial part. The search engine cannot do without the index. It can in theory do without the web crawler. This will be unpractical, but could for example work if a search engine bases its operation on people actively reporting in links; which is the case with human-powered search engines.[114] It cannot do without databases either, but the index is what makes the database work in a practical way. The indexer, which creates the index, will uncompress the gathered data, sort it, and thus make it usable. The database without an index would be similar to making copies of books, translate them to an unknown language and then store the copies in a basement, never to be seen by anybody. In principle this would also be a violation of the right of reproduction, but the damage done to the rightholders would be reduced to a minimum, if there is any damage done at all. The search engine software's main function is to allow the user to communicate with the data stored in the database, through the index.

113 http://www.google.com/librariancenter/articles/0512_01.html - Ranking results, section five [18.02.07].

114 http://www.webopedia.com/DidYouKnow/Internet/2003/HowWebSearchEnginesWork.asp [18.02.07].

Because the indexing-process is the key to the search engines gathering and distribution of data, it is what I will consider in regard to the right of reproduction[115], and not for instance the crawler's mode of operation.

I will discuss this problem in relation to each of the different legal instruments I consider in this thesis.

2.2.2.1 The Berne Convention

The first problem that arises is whether or not the Berne Convention can be said to protect web pages.

While it is obvious that the two other legal sources; the EC-directive and the Norwegian Copyright Act, protects web pages, this question is considered in relation to the Berne Convention solely because of the age of the Convention.

As mentioned in chapter one, the Vienna Convention states that a convention shall be interpreted in *"(...) good faith in accordance with the ordinary meaning to be given to its terms in their context and in the light of its object and purpose"*, cf. Article 31. The Berne Convention's paragraph 1 states that the Union is for the protection of "authors of literary and artistic works". Does a web site fall under the expression *"literary or artistic work"?* Article 2 in the Convention lists what the expression as a minimum shall include.[116] It will include *"(…) every production in the literary, scientific and artistic domain, whatever be the mode or form of its expression, such as books, pamphlets and other writings (…)"*. The specified list that follows this is, as stated above, exhaustive.[117] And nothing that is listed here would naturally and immediately be equivalent to a web page. For example, the text in itself could constitute a literary work, but it is not only the text that makes a web page.[118] And it is the entire web page that must be protected in this regard, not just the text itself.

At the beginning of the 1980ies definitions of what is a literary work expanded to include works that might not seem like a literary work at first glance. USA revised the *U.S. Copyright Act 1976*[119] in 1980 to include computer software under the category *"literary works"*, cf. Sec. 102.[120] This was a beginning for granting copyright protection to works generated by a com-

115 I will however look at the *cache* in chapter 3, which is a more special form of reproduction performed by some search engines.

116 (Sterling, 2003), p. 610, ch. 18.06.

117 However, union countries are free to expand this list if they wish to protect other types of works as well. (Rognstad, 2004), p. 6. They just cannot abstain from protecting the works listed in Art. 2.

118 For instance the composition of pictures, sounds, colors etc. are vital part of a web page.

119 The Copyright Act is codified as *Title 17, U.S. Code.*

120 http://www2.library.ucla.edu/copyright/2123.cfm [10.04.07].

puter. The European Union supported this line by stating in Council directive 91/250/EEC[121] that computer programs are to be protected by copyright *"(…) as literary works within the meaning of the Berne Convention (…)"*.

Furthermore, also the WIPO Copyright Treaty states that *"Computer programs are protected as literary works within the meaning of Article 2 of the Berne Convention. Such protection applies to computer programs, whatever may be the mode or form of their expression"*, cf. Art. 4. The *Agreed Statements concerning the WIPO Copyright Treaty* explicitly states that the scope of protection granted for computer programs is consistent with Article 2 of the Berne Convention.[122]

This has now evolved to be a general principle.[123]

Moreover, one of the purposes of establishing the Berne Convention was to ensure a set of minimum rights for rightholders. The rightholders of a web page have as much need of protection as the rightholders of a book. Because of this the conclusion must be that a web page also is protected by the Berne Convention.

Has a reproduction taken place?

The third paragraph in Article 9 states that any sound or visual recording always will be considered as a reproduction, and thus it clarifies what most certainly is meant by *"any manner or form"*, though not in an exhaustive way. However, the index cannot be said to be a sound or visual recording, and the Art. 9 (3) is therefore not applicable in this respect.

The question that remains is whether this can be considered as a reconstruction that fits the wide definition *"any manner or form"* which is found in Article 9, paragraph 1. There is nothing connected to the wording that states that the reproduction must be easily accessible, or obvious. The information is stored by the search engine, and with the right knowledge and time it can be recreated. The wording *"any manner or form"* is, as previously stated, quite wide, and thus it should be very likely that even the way the index process and stores the web sites must be included.

Moreover, the *Agreed Statements concerning the WIPO Copyright Treaty*[124] *states that the reproduction right, cf. Art. 9 of the Berne Convention, "(…) fully apply in the digital environment (…)"*. It shall also be understood that *"(…) the storage of a protected work in digital form in an electronic me-*

121 Council Directive 91/250/EEC of 14 May 1991 on the legal protection of computer programs.

122 "Concerning Article 4".

123 (Koktvedgaard, et al., 2005), p. 57.

124 "Concerning Article 1 (4)".

dium constitutes a reproduction within the meaning of Article 9 of the Berne Convention".

These factors viewed together sum up to the conclusion that a search engine's indexing process is a reproduction in relation to the Berne Convention.

2.2.2.2 The Copyright Directive

The Copyright Directive states in Article 2 that any form of reproduction, be it in whole or in part, in whatever form or means, is at the discretion of the rightholder. The indexing process will, as stated above, take the full contents of the web pages crawled and make this content searchable, thus creating a full reproduction. The legal requirement *"in whole or in part"* will therefore be fulfilled. No matter which type of software is used to create the index, or the few differences that might exist between the methods search engines uses to create the index.

The claim from different experts of technology that indexing doesn't constitute a copy because no human being will be able to read the material in context because of the way a index is built, is not supported by article 2. This is because it also includes copying in *"(…) whatever form (…)"*. There is no support that it must be readable from a human's point of view, neither from reading the legal text nor the preamble to the directive.

However, the contents indexed are not unattainable for a human, even without the software that automatically searches the index for hits to a query. This software and the systems used to manage the database will allow for the search engine to "put together" a web site again at an extremely rapid pace, because all of the words have been stored with an ID linking them to a certain document. This ID also contains information about where in the document the words were placed, the approximate font size and so on[125], as mentioned above. For a human to do the same would be a daunting task, but with the IDs assigned available it would be possible; it would just take a vast amount of time. It would be possible to link all of the IDs together, and then find the corresponding document that is stored on one of the search engine's servers. It can in some way be compared with the situation where someone is suddenly presented with a work written in Braille. It would be a daunting task to try and read it, but with the corresponding alphabet available; one might say the words' ID in Braille, it would be completely possible. It is in fact done rather often.

The conclusion must therefore be that a search engine's indexing process is a reproduction in relation to the Copyright Directive.

125 http://infolab.stanford.edu/~backrub/google.html , part 4.1 "Google Architecture Overview" [18.02.07].

2.2.2.3 The Norwegian Copyright Act

The right of reproduction is regulated by the Norwegian Copyright Act, Sect. 2. It states that the one that holds the copyright to the work has the exclusive right of reproduction. This right includes both permanent and temporary reproductions. Section 2 further explains that it is also considered a reproduction when the work is transferred to a device that can reproduce the work. A natural understanding of the wording is that every device that might reproduce the work is included, both electronic and otherwise, thus including computers. The device must only have the ability to reproduce the work, it does not specifically state that the device must have the ability to do so alone, i.e. that it does not need additional programs or devices to do so. This would include storage on a hard drive, which would require computer power to show its contents. It also follows of the commentaries in "Karnov" that also devices that can be read by computers must be included.[126]

Now, the word "transfer" might be understood to mean that there has to be a transfer of information from one way of presenting the work, to a completely different way. For example by scanning a book into a computer. However there is no basis to support this, a transfer between equal devices should be covered as well. The indexing-process will therefore constitute a reproduction because it violates the *exclusive* right the rightholder has to decide how and when his work should be reproduced.

Sect. 2 (1) explains what is granted by copyright protection. The understanding of the wording *"Copyright gives, within the boundaries stated in this law, the exclusive right to (…)"*[127] suggests that the copyright does not necessarily belong to the author of the work. It opens up for these exclusive rights to be transferred, in most cases that will happen for a price. This is in contrast to Sect. 3, where the rights granted are connected to the author in person. Section 3, paragraph one, governs the moral rights of a work, and states that the author shall be named in accordance with fair practice on each copy of the work.

Although the rightholder is able to give away the right of reproduction, it is not therefore said that he or she has given away that right by publishing it on the Internet. If that were the case, then all rightholders of web pages had given away this exclusive right. And to whom is it given? All of the users of the Internet? That the rightholder might have given away part of the right, in the meaning that some reproduction must be allowed, is another question which will be further debated below.

126 Note 4 to Sect. 2 (2). Karnov is also called "Norwegian law commentaries" and is published by Gyldendal publishing.

127 " Opphavsretten gir innen de grenser som er angitt i denne lov, enerett til (…)"

The conclusion must therefore be that a search engine's indexing of web pages is a reproduction in regard to the Norwegian Copyright Law.

2.3 Exceptions and limitations to the right of reproduction

It has now been established that the search engines' indexing process is a reproduction in regard to all copyright legislation considered in this thesis. As a starting point it can be said that they therefore need consent from the rightholders in order to continue with their practice. However, the rights granted to the rightholders are the result of a consideration between the need to protect the authors of works, and the need the society has to utilize those works. It is important to let the rightholder have as much control over his or her work as possible, so that further creation is incited. Therefore the different legal sources include somewhat of a downscaling of, and in some cases exemptions to, the otherwise exclusive rights the rightholders have. There are several rules regulating this, and I will only consider those that might be relevant in short, ruling out the ones that obviously cannot be applied.

All the legal sources state the right of quotation as one of the key limitations to the otherwise exclusive right of reproduction, cf. The Berne Convention Art. 10 (1), cf. The Copyright Directive Art. 5 (3) (d), cf. The Norwegian Copyright Act Sec. 22. Even if the subtract of the web page that is presented on the search hits-page might be considered as a quotation, the index itself cannot be considered as such. The limitation does not therefore apply to the index.

There is another important limitation to the right of reproduction that follows from the Copyright Directive and the Norwegian Copyright Act; namely the right of temporary reproductions. However, the index cannot be said to be temporary in nature, and because of this it cannot be considered in relation to this regulation. I will however return to this regulation below in 3.1.1, when discussing the cache-function.

2.4 Is there consent for the indexing done by a search engine?

It has been established above that the current legislation by itself does not provide the necessary legal authority that search engines' need in order to legally index web pages. However, it is completely possible that the rightholders of the web pages either have the possibility of explicitly consenting to indexing, or perhaps already have consented through implicit consent.

2.4.1 Explicit consent

The rightholders can give their consent to their work being reproduced. This follows from every legal source analyzed in this thesis. Both the Berne Convention (Art. 9) and the Copyright Directive (Art. 2) give the *rightholders* the exclusive right to *authorize* reproduction, while the Norwegian Copyright Act give the one(s) that *hold the copyright* the exclusive right to *control* the work, cf. Sec. 2 (1). There are several ways of giving explicit consent, and I will look at the two most prominent ways.

2.4.1.1 Agreeing to terms when publishing the web site

If a person does not buy a domain on the Internet, for example http://www.news.com, he or she can rent or get space through a web hosting service, for example http://home.no.net/news[128]. There are several web hosting-service, some are free while others charge for their services. These services will normally require that a user accepts a set of terms and agreements before gaining access. One normally accepts such agreements by clicking on "I accept" when presented with the terms of service.[129] It is not uncommon that these agreements contain a passage which regulates the access for search engines to index the page.[130]

Geocities

Geocities is a very popular web hosting-service because it's free and well known. It is owned by Yahoo! and as such the terms of service are dictated by them. This is an extract from what a person has to agree with before he gets the opportunity to create a web site at Geocities.

> *"However, with respect to Content you submit or make available for inclusion on publicly accessible areas of the Service, you grant Yahoo! the following worldwide, royalty-free and non-exclusive license(s), as applicable: (…) the license to use, distribute, reproduce, modify, adapt, publicly perform and publicly display such Content on the Service solely for the purposes of providing and promoting the specific Yahoo! Group to which such Content was submitted or made available."*[131]

128 The home.no.net address is provided by www.start.no. [22.03.07]

129 So-called "click-wrap".

130 If someone buys their own domain on the Internet, but does not have a server of their own, they might have to rent space from a web hotel. These hotels might have similar terms of service.

131 from Yahoo! Terms of Service point 9. "CONTENT SUBMITTED OR MADE AVAILABLE FOR INCLUSION ON THE SERVICE"; http://info.yahoo.com/legal/us/yahoo/utos/utos-173.html [22.03.07].

Googlepages

Google offers space for hosting a web site as well, through their service "Googlepages"[132].

Below is an extract of their terms of service:

> *"By submitting, posting or displaying Content on or through Google services which are intended to be available to the general public, you grant Google a worldwide, non-exclusive, royalty-free license to reproduce, adapt and publish such Content on Google services solely for the purpose of displaying, distributing and promoting Google services."*

One will see that both the services reserve the right to *reproduce* the contents in order to promote their own services. Yahoo! limits it to the group the content might be related to, while Google makes no such distinction. These terms of service must be considered to include reproduction in relation to the search engine function of these companies as well, particularly since this is a highly central part of their business. A valid question is however if one by accepting these terms only comply to the reproduction by the business in question's search engine, or also others? In terms of explicit consent it is highly likely that this is a consent for either Yahoo!'s or Google's search engine, and those alone.

2.4.1.2 Registering the URL with a search engine

The most explicit way of giving consent for the search engine to index one's site is to actively register the URL with it. For instance Google has a page of its own where people can do this.[133]

2.4.2 Implicit consent

2.4.2.1 Introduction

Implicit consent is crucial for the society to function as it does today. A person will accept terms and regulations every day through implicit consent. For instance when someone enters a bus and pays for the ticket. By doing this the bus company's terms of service are accepted. Or when a person enters a restaurant, sits down and orders something from the menu. This action will indicate that the price for the meal is accepted, as well as an agreement to pay for it. If

132 http://www.googlepages.com [22.03.2007].

133 http://www.google.com/addurl/?continue=/addurl [22.03.2007].

every offer/accept had to be more formalistic in nature (either oral or in writing), everyday life would be a lot more tiresome and a lot less efficient.

Implicit consent is not established in any formal legislation, but it is accepted as a valid form of consent by the courts. A Supreme Court decision from 2003 regarding illegally distilled spirits is an example of this. Here the Court decided that implicit consent was not given, but they none the less considered it, cf. Rt. 2003 p. 902 Sec. 12. The European Court of Justice has also considered implicit consent. In the case of *Zino Davidoff AS v A & G Imports Ltd and Levi Strauss & Co. and Others v Tesco Stores Ltd and Others*[134], regarding whether the sale of trademarked items within the EEA had been implicitly consented to, ECJ found that it had not. But they still considered implicit consent as a valid form of consent. What these cases show is that although implicit consent is a valid form of consent, it is not easy to establish that such consent is given.

When considering how much shall be demanded before implicit consent is considered given, one must probably take notice of the nature of the situation. The cases tried before the courts are for the most part of a more serious nature, and it should and will be more difficult to establish implicit consent in those cases than in situations from everyday life.[135]

The conclusion drawn from the cases mentioned above are further supported when considering the practice of The International Court of Justice, which is *"(…) increasingly rigorous in ascertaining the existence of consent, especially when the consent is alleged to have been given implicitly (…)"*.[136]

In the following I will analyze whether implicit consent can be considered given in the relation considered in this thesis.

2.4.2.2 The nature of the Internet

The Internet cannot exist in its present form without search engines. They are equally as important as hyperlinks, which also are a requirement for the Internet's continued existence.[137] The importance the Internet has for society as we know it today cannot be emphasized enough, at least for the western part of the world. This would mean that perhaps the presumption of the existence of implicit consent is stronger in this field than in others.

134 Joined cases C-414/99 to C-416/99.

135 As for instance in the example with the restaurant.

136 Leiden Journal of International Law, 16 (2003), pp. 701-713. *http://journals.cambridge.org/download.php?file=%2FLJL%2FLJL16_04%2FS0922156503001390a.pdf&code=a63c20fc11d3c28d91802b67ab63f032* [15.03.07].

137 (Udsen, et al., 2006), p. 1. "Links are the life nerve of the Internet". ["Links er Internettets livsnerve"]

2.4.2.3 Can consent be considered given implicitly?

The Internet is a public place. People use the Internet for a vast amount of purposes, from attaining the strictly personal goals, to solutions for major multi-billion dollar businesses. There are a vast amount of users of the Internet, and it has been designed to ease communication and cooperation between users.[138] Keeping something private on the Internet is in many ways a contradiction in terms. If one wants to ensure that information is not being accessed, exploited or misused, one should keep it off the Internet. However, this cannot be said to be a valid argument anymore. The Internet, with its increased importance in the community must be accessible for everyone to make use of. More and more companies build their business around Internet-solutions. For instance some banks are completely based on the Internet and do not offer any "physical" bank branches; for example Skandiabanken[139]. But also the use of the Internet on a smaller scale must nowadays be said to be a *right*. It's a way of expressing ideas, concepts, projects and material that is unique, from a historical point of view. A clear-cut "accept or decline" attitude is therefore unfortunate at best.[140]

Another argument stated in favor of more or less free exploitation of works not explicitly protected is that most people are concerned with how to rank as high as possible on the search-hit list[141], and not how to avoid the search engines. This is probably true; one does not have to look any further than to all the companies that works solely on helping others ascending the search-hit ranks. However, is it so that even if a majority feels it is okay to relinquish some rights, then the minority must follow? Or, turned around, should a minority dictate what's acceptable when the majority of the people want the opposite? It is found in several other areas, both legal and in general society issues, that the majority will suppress the minority's interest in order to preserve their own interests. And it is also found that minorities' interests will overturn the majorities' interest. In order to reach a conclusion one must probably ponder the question in each concrete case, and therefore there is not a general answer to this.

What seems to be the most popular and lasting argument for why consent must be considered given is that it is so easy to stop the web crawlers from visiting one's web page, and therefore in turn preventing indexing. This can be

138 Originally a military project, and then to ease cooperation between Universitys.

139 http://www.skandiabanken.no [22.03.07].

140 This would perhaps be different if the Internet was in its early beginning; only a few people used it and it was not really all that important. But this cannot be said to be the case now.

141 Danny Sullivan mentions this in his blog: http://blog.searchenginewatch.com/blog/051021-113341. However, he adds that whether the indexing is legal still has not been tested.

done by adding an exclusion protocol to the web page. The exclusion protocol will be recognized by a crawler as a message that this page is not to be indexed. The most common way of doing this is with the use of the robot.txt exclusion protocol.[142]

The robot.txt protocol can be communicated to the web crawlers in two ways. The first is through html[143]. The rightholder of a web page needs to create a page called *robot.txt*, and place it as a separate page on the top level. For example http://www.news.com/robot.txt. This page cannot be placed anywhere else, for instance as a sub-page: http://www.news.com/miscellaneous/robot.txt. If it's placed here then the crawler will take no notice. The same is the case if the page does not contain the correct codes; it's either empty or the codes are wrong. An example of a robot.txt file that will keep every robot (that heeds the exclusion protocol) is this:[144]

```
# go away
User-agent: *
Disallow: /
```

If one rents space, i.e. does not own a domain of one's own, it can be problematic to place the robot.txt file on the top domain. Only the server administer can do this, and they might not be willing to help. Another way of solving it is then to use Meta-tags[145] in the header. The header is where a rightholder of a web page will write what the title is and so on. An example of how this can be done is: <META NAME=»ROBOTS» CONTENT=»NOINDEX, NOFOLLOW»>.[146] This command would notify the crawlers to neither index the page nor follow the links. The problem is that only very few crawlers understand Meta-tags, so this will not work as well as with the html-version.

The robot.txt is claimed to be an easy way of avoiding crawlers. But this is for the most part claimed by people with some, or extensive, background in

142 http://www.robotstxt.org/wc/exclusion.html [21.03.07].

143 HTML = hypertext markup language. It is the language in which a web page is written, and contains pre-defined standard codes that most browsers will understand. For instance if one wants to have a line break after a sentence, the following code is inserted: <*BR*> http://www.web-source.net/html_codes_chart.htm [22.03.2007].

144 http://www.robotstxt.org/wc/norobots.html [22.03.07].

145 Meta tags is a way of including information in the header that a human user might not see, but the computers used will see and understand them. For instance which character set to use, or how to get a logo to appear next to the URL in the browser's address field. http://searchenginewatch.com/showPage.html?page=2167931 [21.03.07].

146 http://www.robotstxt.org/wc/exclusion.html#meta [21.03.07].

computer technology.[147] By accepting that a rightholder must take these measures to protect his or her work, do we then accept that only people with a certain minimum of technological knowledge shall be granted protection of their works?

One of the most important regulations in the Berne Convention is Art. 5 (2); which provides that it cannot be a condition for protection that works fulfill a certain formality.[148] By accepting robot.txt as an adequate solution one introduces an *opt-out* principle[149], when in the rest of the copyright field the *opt-in* principle is the valid one. Article 5 (2) is, alongside the principle of national treatment, one of the reasons why the Berne Convention has been so successful.[150]

So why should the Internet and the search engines be treated any differently? Perhaps because of the immense interest society has in keeping the Internet up and running. And if the search engines' have to ask for explicit permission every time they want to index a new page, then the number of indexed pages would suffer a marked drop. If this happens then the Internet would not exist as we know it now. A valid objection to this consideration is if society's interests should be allowed to overturn or limit the interests of a private party. However, this has happened in other aspects of the copyright field. The society has an interest in exploiting created works, so that it may advance in respect of technology, art and culture. This is for instance the background for the rules stating that private use of protected works is not illegal. The Norwegian Copyright Act also has a regulation which establishes that the holder of the copyright for a work cannot oppose someone using this work in order to create a new and independent one, cf. Sec. 4. However, these rules are based on the consideration of encouraging new creations, and the access for private, non-commercial use. The search engines don't really fit into any of these considerations. This is because they are highly commercial, and their function is not really to encourage new creations. They function more to enable communications on the Internet through allowing people to easily find relevant information. Although there are also other exceptions and limitations to the exclusive right of authorizing reproduction, and that is in the case of

147 Gisle Hannemyr comments the Belgium ruling against Google, in his blog: "*As always*, there is no mentioning in the ruling of the fact that the robots.txt (Robots Exclusion Standard) allows publishers to opt-out from being indexed by search engines" (italicizing done by me), http://heim.ifi.uio.no/~gisle/blog/?m=200702.

148 This is an Article that every member of the Union must heed, and therefore it is quite established in the copyright legislation of different countries around the world.

149 One would have to "opt-out" of making one's works available for reproduction, instead of giving explicit permission for reproduction (which would be to "opt-in").

150 (Rognstad, 2004), p. 4.

temporary reproduction, which I will analyze in chapter 3. The consideration behind allowing temporary reproductions is to enable a more efficient mode of communication on the Internet, sparing band-width and time. But there is a very limited access to this limitation. And while the indexing process definitely does not fulfill the requirements, it probably does not comply with the considerations behind the regulations either.

There are now several programs on the market that helps people create their own web pages. This lowers the requirement of technological knowledge even further in regard to who can create a web page. And the use of robot.txt is difficult for an increased amount of people. For some it might even be close to impossible, because the web editing program used does not allow for the commands needed to be written in a correct way. If a program offers the possibility of such commands, then it would probably be more in the lines of a menu-choice offering "create robot.txt".

Many web editing programs allow for the text to be written, and pictures and links to be inserted into the page. They will not allow for direct html-coding to take place. If a person tries to create a page called robot.txt and only writes the codes above, then this would probably not work. This is because the text defining the page as a robot.txt exclusion protocol is html-codes themselves. And if one tries to insert these through the text function in a web editing program, the program would automatically insert additional html-codes in from of what's written, defining this as a text (with the chosen character size, type, color of the text and so on) in the body of the page. For instance an extract could be: <h1>This is what I have written</h1>.[151] If these codes are written on each side of the code needed for the robot.txt protocol, in addition to other codes that will appear, then the crawler would *probably* not recognize it as valid.

This would mean that if the opt-out principle is accepted, then people with limited technological skills would either have to accept lack of protection, or relinquish the possibilities offered by the Internet. On the other hand, perhaps the possibilities offered by the Internet more than compensate for the lack of control one has with the material published.

Moreover, the robot.txt exclusion protocol is not mandatory to follow. Most search engines of a certain size will follow it because they rely on a reputation as upstanding and loyal. But what of those that does not? Must consent still be considered as given, because one has after all published the material on the Internet? In order to shut these out, the administrator of a web page must actively exclude requests from certain IP-addresses. Is this something that eve-

151 H+a number defines the size of the letters. The scale goes from 1-6, where 1 is the largest size.

ryone which makes use of the Internet must expect and accept? If it's accepted that certain technological measures must be taken if one wants to protect a web page from search engines, and others, where do you draw the line? What is too complicated, and therefore unreasonable to demand?

And what of those that does not know of robot.txt? With the extensive amount of users on the Internet and the relatively unknown existence of robot.txt, many will not know of this possibility. So they might write something on their page along the lines of "This page is copyrighted, and shall not be indexed". Is this enough of an effort? They have clearly stated that they do not consent to indexing. And human users will have to follow it, but what of search engines? As pointed out of Sergey Brin and Lawrence Page, the web crawlers will not recognize this.[152] They do not have the human ability to understand words in context. They will see the above sentence as nine words, and gather them for indexing along with the rest of the words on the page.

Conclusion

The objections against implicit consent in this regard are many and valid. However, in this author's opinion the considerations in regard to the importance Internet, and search engines with it, have to society is probably stronger. After all, many of the limitations to the apparent exclusive right of reproduction that authors have are made in consideration to society's best interest. And, in a way one can say that even the exclusive right given to rightholders also is a product of society's interests. The right of reproduction is mainly a right which ensures economic benefits for the rightholder. And if this right weren't granted, no one would bother creating anything. This of course is because rightholders probably would have difficulties making money from their works. And if they can't make a living off it, they will stop creating. And a situation like that would definitely not benefit the society.

However, the indexing done by search engines will probably *benefit* the rightholders economically, and not cause them any losses. Therefore the conclusion must be that by publishing something on the Internet, one has given implicit consent to the institutions that are fundamental for the existence of the Internet; i.e. people can link to the contents, and search engines may index the contents. Protective measures like an exclusion protocol in the form of for instance robot.txt must in this author's opinion therefore be considered as a *withdrawal* of the consent already given. And because of the nature of the consent given, there must be an opening to withdraw consent; unlike other legal

152 http://infolab.stanford.edu/~backrub/google.html - 4.3 -crawling the web. [21.03.07].

areas where a consent in most cases are binding and cannot be withdrawn by one of the parties alone.

In the cases where a page is published with a robot.txt exclusion protocol, it might be viewed as a paradox that consent is given and withdrawn at the exact same time. However, this would be the most reasonable and logical way to view this situation. Because if everyone is considered as to have given consent the moment they publish a page on the Internet, and countermeasures as for instance robot.txt is seen as a withdrawal of consent, then the concerns relating to the rightholders technological level would be lessened. This is because then the line is not drawn between who has given consent and who has not, but between the ones that withdraws their consent and the ones that do not. Always viewing consent as given the moment a page is published would also help emphasize that if someone wants to make use of the Internet, then he or she has to accept some of the services and methods that make the Internet work. But to avoid this being too unreasonable, the access to withdraw consent in some relations, as with search engines, is granted without reservation.

The consent withdrawn in relation to search engines must be considered as a withdrawal with retroactive effect. This means that when a search engine for instance discovers that an already indexed page now has a robot.txt exclusion protocol, then it should delete it from its index, and not only heed the exclusion by not updating the entry in the index.

Implicit consent as legal grounds are however somewhat shaky, and is not really a satisfactory solution. The only reasonable and sensible course of action is to establish new International regulations that allow the search engines' conduct; at least their conduct on such a basic level as indexing really is.

3 IS THE SEARCH ENGINES' TEMPORARY STORAGE OF PAGES A REPRODUCTION?

3.1 What is allowed and what is not; a look at the cache

When referring to storage in a computer-related context, most will think of the user's storage of files on a hard drive. However, in addition to this there exists something called *temporary* storage, which is equally important for the functioning of computers locally and networks in general. The temporary storage is by technical terms often called a *buffer* or a *cache*, depending on which purpose it serves. For instance it is called a buffer in relation to TCP/IP, and a cache in relation to web browsers or search engines. The focus of this thesis is of the search engines, but in order to get the complete picture of the nuances that decides whether a form of temporary storage is legal or not, I must explain several other technological aspects in addition to the search engines' cache. This is because the label "cache" is not a uniform term; it does in fact represent several very different technological solutions. The only common feature is in fact that it is a type of (presumed) temporary storage of information.

TCP/IP stands for *Transmission Control Protocol (TCP)* and *Internet Protocol (IP)*.[153] This is the protocols most commonly used to establish and manage connections between computers[154], so that information can be exchanged on the Internet.[155] The IP is a protocol which provides addresses for the different computers on the Internet, so that the source computer knows where to address the data. The TCP is a protocol that makes sure that the data that arrives are the same that was delivered. This is important, because the when the data is sent from location A it does not "travel" as a single file, but is divided up into several small *packets*[156] which doesn't necessarily all take the same route to location B. The TCP makes sure that the packets are put together in the same way as it was when it was sent, and that all the packets

153 http://en.wikipedia.org/wiki/TCP_IP [09.03.07].

154 http://www.searchandgo.com/articles/internet/internet-practice-4.php [09.03.07].

155 The original protocol specification from 1981 is still relevant today. It's defined in document RFC793 which can be found here: http://tools.ietf.org/html/rfc793 [09.03.07].

156 The data sent will be distributed between the packets, although each packet will contain the same information in its header for identifying the data, such as proper protocols, the originating address, the destination, and the packet number. http://computer.howstuffworks.com/question525.htm [09.03.07].

arrives. If one or more packets are delayed or held up, the protocol will request a resend of that packet from the originating computer.

When the data is sent between the different end hosts (sending and receiving computer) there might be a delay. If the output from the receiving computer is slower than the input to that computer, there will be a stop in the dataflow. In order to prevent the data from being lost the TCP makes use of a buffer. This is a designated area on the end hosts (in this case the receiving computer, for instance a *router*[157]) which enables data to be temporarily stored until it can be sent forward. The data in this buffer is substituted continuously. When a packet is sent forward, it is deleted from the buffer and a new packet enters. The size of the buffer is what defines the *receive window*.[158] This is a form of temporary storage. The purpose of this buffer is to ensure that the data flow is even and accurate.

Since the data is fixed in a storage device, even for a brief moment, it must be said to constitute a reproduction, cf. chapter 2.2.2 regarding reproduction in relation to indexing.[159] However, while indexing will mainly serve the same purposes, and function in the same way no matter in which context it is used, this is not the case with temporary storage.

A very different type of temporary storage is the ones maintained by the *proxy*[160]. The proxy is most commonly used in relation to servers ("proxy servers"), or web browsers.[161] This type temporary storage is called the *cache*.

A proxy server will in most cases serve two purposes.[162] Firstly it is a security measurement, because it is "placed" between the internal network a desktop computer is connected to, and the Internet. When a user request for instance a web page, the proxy server will apply it's already configured filters[163] to that web page in order to decide if it shall be let through or not. Secondly, it will improve transfer speeds. This is because it creates a local copy –in the cache – of web pages visited. The University of Tromsø's internal network can be used as a good example. There are possibly several thousand users connected at the same

157 "A router is a computer networking device that buffers and forwards data packets across an internetwork toward their destinations, through a process known as routing" http://en.wikipedia.org/wiki/Router [09.03.07].

158 http://itpro.no/art/10007_side2.htm [09.03.07].

159 The Berne Convention art. 9 (1), the Copyright Directive Art. 2 and The Norwegian Copyright Act § 2, regarding the right of reproduction.

160 The proxy is a process used to for instance maintain the cache. (Tanenbaum, 2003), p. 657

161 (Tanenbaum, 2003), p. 658.

162 http://www.privatenavigator.com/GetAnonymous/faq.asp [09.03.07].

163 Such filters can be that certain pre-defined web pages are disallowed, or it can establish that web pages containing certain words shall not be allowed. This is to prevent users from accessing for instance illegal or unwanted material, and prohibiting them to access pages that possibly could be a treat to the internal network connected to the proxy server.

time, either directly to the network or through *vpn-clients*[164]. If all were to visit for instance http://www.dagbladet.no at the same time[165], this would possibly overload to the server hosting this site. What the cache-function in the proxy server does is that when the first user has accessed the page in question it creates a local copy of the page –in the cache. When a user at a later time requests to access that page, the proxy server will first check to see if the page is stored in the cache. If it is, then it's the local copy that's returned to the user. If it's not, then the proxy will request the page from the server and after it has stored the page in the cache for future use, it will return it to the user that has requested it.[166]

The main question in regard to the proxy is how long the pages should be stored in the cache. One way of doing it is to set a timer on how long a page is to be stored in the cache before getting evicted.[167] There are drawbacks with both setting the timer too long and too short. If the timer is set too long, then the copy might become *stale* and the user gets an outdated page returned, containing data that's obsolete. If the pages are thrown out too quickly the cache will not be very effective because it will contain few pages and would have to contact the originating server more often. Another way of solving this problem is to use the RFC 2616's special features that deal with cache management.[168] The proxy will then send out a minor header to the server, for instance the *If-modified since* request header. This will specify which page the proxy wants and when the page in the cache was last modified. If the page has been modified since then the server will send back the new page, if it has not then the server will send back a short *Not Modified* message.[169] This will require the proxy to contact the web page's server, but requires much less bandwidth than if the server were to return the entire page every time a user wants to access it. This is very useful for the more popular sites like http://www.dagbladet.no or http://www.nytimes.com which has a great deal of users every day. These two methods can also be combined, where a timer is set to decide with which intervals the proxy is to send the *If-modified since* request to the server in question. This means that the proxy will not have to send the request every time a user requests to access the page, and also lessens the risk of the cache containing stale data.[170]

164 Vitual Private Network. http://en.wikipedia.org/wiki/Vpn [09.03.07].
165 Which is a very likely scenario, especially in the morning when everybody arrives and wants to read the news before starting to work.
166 (Tanenbaum, 2003), p. 657.
167 (Tanenbaum, 2003), p. 658.
168 RFC 2616, sec. 14.9 "cache control". The RFC 2616 is a document for the specification of the *Hypertext Transfer Protocol (HTTP)*. http://rfc.net/rfc2616.html
169 (Tanenbaum, 2003), p. 659.
170 (Tanenbaum, 2003), p. 659

The proxy server's cache is a more permanent type of cache; the data stored here will not be replaced or deleted until the pre-defined algorithms for updating the cache is evoked. It may very well be that a web page exists for as long as for instance a year, if the page is of such nature that it is not updated very often ((Tanenbaum, 2003):659 uses a web page listing the gods from the Greek and Roman Mythology as an example of a stable page).[171] The stored copies of the web pages can also be copied to a hard drive and permanently stored there as files, completely accessible long after the cache is updated. This is in stark contrast to the buffer created by TCP/IP where each packet is almost worthless on its own, and is continuously updated as information is sent.

The cache created by a web browser like Microsoft Internet Explorer or Mozilla Firefox is in most regards similar to the proxy cache, only that it is stored locally on the desktop computer and is only accessible to the user of that computer and not an entire network.

The parameters governing the proxy in for instance Internet Explorer can be defined more closely by the user.[172] The user can also get easy access to the list over data cached by Internet Explorer. Here the user gets to choose how, and how often, the cache is to be updated. Internet Explorer v. 7.05 provides the user with four choices as to when it shall check for newer versions of stored pages: 1) *Every time I visit the webpage,* 2) *Every time I start Internet Explorer,* 3) *Automatically* 4) *Never.* The default setting for Internet Explorer is *automatically*[173], but as shown here a cached web page can in theory be stored locally on the computer for as long as the current installation of the operating system on that computer exists.[174]

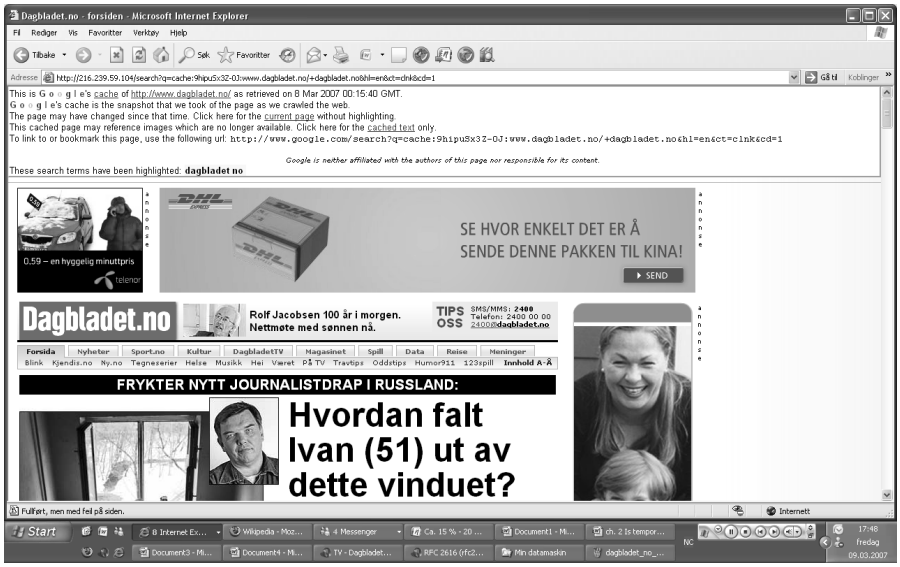
171 The proxy does not recognize the contents as a stable nature, but checks the *Last-Modified* header to see how long its been since the page was updated. If it has not been updated for instance the last year, this is a sign that this is a very stable page.

172 This access is gained by entering *Control Panel* and accessing *Internet Options*. Under the *General* tab the *Settings* for *Temporary Internet-files* is found under *Browsing History*. Clicking on *Settings* will grant access. Common users of a proxy server will not be able to decide the update-frequency of the proxy.

173 Even if it is set to update automatically, stale pages may appear. Users may experience this from time to time – an outdated version of the web page appears when requesting access, and manually requesting the proxy to access the originating server by use of a predefined key; most often F5, will cause the page to update.

174 A new installation of an operating system will in most cases require a formatting of the hard drives (thus deleting all data), at least in the case of Microsoft Windows.

The caches created by the search engines[175] are different in several ways from the caches mentioned above. The first being that search engine does not create a local copy, but a copy on a server that all its users have access to, thus making it different from the web browser cache and more like the proxy server's cache. Furthermore, the search engines cache will not on the whole either evict the page from the cache after a certain amount of time, as the proxy server might do, or check for updates when the user access it again, as the web browser's cache might do. The cache the search engine has created will only be updated when the web crawler visits that web page again. The page the user then gets access to will be located on the search engines own server, with its own URL[176], unlike the cache on both the proxy server and the web browser, which operates without the users even knowing about the process. In the following there is an example of Google's cached copy of www.dagbladet.no, and the original page access directly through the web browser on my desktop computer (in the above order). Both pages were accessed at exactly 17:48 on the 9th of March 2007.



175 The creation of a (public) cache is not done by all search engines; in fact most do not. The largest search engines that offer the possibility of accessing a cached version of a web page are Yahoo!, Google and Windows Live Search.

176 Not like the proxy, which most users does not know exist because they see the web pages' own URL in the address field of the browser, for instance http://www.dagbladet.no. Access to www.dagbladet.no through the search engines' cache will generate an URL that looks something like this: http://66.102.9.104/search?q=cache:9hipuSx3Z-0J:www.dagbladet.no/+dagbladet.no&hl=no&ct=clnk&cd=1&gl=no

This means that while some web sites, like the ones of for instance news sites of some size might be visited frequently by the crawler, and thus gets the cache updated with equal speed. Others might be crawled less often, enabling users to access potentially outdated information. This would include information meant to be removed; either because it was moved to an area only accessible for paying users, or because it was wrongful and not meant to be accessed at all.

If a user clicks on a link in the cached copy, the link will lead to the page's location on the original server. So if this is taken offline, an error message will appear stating that the page cannot be found. However, if a user experiences this, all that has to be done is to search for the complete headline, preferably by using inverted commas, and then accessing the cached copy of it. Thus the main page of the original source is evaded all together. This has already been claimed as an infringement on the exclusive right of reproduction by certain right holders. While they might to some extent be prepared to accept indexing, they have a harder time accepting the cache. This is especially the meaning of different news corporations'. This opinion has been expressed for example by "Mediebedriftenes landsforening", a Norwegian organization for media corporations. Via their representative Pernille Børset the following was said to www.digi.no (quotation translated by me): "*Articles and other material are stored in full by Google, and if you click on the cache you won't come to the source, but to the archive of the search engine, no matter if the original article is removed from the net, or put in closed archives, says Børset. In our opinion,*

this clearly violates the rules of the Norwegian Copyright Act, and we will raise the issue with the search engines in question"[177].

It is worth mentioning in this respect that while the "TCP/IP-buffer" and the cache related to the proxy server and web browser are user-defined, the search engines' cache is not. Users have to access pages or otherwise transmit data for it to be stored in the buffer and cache, while the search engines decide for themselves if they wish to cache pages and/or publish these copies.[178]

3.1.1 Exceptions and limitations to the right of reproduction

There can be no doubt as to if the copy created in the cache is a reproduction of the original web page, cf. the analysis of "reproduction" in ch. 2 above. The question that arises is therefore if any of these methods of temporary storage can be regarded as exempt to the exclusive right of reproduction. The discussion if certain acts of temporary reproduction should be exempted to this right began with the WIPO Copyright Treaty.[179] This debate started much later than the Berne Convention was passed, and because of this there is no legal authority in the Convention that may include these particular exceptions. For this reason I will limit this analyze to only two of the legal sources – The Copyright Directive and the Norwegian Copyright Act. However, each occurrence that might be exempt from the exclusive right must also be subjected to the *three-step test*[180]. This "test" originated with the Berne Convention, and is found in Art. 9 (2). The reproduction must only happen in "(...) ***certain special cases***, *provided that such reproduction does* ***not conflict with a normal exploitation of the work*** *and does not* ***unreasonably prejudice the legitimate interests of the author***"[181]. In this way the Berne Convention has an influence also in this area. The Directive states it explicitly in Art. 5 (5)[182], while it in regard to The

177 NO: "Artikler og annet stoff lagres i fulltekst hos Google, og trykker man på hurtigbufferen kommer
man ikke til kilden, men til søketjenestens arkiv, uansett om den opprinnelige artikkelen er fjernet fra nettet, eller lagt i lukkede arkiv, sier Børset. Dette er etter vår mening klart i strid med åndsverkslovens regler, og vi kommer til å ta det opp med de søketjenestene det gjelder."
http://www.digi.no/juss+samfunn/avisene+vil+drøfte+søk+og+opphavsrett/art289622.html [09.03.07].

178 A web page crawled and indexed by a search engine might very well never be accessed, but still the cached copy is created.

179 (Sterling, 2003), p. 712.

180 View ch. 1.2.1.1.

181 Bold types created by me.

182 Explicitly in the meaning of that the different criterions are stated in Art. 5 (5). The Directive does not say that it follows the "three-step" test, but the criterions listed are equivalent to those in the "three-step" test.

Norwegian Copyright Act is found in the legal history, cf. Ot.prp.nr. 46 (2004-2005) page 12.

The Directive states the exception for temporary storage from the right of reproduction in Art. 5 (1). This exception is moreover the only one that is mandatory for the member states, i.e. while the rest of the exceptions functions as an exclusive list over which exceptions the member states are *allowed* to implement in their own legislation, Art. 5 (1) *must* be implemented. This follows of the wording in Art. 5 (1); *"Temporary acts of reproduction (...) which are (...) part of a technological process (...)* ***shall be exempted*** *(...)"*. This stands in contrast to the wording in Art. 5 (2) *that "Member states* ***may provide*** *for exceptions or limitations (...) in the following cases (...)"*[183]. Most European Countries will have such an exception in their national laws since the deadline for implementing the Directive was 22nd of December 2002, cf. Art. 13 (1).

Article 5 (1) allows for the temporary acts of reproduction provided that they are an "(...) *integral and essential part of a technological process (...)"* as well as being *transient* or *incidental*. Furthermore, these acts must have no independent economic significance, and their sole purpose must be to enable either *"a transmission in a network between third parties by an intermediary"* or *"a lawful use"* of a creation subject to the right of reproduction under Art. 2.

In relation to the "three-step" test it can be claimed that when all of the criteria in Art. 5 (1) is fulfilled, the criteria of the "three-step"- test is fulfilled as well. As long as the reproduction is temporary or incidental it can be said to fulfil the requirement of being a *special case*. And all the while it is an integral and essential part of a technological process it cannot be said to be *in conflict with a normal exploitation of the work*. Following that the most important consideration behind the exclusive right of reproduction is the authors' economic interest; it is highly likely that the demand that the act has *no independent economic significance* is enough for it to be considered as *not unreasonably prejudicing the legitimate interests of the author.* Therefore the "three-step" test must not necessarily be considered in addition to the Directive Art. 5 (1).

To illustrate the exception I will consider the buffer dedicated to the TCP/IP-process up against the criterions in Art. 5 (1). The buffer is necessary in order to keep packets from being lost, and from creating "bottlenecks" at the end hosts. It is dedicated to this process, and is because of this an integral and essential part of the function of the TCP-protocol. The buffer contains packets that have no independent value; they cannot be made sense of separately, and is therefore without independent economic significance. The data in the buffer is continuously replaced with new, different data, and is therefore highly tran-

183 Bold types created by me.

sient. The buffer is moreover not always in use, if the input of data in the end host is no greater than the output then the buffer will not be in use. The use of the buffer is because of this dependant on the amount of traffic, and can therefore also be said to be incidental to some extent. The transmission of data, for instance a web page, will happen between the server and the web browser. The transmission itself will be provided for by the *Internet Service Provider (ISP)* with the use of the TCP/IP protocol. It is therefore a transmission between third parties by an *intermediary*.

Because of this it must be concluded that the buffer in connection to TCP/IP fulfil the requirement of Art. 5 (1), and is as such exempted from Art. 2.

Section 11a (1) in the Norwegian Copyright Act is equivalent to Art. 5 (1). The requirements are listed differently, but otherwise directly translated, and because of this it follows that the buffer created to support TCP/IP, must be considered to fall within the limitation to the exclusive right of reproduction, provided for in Sec. 2. In regard to the "three-step" test, the same conclusion as above must apply here as well.

Considering the cache maintained by proxy servers and browsers, the situation might be different. First of all, the process is not necessarily transient. In most cases it will be, especially regarding the cache in proxy servers, but it is also quite easy to create permanent copies with the use of the cache, as shown earlier in the case of the web browser's cache. Since the requirement of the act being *transient or incidental* cannot be said to apply with certainty to proxy managed caches, the preliminary conclusion must be that the exception does not apply for either the proxy server's cache or the web browser's cache. However, recital 33 in the preamble to the Directive states that the *"(...) exception should include acts which enable* ***browsing*** *as well as act of* ***caching*** *to take place including those which enable transmission systems to function efficiently,* ***provided that the intermediary does not modify the information and does not interfere with the lawful use of technology*** *(...)"*[184]. As the cache, both in relation to proxy servers and web browsers, stores a complete and unaltered copy of the web page, it could seem that it was meant to be exempt from the exclusive right of reproduction after all. However, the recital also has a criterion that these acts of caching or browsing only is exempt if other requirements are met, namely the ones in Art. 5 (1). Thus the cache in relation to the proxy server and web browser does not fulfill the criterions of the exception, and must be considered to be in violation of Art. 2.

There is however another rule that allows exceptions and limitations to be applied to the otherwise exclusive right of reproduction; namely if it is done for private use.

184 Bold types created by me.

The right to private copy is stated in the Directive's Art. 5 (2) (b). It follows from this Article that member states may *"(…) provide for exceptions or limitations to the reproduction right (…) in respect of reproductions on any medium made by a natural person for private use (…)"*. The rightholders must however receive fair compensation. As this is an optional rule for the member states, it is not certain that many European states will have implemented this rule; though it is highly likely that they will have. Since Norway had to implement this directive, I will now look at how this was solved in the national law.

The Norwegian Copyright Act limits the rightholder's right of reproduction in regard to the reproduction for private use, cf. Sec. 12. The requirement of fair compensation to the rightholders is solved by granting means through the state budget on a year to year basis.[185] The reproduction must not happen for commercial purposes, which is also in accordance with the Directive, art. 5 (2) (b). The cache in relation to web browsers cannot be said to happen for commercial purposes. Section 12 establishes certain exceptions to the right of reproduction for private use, cf. Sec. 12 (2)-(4). However, none of these can be said to apply for the process that is caching in the above mentioned contexts. The "three-step" test should also be applied here, but the legal history of the law explicitly states that browsing is exempt from the exclusive right of reproduction when conducted for private use, cf. Ot.prp.nr. 46 (2004-2005), page 17.

The conclusion is therefore that caching in relation to web browsers is exempt from the exclusive right of reproduction; not through the regulations that governs temporary reproductions, but through the rules of private use. The question is then if the proxy servers can be said to constitute private use. The Directive solves this by stating that the exceptions or limitations only apply for *natural persons*, cf. Art. 5 (2) (b)[186], thus ruling out proxy servers since there are few, if any, natural persons that use a proxy server for their own local network in their homes. The Norwegian Copyright Act leaves the question more open though. It only requires the use to be *private*. While the cache used by the proxy server at the University of Tromsø cannot be considered as created for private use, which could easily be the case for any private company; for instance a law firm. While the use of the word 'private' in this context naturally would mean natural persons' private use, the text does not give a clear answer. The legal history of Sec. 12 (1) does not provide an answer; apart

185 NO: "Opphavsmennene gis en rimelig kompensasjon gjennom årlige bevilgninger over statsbudsjettet.", cf. Sec. 12 (1).

186 This is however not mandatory for the member states to implement, so many European Countries might not have this limitation in their legislation.

from that it states that the law has been revised in order to comply with the Copyright Directive. This would further support the conclusion that *private use* means *natural persons* private use. However, it is not always clear that directives are implemented correctly, cf. Rt. 2000 p. 1811 (the Finanger-case).

A Supreme Court-ruling from 1991 states on page 1299 that *"copying for private use in accordance with Sec. 11 is not limited to his or hers* ***own personal use****, but can also take place to the benefit of family or friends (…)"*[187], cf. Rt. 1991 p. 1296.[188] This must be understood as to the fact that private use is limited to *natural persons*; although both the person doing the actual reproduction and his or hers family and friends. Private use cannot be understood as to apply for private companies as well, or to natural persons if the reproduction occurs in relation to their work, cf. an understanding of the 1991-ruling.

This means that while the process of caching, and the purpose with it, is the same for both the proxy servers and the web browsers, only the browsers are exempt from the rightholders' exclusive right of reproduction.

The cache created by search engines cannot, like the cache created by web browsers and proxy servers, be said to meet the requirements of the Directive's Art. 5 (1) or the Norwegian Copyright Act, Sec. 11 (a). It is not created for private use, and might very well have a commercial purpose. It can therefore not be deemed to fulfill the requirements stated by either the Directive's Art. 5 (2) (b), or Sec. 12 (1) in the Norwegian Copyright Act.

3.1.2 Consent

The last possible way for search engines to gain legal grounds for their caches, are through consent; either explicit or implicit.

Above it was concluded that the rightholders have consented implicitly to the indexing done by the search engines. The defining point for this conclusion was the Internet is vital for today's western society, and the search engines are vital for the Internet's existence, while indexing is vital to search engines. Thus society's interests must take precedence over the rightholders' interests. Since the search engines that cache web pages will do so with every page indexed, one might reach the conclusion that if one has consented implicitly to indexing this will include caching as well. After all it is fully possible to prevent the search engines from publicly presenting a cached page; the most common way

187 NO: "Kopiering til privat bruk etter § 11 er ikke begrenset til egen personlig bruk, men kan også skje til fordel for familie- eller vennekretsen (...)". Today's Sec. 12 (1) is equal to what before was Sec. 11, cf. Ot.prp.nr.15 (1994-1995) page 109-110.

188 Translation and bold types done by me.

is with the use of Meta tags or the robot.txt exclusion protocol. For instance Google has a page where the process is explained.[189]

One situation where consent to the cache may be considered as explicitly given is where the rightholder have consented to terms stated by a web hotel, as mentioned above in 2.4.1.1. The other method of explicitly consenting to indexing was to actively registering the page's URL with a search engine. If a rightholder does this without prevent the cache with the methods explained by the search engines themselves it might be viewed as consent for caching. However, this cannot in this authors opinion be viewed as an explicit consent, but instead as a form of implicit consent.

In regard to the implicit form of consent it must be noted that society has a stronger interest in the indexes than in the caches. Making these caches available for the public, i.e. the users, is not vital or even necessary for the operation of the search engines.

If search engines create the cache, but do not make it available, it will technically be a copyright infringement, but it does no real harm. And if they need the caches to keep the indexing process as good as possible, it might be argued that un-publicized caching is consented to, since consent is considered given to the indexing of the pages.

For the common user, and the society as a whole, the cache is no more than a useful function. Considering the possible economic ramifications it might have for the rightholders combined with the rather little importance the search engines' caches has for society, implicit consent cannot be considered given for the cache as well. The caching done by search engines must therefore be considered as a copyright infringement.

3.1.3 What may the consequences for the search engines be?

The legislation provides for several possible sanctions in relation to copyright breaches. The Copyright Directive require that the national states bound by the Directive shall provide for *"(…) appropriate sanctions and remedies (…)"* which are *"(…) effective, proportionate and dissuasive"*, cf. Art. 8 (1). The Norwegian Copyright Act has fulfilled this in Sec. 54. Anyone who *willfully*[190] or *involuntarily*[191] breaks the Copyright Act may be punished by fines or by prison up to three months, cf. Sec. 54 (1) (a-e).The caches created by the search engines are a violation of chapter one of the Copyright Act, and as such Sec.

189 http://www.google.no/intl/en/remove.html [26.04.07]. "Remove cached pages."
190 NO: "Forsett".
191 NO: "Uaktsomt".

54 (1) (a) is applicable if their actions are found to be involuntarily or willful. An argument might be made that large international search engines cannot be expected to know of purely national legislation concerning this. However, the legal history of the Act states that involuntary mistake of law will not exempt someone from punishment[192], cf. Ot.prp.no. 46 (2004-2005) page 158. Actual misapprehension may be judged differently, since it might in other criminal cases as stated by the Penal Code Sec. 42, cf. Ot.prp.[193] page 158. However, a search engine that systematically caches every page it indexes must know or understand that the probability of some of the information cached will be protected by copyright legislation. Even involuntarily actual misapprehension of the circumstances[194] cannot be ruled as likely in this case.

Moreover, if the infringement is willful *and* the matter is especially aggravating then the transgressor may be convicted to pay fines or serve up to three years in jail, cf. Sec. 54 (4)[195]. When considering if the matter is especially aggravating one shall especially take into account the damages done to the rightholders and others, as well as the gain the transgressor had and the scope of the transgression besides what's mentioned above, cf. Sec. 54 (4)[196].

These are regulations that may only be prosecuted by official authorities, cf. Sec. 54 (7).

Furthermore, the Copyright Directive requires that the national states takes steps to ensure that rightholders can *"(…) bring an action for damages and/or apply for an injunction and, where appropriate, for the seizure of infringing material as well as of devices, products or components (…)"*, cf. Art. 8 (2). The possibility for rightholders to bring such action as mentioned above is found in the Norwegian Copyright Act Sec. 55. This regulation provides that rightholders may seek compensation for damages as stated in Sec. 54. The rightholders may also be restituted if the action taken by the transgressor is willful or grievous negligence[197], cf. Sec. 55 (1).

However, there exists a possibility that the search engines may not be held liable, even if they are found guilty of violating the rules of the copyright legislation. As mentioned in chapter one, the e-commerce legislation contains exemptions from liability that might be applicable in this regard.

192 This is moreover the "common rule" in criminal law, cf. the Norwegian Penal Code Sec. 57.
193 Ot.prp.no.46 (2004-2005).
194 NO: "uaktsom faktisk villfarelse."
195 NO: "(…) foreligger særlig skjerpende forhold (…)"
196 NO: "Ved vurderingen av om særlig skjerpende forhold foreligger, skal det først og fremst legges vekt på den skade som er påført rettshavere og andre, den vinning lovovertrederen har hatt og omfanget av overtredelsen for øvrig"
197 NO: "grov uaktsomhet".

Because of this I will consider if these exemptions apply for the search engines which caches web pages before I consider the sanctions further. If the exemptions are found to apply then I will not consider the possible sanctions further.

3.2 Can the E-Commerce legislation prevent the search engines from being liable?

Electronic Commerce; EC Directive and Norwegian Law

In this chapter I will analyse two different legal sources together; The E-commerce Directive and the E-commerce Law. The reason for this is that the Norwegian E-commerce Act was passed as a direct result of the E-commerce Directive. Many things will be similar, and a comparison of them will emphasise the similarities and possible differences in a clearer way than a divided analyse would. In the following I will refer to the E-commerce Directive as the "Directive", and the Norwegian E-commerce Act as the "E-Act". When referring to the "E-commerce legislation", both these legal sources are included. They will only be mentioned with a more complete name when necessary in order to avoid confusion.

The E-commerce legislation provides for exemptions from liability for certain service to ensure the free movement of information society services in the EEA[198], cf. The Directive Art. 1, and the E-Act Sec. 1. In the following I will analyse whether these exemptions from liability applies for the cache provided by some search engines. The Directive does not regulate *when* a service provider might be held liable, only when it might *not*. The Articles 12-14 has a wording consistent with the interpretation that this is a set of minimum requirements; a service provider may never be held liable if the requirements of any of these Articles are fulfilled. This is also supported by Recital 10 which states that the "(...) measures provided for in this Directive are strictly limited to the minimum needed to achieve the objective of the proper functioning of the internal market (...)". However, the Directive does not contain any word-

198 The EC-directive is aimed to ensure the free movement among Member States, cf. Art. 1, while the Norwegian legislation expands this to the entire EEA, Sec. 1. But since the EEA decided to implement the Directive into the EEA-agreement, one can conclude that the Directive as a whole also applies for the entire EEA, cf. *the Decision of the EEA Joint Committee No. 91/2000 of 27 October 2000 amending Annex XI (telecommunication services) to the EEA-agreement*.

ing indicating that these are the only situations where limitations from liability can be applied. This means that the national states may grant exemptions from liability *also* to other service providers than the ones covered in Art. 12-14.

3.2.1 Does the E-commerce legislation cover search engines?

The legislation encompasses *information society services*, cf. Art. 1 (1) of the Directive, and Sec. 1 of the E-Act. The exemptions from liability provided by the legislation are for the service providers of such information society services, cf. The Directive's Art. 2 (b) cf. Art. 12-14, and the E-Act Sec. 3 cf. Sec. 16-18.

There can be no doubt that a search engine is a *service provider*, as defined in the Directive, Art. 2 (b) and in the E-Act, Sec. 3 (a). Nor is there any doubt connected to the fact that any user of a search engine is a *recipient of the service*, as defined in the Directive Art. 2 (d), cf. the E-Act, Sec. 3 b).

The first issue raised is therefore if a search engine's service is an *information society service*.

The Directive defines information society services with a referral to Art. 1 (2) of Directive 98/34/EC as amended by Directive 98/48/EC, cf. The Directive Art. 2 (a). Information society services shall include *"(...) any service* ***normally*** *provided for remuneration,* ***at a distance, by electronic means*** *and at the* ***individual request of a recipient of the service*** *(...)"*[199]. The E-Act has this exact wording as well, apart from that "*by electronic means"* is listed before *"at a distance"*, cf. Sec. 1 (2) (a).

The E-Act and Directive 98/48/EC require that a service is *"(...) normally provided for remuneration (...)"*. An understanding of this would be that the service will have to charge the recipients in some way, in order to fulfil this requirement. However, the requirement contains the wording *"normally"*, and this indicates that the legislation does not exclude every service that is *not* provided for remuneration. Recital 18 to the E-commerce Directive helps to clarify this; a service need not be paid for by the recipients, but must represent an economic activity. Search engine services are for the most part offered free of charge for the user[200], and get their income from advertising. They are therefore representing economic activity. Concerning the E-Act it is made clear through the legal history that also those that offer tools or services for searching the Internet are considered as covered by the expression *information society services,* cf. Ot.prp.no.31 (2002-2003) page 55, sec. 3. As a result search

199 Bold types created by me.

200 Search engines which limits some of their services to paying customers, are normally smaller in size and often concentrated around certain limited subjects. An example of this is Lovdata (http://www.lovdata.no). [22.03.07].

engines must be considered as fulfilling the requirement concerning remuneration, even if most of them provide their services free of charge for their users.

At a distance is considered to mean that neither the service provider nor the recipient is simultaneously present when the service is provided, cf. Directive 98/48/EC Art. 1 (2) (a).[201] This interpretation also follows from the understanding of the E-Act Sec. 1 (2) (a), and is supported by the law's legal history, cf. Ot.prp.no. 31 (2002-2003) page 55. The search engines' considered in this thesis are completely Internet-based. This means that the user (the recipient of the service) will never conduct his or her search from any office or location owned or offered by the search engine company. The requirement that the service must be provided *at a distance* must therefore be considered as fulfilled.

The service must be performed *by electronic means*. A common understanding of this would be that if the provider performs the service electronically, then the requirement is fulfilled. However, "*by electronic means*" is defined in Directive 98/48/EC Art. 1 (2) (a) to mean that *"(...) the service is sent initially and received at its destination by means of electronic equipment for the processing (including digital compression) and storage of data (...)"*. Furthermore, the entire transmission of the data must have happened by electromagnetic means, cf. Directive 98/48/EC Art. 1 (2) (a). Annex V of this Directive considers that if a service has some form of *material content*, even if it is provided through electronic devices, it cannot be considered as provided *by electronic means* cf. Sec. 2. Examples of such services are stated in the Annex, and one example is automatic cash machines; here the service is provided electronically (giving access to the account through the machine), but it takes material form at the final destination (the cash withdrawn).

Concerning the E-Act, its legal history translates this definition directly in the section concerning explaining the expression *by electronic means,* cf. Ot.prp.no. 31 (2002-2003) page 56, Sec. 3. In the same section the legal history states that this will include the electronic communication that takes place in the Internet. The wording chosen in Directive 98/48/EC must also be said to include services that offers transmissions based solely on the Internet. The preamble does not offer any clarity to this issue. Considering that search engines do not offer their results and caches as *material* content (a computer is needed to view them), they must be considered to fulfil the requirement of providing services *by electronic means*.

201 In Annex V, the same Directive lists an example of services which are not *provided at a distance*. The common feature of these examples are that the service provided has happened in the " (...) physical presence of the provider and the recipient, *even* if they involve the use of electronic devices", cf. Sec. 1. (Italicizing done by me).

The requirement that the service has to be provided *at the individual request of a recipient of the service* must be understood as exactly that. A concrete and individual request must be directed to the service from the recipient of the service. This means that for instance television broadcasting or teletext is not included, cf. Annex V of Directive 98/48/EC, Sec. 3. This is because the broadcasting and the teletext is not provided after an individual demand, but rather created for the simultaneous reception of an unlimited number of recipients, so-called "point-to-multipoint" transmissions. On the other hand, for instance "video-on-demand" services will be included, as they are "point to point" transmissions, cf. the legal history[202] of the E-Act page 56, Sec. 2. Considering the wording of Directive 98/48/EC and the specific exclusion of "point-to-multipoint" transmission services in Annex V, the inclusion of "point-to-point" transmissions must be unquestionable.

The search engines must be considered as "point-to-point" transmission-services. They will not offer a search on their own initiative, and does not have a pre-defined set of results. Every result will come as a direct result of the individual request of a user. The search engines' must therefore be considered to fulfil the requirement of supplying services on *individual requests*.

It must therefore be concluded that search engines are an information society service, and their services can be regulated by E-commerce legislation.

3.2.2 Do the exemptions from liability apply for caching performed by search engines?

The E-commerce legislation provides exemptions from liability for three different types of service providers; those that offer services consistent of a mere conduit of information[203], the cache-process[204] and finally the hosting services[205]. As stated above in 3.1 the word *cache* is not a unanimous term, but rather a term that covers a wide range of technical aspects. Even if the regulations use *caching* as a head line it does not mean that it must be presumed that every technical solution described as a cache is covered by them. However, as it is apparent that the caches created and offered by the search engines are more than a mere conduit of information[206], and are not a hosting service either, it is the regulation concerning *caching* that is interesting in this regard.

202 Ot.prp.no. 31 (2002-2003).
203 Cf. Directive, Art. 12, cf. E-Act, Sec. 16.
204 Cf. Directive, Art. 13, cf. E-Act, Sec. 17.
205 Cf. Directive, Art. 14, cf. E-Act, Sec. 18.
206 For instance the TCP/IP buffer explained above would most likely fulfill the requirement stated in regard to the *mere conduit* regulation.

The Directive provides exemptions from liability without emphasising which types of liability that are meant. The wording however indicates that it is every form of liability, and in practise the most important forms would be liability for damages (civil liability) and criminal liability. The E-Act elaborates this by specifying that the exemptions are for both liability for damages and criminal liability, cf. Sec. 16-18.

There are several requirements which all must be fulfilled in order to obtain the exemption from liability, as stated in the Directive, Art. 13 cf. the E-Act, Sec. 17. For the most part Section 17 of the E-Act is a direct transcript from the Directive, but there are some things that might be different, and I will elaborate on this in the relevant parts.

The Directive states that a service provider, who stores information in order to transmit it in a communications network, shall not be held liable, presumed that the storage is *"(...) automatic, intermediate and temporary (...)"*, and the information is provided by a recipient of the service, cf. Art. 13 (1). The wording "*information provided by a recipient of the service"* might indicate that the information must be given to the service provider, by a user of that service. However, this would not correspond well with the way for instance a proxy server's cache works[207], as explained above in 3.1[208]. However, Recital 20 states that the definition *recipient of a service* also covers persons who provide information on the Internet. This means that a service provider which also, in addition to transmitting information directly provided for it, transmits information located on the Internet, could be covered by Art. 13.

The E-Act has somewhat different wording; *"a service provider that transmits information for a recipient of a service in a communications network (...)"*[209], cf. Sec. 17 (1) cf. 16 (1). This could be interpreted in two ways, either that the provider transmits information on behalf of the owner of that information, or that it transmits information for someone who requests it. The legal history states that Sec. 17 is passed with proxy servers' cache in mind, cf. Ot.prp.no.4 (2003-2004) ch. 8.1. And considering the functions these caches fulfil, the second interpretation must be the correct one. This interpretation will also support the way the search engines operate with their caches, and thus they will fulfil this provision as well.

207 The recipient of the service (the cache) is the user connected to that particular network. The information provided comes from a third party, for instance an online news paper.

208 The preliminary indication might correspond better with a proxy server cache if one can view the owners of the cached web pages as recipients of the proxy server cache's services. This is because the cache will take redirect some of the traffic away from the owners servers and thus freeing bandwidth.

209 NO: "En tjenesteyter som overfører informasjon for en tjenestemottaker i et kommunikasjonsnettverk (...)".

However, the matter at hand in this thesis is not of the caches contains illegal information, but whether the very creation of the cache is illegal. The court in the case against Google in Belgium[210] mentioned these rules briefly, but decided that they cannot be applicable since "(...) *it is the behaviour of Google itself that is incriminated and not the content of the sites on which Google permits access (...)"*, cf. Sec. 8 of the "Discussion" in the ruling. Considering Article 14 regarding *hosting* there will always only be the matter if the contents of the pages are legal or not, not if the service provider has gained legal access to them, because it is the recipients of the service that *actively* uploads the contents to the providers' servers. In regard to the caches that might be regulated by Article 13, they will always either gather information on their own accord, like search engines do, or because of repeating requests for certain pages, as proxy servers do. The proxy servers are a good example in this regard. The electronic commerce legislation; hereunder the exemptions from liability, has been adapted in order to ensure the *"(...) free movement of goods, services and the freedom of establishment (...)"*, as stated in Recital 1 to the Directive. The service providers shall be able to offer their services to the internal marked without consideration to borders, cf. Recital 3. To gain this objective the Directive aims to support the development of information society services by establishing a *real area* without internal borders, cf. Recitals 2 and 3.

If the exemption from liability, as stated in Art. 13 cf. Sec. 17[211], merely were to apply for the actual *contents* stored, and not the way the service providers actually obtain this information, then this exemption would not have much practical worth. The service providers running the proxy servers would then not risk being liable in terms of storing otherwise illegal contents, but they would risk being liable for their very practice. The objective of the Directive would in that case not be fulfilled. It is this author's opinion that the regulation regarding caching services may be interpreted in a way that includes the situations where the service providers themselves caches copyright protected material from the rightholder's web page. It is therefore this authors humble opinion that the Belgium court perhaps were too quick to dismiss the possible grounds found in the international electronic commerce legislation, which Belgium in this case also are bound by[212].

210 Google vs. Copiepresse. Ruling by *the court of first instance in Brussels*. No. 06/10.928/C. "Copy. Issued to the Party Copiepresse"

211 Art. 13 of the Directive and Sec. 17 of the E-Act.

212 Belgium was one of the founding members of the European Union. http://www.europa.eu/abc/european_countries/eu_members/belgium/index_en.htm [13.04.07].

The conclusion must therefore be that also the actual caching of pages that search engines conduct may be covered by the Directive Art. 13 cf. the E-Act Sec. 17.

The requirement of automatic, intermediate and temporary storage

The storage of information to be transmitted must be *"(...) automatic, intermediate and temporary (...)"*, cf. the Directive Art. 13 (1) cf. E-Act Sec 17 (1).

When the proxy servers caches pages it is because the page in question has been accessed by users connected to the server network. Therefore the proxy server never actively chooses which pages to cache; the administrators of it have no control over which pages are cached and the process must be said to be completely *automatic*. The search engines do not create caches as a result of users accessing the indexed page; they do so by their own decision. In a way this might be viewed as a more controlled form of caching than what's the case with the proxy servers. However, the search engines that caches pages and makes these availably, does so for *every* page indexed. This means that the administrators of the search engines do not actively choose that certain pages shall be cached.[213] The caches created by search engines are *automatically* produced when a page is indexed, cf. ch. 3.1 above. The first condition is therefore fulfilled.

Furthermore, the cache is neither the place of origin for the information, nor the final destination. The cache is created and made accessible so that users might access the information stored there, and thus it is *intermediate* in nature. The requirement of the storage being *intermediate* must be considered as fulfilled because of this.

Moreover, the information must only be stored *temporarily*, and it is more uncertain that the search engines' fulfil this premise. As stated in ch. 3.1 the cache is updated every time a web crawler visits the page anew. This means that the information gathered from a page and stored in the cache might be updated or deleted every hour, day or year, depending on how often the crawler is set to visit the page in question. The question is thus what is meant by "temporary". The common understanding of the term would be that the information stored is intended to be removed, and will be removed, replaced or updated. However, how long it is acceptable to store information before it must be considered as a more permanent form of storage is uncertain. The preamble to the Directive does not clarify this problem, nor is there any case law that might be applicable.

The E-Act contains the same requirement, and its legal history[214] relates the requirement of temporary storage to the rules regarding the updating of the

213 Such decisions could for instance rely on type of page, how popular it seems to be etc.
214 Ot.prp.no. 4 (2003-2004).

information, cf. page 18, sec. 5. By this it is meant that if a copy is stored "forever" under the pretence of being a cached copy, it will not fulfil the requirements relating to updating the information, as well as the premise of being of a temporary nature.

A question that might be asked in relation to this is if a search engine checks the original copy dutifully, and this remains the same all along so that the information stored in the cache will remain there unchanged for years, is it still a temporary copy?

If compared to Art. 12 of the Directive, cf. Sec. 16 of the E-Act, the requirement is that the storage of information is *transient*, this means that the deliberate use of the word *temporary* in Art. 13 cf. Sec 17, indicates a more prolonged storage. The crux of the matter must be said that it doesn't only depend on how long the information is stored, but also on a broader assessment of the situation and why the information is stored.

The search engine will as mentioned above create the cache from the information gathered for the indexing process, and will be programmed to automatically update this information. The frequency of these updates will depend on how often the crawlers are programmed to visit the web pages stored; which will furthermore depend on how often the pages seem to be updated.

Since the purpose of the search engines' cache never are to create a permanent copy, the requirement of *temporary storage* must be considered fulfilled.

Yahoo! for instance has a cached copy of each page indexed, but will also provide a link[215] for *"previous versions at the Internet Archive"*. The pages stored at the Internet Archive[216] are permanent copies and the Archive can therefore not be considered as fulfilling the requirements of Art. 13 of the Directive or Sec. 17 of the E-Act, even if their copies of web pages might seem to be exactly the same as the ones created directly by the search engines.

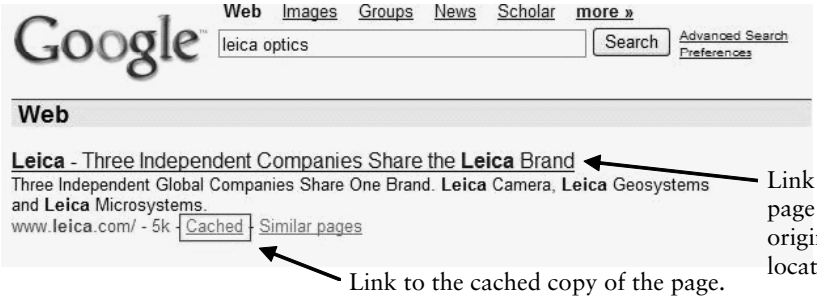
The requirements of sole purpose and individual requests

The storage of information must be performed for the *sole purpose* of making the onward transmission to other recipients more efficient, cf. Directive Art. 13 (1) cf. E-Act Sec. 17 (1).

215 An example may be found here: http://216.109.125.130/search/cache?p=dagbladet&fr=yfp-t-501&toggle=1&ei=UTF-8&u=www.dagbladet.no/&w=dagbladet&d=XMKwGBIeOi_C&icp=1&.intl=us [02.04.07].

216 http://www.archive.org/web/web.php [02.04.07].

Also in this regard the cache created by the search engines is different from the proxy servers' cache. The proxy servers' cache is undoubtedly created for the sole purpose of making the onward transmission of information more efficient; if they had not existed then the more popular web sites would spend much time "down" and out of service because the available bandwidth and server capacity would be exceeded. The search engines' cache does not however seem to have efficient transmission as their purpose; the cache is not the main choice provided when a search is conducted. It is on the contrary a choice somewhat hidden in small fonts below the snippet of information from the web pages in question. Below is an example of how Google presents the link to the cache:[217]



However, since the cache is routinely updated whenever a crawler visits the page, and therefore does not fulfil a purpose as a sort of historical archive, the initial assumption might be wrong. The only purpose with the cache, or at least the only purpose with giving users access to the cache, is to enable access to pages that might be offline for any reason. For instance because of a server overload, or that they are taken down for maintenance.[218] Enabling access to users in these cases would lead to a more efficient transmission of information, and if the pages are down for server reasons[219] it will help the situation by diverting traffic away from the original page's server. Recital 42 of the Directive states that if this requirement is fulfilled then it will imply that service provider has *"(...) neither knowledge of nor control over the information which is transmitted or stored"*. This implication can be turned around; if the service provider has neither knowledge of nor control over the information which is transmitted or stored, then this implies that the requirement is fulfilled. The

217 Yahoo!'s and Windows Live Search's way of presenting links to their cached copy are more or less alike this example.

218 By maintenance it is mean the structure of the original page, not updates of the information itself.

219 As is often the reason, especially for very popular pages.

cache is as stated above created automatically; the search engine companies caches all pages indexed, and thus has no knowledge or control over the information stored. They may acquire this knowledge or control manually if they wish, but this is the situation also for the administrators of the proxy servers, and should not be the defining point.

Although on uncertain grounds, the only practical reason for making this type of cache public is to promote efficient transmission; primarily by diverting traffic from servers that might be down for some reason. The requirement of efficient transmission as their sole purpose must therefore be considered fulfilled.

The transmission must also happen only at the *request* of the recipients of the service; i.e. the users, cf. the Directive Art. 13 (1), cf. E-Act Sec. 17 (1). The cached copy is not sent out to users, neither randomly nor to certain groups of people. The user must actively access the cache, and the requirement is undoubtedly fulfilled.

The requirement of no modification of the information

The provider must not modify the information, as stated in Art. 13 (1) (a) of the Directive, and in the E-Act Sec. 17 (1) (a). The cache created by the search engines' is an exact copy and is thus never modified by them. The search terms are often highlighted in the text, but the information is still presented in the same way as the original and the contents are the same. The requirement is therefore fulfilled.

The requirement of service providers' compliance with conditions stated for access to the information

This provision is stated in Art. 13 (1) (b) cf. Sec. 17 (1) (b). [220] The legal history of the E-Act lists as an example that the service provider must pay heed to for example password protection, and hinder users from gaining access to this information without the password, cf. Ot.prp.no.4 (2003-2004) page 38, last section. The search engines' cache is created with the use of the information crawled for the index. By this it is meant that the page is crawled with the purpose of collecting information and data for the indexing, and the cache is simultaneously created. The creation of the cache might be in question from the copyright legislations point of view, but the access to the information must follow the rules and regulation concerning indexing. As stated above in 2.4.2.3, consent for indexing of web pages must be considered given implicitly if counteractive measures, such as robot.txt, are not taken. This means that if

220 Of respectively the Directive and the E-Act.

such measures are not taken, and the information might be legally gathered for the index, then access to the page has been opened for the cache as well, even if the way the cache itself is created and presented might constitute a copyright breach. So that even if a page with password protection is cached, this is not a violation of the above requirement, as long as the page is not also equipped with solutions for keeping the web crawlers away. Because of this the search engines' cache must be considered to have fulfilled this requirement as well.

The requirement of service providers' compliance with used and recognised rules for updating the information

This requirement is stated in Art. 13 (1) (c) cf. Sec. 17 (1) (c).[221] The rules must be *widely recognised* and used by the *industry.* Industry must here be assumed not to be the industry the owners of the information belongs to; i.e. that if it is information provided by an online news paper, then it is not rules used and recognised by the news industry that are defining. On the contrary, it should be the industry that the service provider belongs to; in this case the technological information society industry.

The legal history of the E-Act mentions *Meta-tags* as a form of rules widely used and recognised by the industry, cf. Ot.prp.no.4 (2003-2004) page 18, sec. 5. Meta-tags are not visible for the regular users of a web page, but are located in the code of the page. Meta-tags might contain info on for instance how long a page is valid, or how often it should be checked for updates. The crawlers must be configured to recognise meta-tags, and not all are. If a service provider takes meta-tags into account, and respects the information provided here, then they are considered as fulfilling this requirement, cf. page 18, sec. 5 of the legal history. It is mentioned in the legal history that a service provider ought to take such tags into account, but does not state it as a requirement as long as other sufficient action is taken to fulfil the requirement.[222]

Recital 47 of the Directive prevents the states from imposing general monitoring obligations on service providers, but considering the requirement in Art. 13 (1) (c) this recital must be regarded as a prevention from obligations to monitor for illegal contents; not monitoring for updates.

If a search engine chooses not to take Meta-tags into account, then it must be configured to locate which pages that should be monitored more closely for updates in another way. For instance online news papers should be visited and the cache updated daily, or almost daily, for the requirement to be fulfilled,

221 Of respectively the Directive and the E-Act..

222 Stated on the same page in the legal history as above.

while pages that are more static of nature[223] can be visited more sparingly, and the requirement will still be fulfilled.

It is difficult to conclude whether or not the search engines' are complying with this requirement, it must really be considered from case to case. However, since most search engines will take meta-tags into consideration they must probably be considered as fulfilling this requirement. A search for web pages that are updated both daily and otherwise will show that at least the cache of most online news papers are updated often, and thus the requirement must be considered as fulfilled.

The requirement that service providers must not interfere with the lawful use of technology to obtain data on the use of the information

This provision is stated in Art. 13 (1) (d) cf. Sec. 17 (1) (d)[224]. The legal history of the E-Act states hit counters as an example of technology which mustn't be interfered with, cf. page 39, sec. 2.[225] A hit counter will register the number of visitors on a web page; the data can for example be used as a foundation for calculating how much advertisers shall pay for the access to advertise on the page. The wording of the Directive is also consistent with an example of this nature. As correctly pointed out in the legal history, the technology for heeding this requirement exists, and more sophisticated ways will probably be developed. A way of solving this would be that for instance the proxy server's cache would register how many hits a cached page has, and then communicate this number to the original page.

The problem is that it is nearly impossible to prove if a service provider takes notice of this requirement. And the question that arises is if *it is the provider* that has to prove that the requirement is followed, or if it is for instance *a rightholder* that has to prove that it is not. In the criminal area the leading principle is the presumption of innocence. It is always the accuser that has to prove guilt, and not the accused that has to prove innocence. This principle goes beyond the area of criminal law. If someone tries to hold another liable for damages, then they must prove that the latter is liable; not the other way around. Therefore the principle must be considered as applicable here as well. The technology exists, and search engines' are for the most part very technological advanced; especially the large companies which are in question here.[226] The probability of them applying such technology, and thus heeding

223 For instance an historical page conserning the American Civil War.
224 Of respectively the Directive and the E-Act.
225 Ot.prp.no.4 (2003-2004).
226 Google, Yahoo! and Windows Live Search.

this requirement, is profound. And therefore they must be given "the benefit of the doubt" until proven otherwise. Because of this the search engines must be considered to fulfil this provision as well.

The requirement of service providers' removal of, or disabling of access to, information

The final requirements for the exemption from liability are the expedited removal of information, or the disabling of access to that information, if the service provider is notified to do so, cf. the Directive Art. 13 (1) (e), cf. the E-Act Sec. 17 (2).

The Directive requires the service provider to *"(...) expeditiously (...) remove or disable access to the information it has stored (...)"* when they acquire ***actual knowledge*** that either *"(...) the initial source of the transmission has been removed from the network, or access to it has been disabled (...)"*, or *"(...) a court or an administrative authority has ordered such removal or disablement"*. This is a requirement that cannot be answered in advance. What can be discussed however, is if this also requires the service provider to show due care. The prevention from imposing *general* monitoring obligations, cf. Recital 47, was above found not to affect the requirement of checking for updates. It probably cannot be said to affect these requirements either, since these are imposed by the Directive itself, while the prevention is aimed at requirements imposed by the national states themselves. Moreover, even if Art. 15 (1) of the Directive also states the prevention of such a general monitoring of information, this must be related to obligations imposed by the national states themselves.

However, the wording states that the service providers must comply with this requirement when they have *obtained actual knowledge*. The common understanding of this would be that the service providers do not have to check for such information on their own accord. And that they can wait until they receive notification of orders from the court or an administrative authority. Moreover, Recital 48 states that the Directive "(...) ***does not affect*** *the possibility for Member States of requiring service providers, who* ***host*** *information (...) to apply* ***duties of care*** (...)".[227] This is explicitly directed at hosting services, cf. Art. 14, and not the services regulated through Art. 12-13. This means that the Directive has opened for imposing a general obligations of due care on hosting services, but not necessarily on the services providers in regard to caches or mere conduit.

227 Bold types created by me.

The other possibility is if the initial source has been removed or access to it has been disabled. The one must probably conclude with that if the requirement of checking for updates is complied with, then this requirement will also be fulfilled. When the crawlers visit the page again, they will discover if the access has been disabled, or the contents removed.[228]

The E-Act has divided these two possibilities. While the Directive requires the obtaining of *actual knowledge,* the E-Act requires *gained knowledge*[229] of a court or administrative authority's decision, and *received notification*[230] concerning if the initial source has been removed or access to it disabled. A common understanding of the wording would indicate that if the search engine company has received some sort of knowledge that there is for instance a court order, then they would probably have to find out if there is any truth to this. This is because knowledge may be gained in any number of ways, from "hearsay" to actual notification. And on the other hand, the understanding of received notification would suggest that there must be an actual and concrete notification before the search engine must take action. The plight to investigate further seems therefore to exist in relation to the order by a court or administrative authority, but not in relation to the situation where the information has been otherwise removed or access disabled.

The legal history does not however impose such an obligation for investigations, except for when the service provider receives *actual knowledge,* cf. Ot.prp.no.4 (2003-2004) page 39, sec. 5. The reason why the information must be provided by the one that wishes for the information to be removed, is that it will be easier for him or her to acquire the necessary information (for instance regarding a court order), than it would for the service provider, cf. the same page as above. If there is still need for investigation after acquiring such knowledge, for instance to check if the court order is genuine, the requirement will be fulfilled if such an investigation is instigated expeditiously and the contents are immediately removed after the necessary clarifications are done, cf. page 18, sec. 8 of the legal history. It is also recognized in the legal history that when automatic checks for updates happen very often, the service provider does not need to do anything extraordinary in order to remove the information; it will be removed automatically within the time limits stated, cf. page 18, sec. 9. If the web crawlers visit a page for instance every day or every other day, like they do with most news sites, they can be considered to always be in compliance with the requirement.

228 The search engine must of course take appropriate action when the crawlers discover this, or else they will not have complied with the requirement.

229 NO: "fått kunnskap."

230 NO: "mottatt underretning."

There are no court decisions regarding Art. 13 of the Directive or Sec. 17 of the E-Act.[231]

Conclusion

It has been established above that when caching pages, the search engines fulfil all of the requirements that may be considered before an actual situation has occurred. The requirements that may not be answered in advance, such as the removal of access to pages after a court order, are not of such nature that search engines cannot fulfil them. On the contrary, there is a high degree of probability that they may and will fulfil these requirements. Therefore the conclusion must be that the search engines will benefit from the exemption from liability in regard to their caches, cf. Art. 13 of the Directive, and Sec. 17 of the E-Act. This means that even though they violate the right of reproduction, they cannot be held liable. This does not mean that they have the necessary legal grounds to operate the way they do, but as long as they comply with the above regulations, then there is no legal sanctions that can be imposed on them. This is an unfortunate situation, since the search engines' cache is practical, but not completely necessary. While for instance the proxy servers' cache are highly necessary for efficient use of the Internet, the search engines' cache are not that necessary. It might seem that the legislators have not had the search engines' in mind when they passed these statues. This author thinks that if they had, then they would have taken the necessary steps to exclude the search engines' from this exemption from liability.

The objective of the E-commerce legislation is to ensure the free movement of information society services within the EEA, as a contribution to the proper functioning of the internal market, cf. the Directive Art. 1 (1) cf. the E-Act, Sec. 1 (1). Granting the search engines' limitations from liability in relation to their caches, are not necessary for the achievement of this objective. Nevertheless, it is the result of the way the legislation is formed. The legal history of the E-Act explicitly states that the draft legislation presented will not regulate browser-caching (i.e. client-caching and the memory cache), cf. page 7, ch. 3.3.[232] A similar statement concerning the search engines' cache could be appropriate. This is because there is no real reason for why the search engines should enjoy such protection from liability in regard to their cache-function.

231 The closest court ruling is from Oslo District Court, cf. RG 2005 p.1627. It is however, not applicable in this regard since the ruling concerns Sec. 16 of the E-Act, and the court is very specific in its deliberation. Thus it cannot be viewed in context with Sec. 17.

232 Ot.prp.no. 4 (2003-2004).

Their cache can be very practical for many users, but unlike the proxy servers' cache, it is not necessary for ensuring that the "traffic" goes smoothly.

If such a statement is clear enough, then it would have more weight in the interpretation of the regulations, than the actual wording of the legal text. However, since this is a regulation implemented because of a Directive this could be problematic. There is nothing in the Directive, neither in the actual text nor in the preamble, that indicates an exclusion of search engines' from it. And since one of the purposes of this Directive is to establish a set of certain minimum rights for service providers, a statement of this nature would mean that Norway had not fulfilled its obligations. On the other hand, the Norwegian legislation is to be interpreted with the presumption that it is in accordance with Norway's international obligations. If the statement is clear enough, then there is nothing that can be interpreted in any other way from the point of view of Norwegian legislation. The question is then if there is any way that at least one of the requirements of the Directive can be said not to have been fulfilled by the search engines', in regard to the exemptions from liability. In this author's humble opinion that cannot be done. And therefore the exclusion of search engines' from these exemptions from liability cannot happen at a national level. It must happen at an International level. It would be fortunate if this legislation were to be revised, because it is this author's opinion that the legislators have passed the Electronic Commerce legislation without fully comprehending the reach some of its regulations would have.

4 ALTERNATE SOLUTIONS

4.1 Introduction

In this chapter I will briefly look at alternate ways for establishing a clear framework of rules governing the search engines' practice. This will include "pure" legal solutions as well as legal solutions which presuppose certain technical methods. The latter will first and foremost be considered with the search engines' caches in mind.

As concluded above in chapter 2.4.2, implicit consent must be considered given. However, to completely rely on consent given implicitly is not a very tenable situation. It is this author's opinion that the legislators should adopt explicit legal grounds allowing for the *indexing* of web pages. At the very least they should openly recognize the problem and take a stand, either allowing or disallowing the search engines' practice.

4.2 U.S. Law

4.2.1 Introduction

The legal system in the USA is very different from continental Europe's legal system. USA is a part of the common law-system, while most of Europe is part of the civil law-system.[233] The civil law-system will first and foremost look to the formal legislation, while the common law-system will focus on case law. However, in recent years the U.S. legal system has created several legal statues, presenting a more similar system to the one in Europe.[234]

The Norwegian legal system is a sort of hybrid between these systems. It has an extensive regulation through formal law, but court rulings are used extensively as a factor of interpretation; in particular those of the Supreme Court. This is not only a Norwegian phenomenon; also the other Nordic countries have a similar system.

However, the vast application of court decisions as sources of law, and the fact that the legal thinking in the U.S. will be different from the European way

233 The most important exeption to this is the United Kingdom, which also have the common law-system.

234 Burnham, William; *Introduction to the law and legal system of the United States*, (St. Paul 2002) p. 49.

of thinking, requires a certain critical sense when considering solutions from U.S. law in relation to renewing European legislation.

Similarities

There are several similar principles in terms of copyright related thinking. One is the consideration what is to be protected by copyright legislation. In the U.S. one will find that copyright protection does not extend to *"(…) any idea, procedure, system, method of operation, concept, principle of discovery (…)"*, cf. Facts v. expression of facts (17 U.S.C.A § 102 (b))". The same preclusion may be found in the WIPO Copyright Treaty; *"Copyright protection extends to expressions and **not to** ideas, procedure, methods of operation or mathematical concepts as such"*[235], cf. Art. 2. As mentioned in chapter one; the WCT is founded on the Berne Convention, and the European Union approved of it through a Council decision. This is therefore clearly a European principle as well as a U.S. principle.

Differences

There are of course many differences between the two legal systems. One of the more important in terms of copyright is how the rights are granted to the rightholder. In the U.S. the rights are viewed as **given by** the legal system to someone; be it a natural or legal person.[236] For the most part the rightholder's rights are viewed as the right of reproduction. This can be illustrated by the Supreme Court decision *Fogerty v. Fantasy*[237]. The court states that *"(…) the ultimate aim of copyright is not to reward the labor of the authors, but to promote the Progress of Science and useful Arts"*. In the European way of thinking the rightholder achieves his or her rights the moment the work is created. And while the U.S. legal system might view for instance a corporation the creator of the work[238], the European legal system will always view a natural person as the creator. If the rights are held by a corporation then there must have happened an actual transfer of rights. This can be illustrated with a look at the Berne Convention Art. 1; *"The countries to which this Convention applies constitute a Union for the protection of the **rights of authors** in the literary and artistic works"*[239].

Moreover, the view of personal rights such as the right to be named does not naturally belong in the U.S. way of thinking. And this difference of think-

235 Bold types created by me.
236 (Andersen, 2005), p. 280.
237 510 U.S. 517, 526 (1994).
238 An example of how this might work in practice is Microsoft Word 2007. When accessing "managing sources" for the bibliography fields, the option "Author as Organization" exists.
239 Bold types created by me.

ing was also part of the reason why the U.S. waited so long before ratifying the Berne Convention.[240]

Can inspiration be drawn from U.S. Law?

In the legal history of the Norwegian E-commerce Act, the original proposal suggested that the system of *"Notice and takedown"*, from the Digital Millennium Copyright Act (DMCA) in U.S. Law, should be applied.[241] Ultimately it was decided against it after the consultative round, but the issue was after all raised. This shows that it is not at all unheard of to look to U.S. Law for inspiration and ideas. Although, as stated above, the Norwegian legal system is a sort of a hybrid between civil law and common law, so this point of view might not be tenable in regard to the rest of Europe. However, the European Union also uses case law to a large extent, so it cannot be unheard of.

The conclusion must therefore be that as long as one remains critical, and is conscious of the differences in legal thinking, inspiration may be drawn from U.S. Law.

4.2.2 Fair use

The Digital Millennium Copyright Act of 1998 contains a provision of *fair use*, cf. 17 U.S. Code § 107. Below is the legal text in its entirety.[242]

§ 107 · Limitations on exclusive rights: Fair use

Notwithstanding the provisions of sections 106 and 106A, the fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include—

(1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;

(2) the nature of the copyrighted work;

(3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and

(4) the effect of the use upon the potential market for or value of the copyrighted work.

The fact that a work is unpublished shall not itself bar a finding of fair use if such finding is made upon consideration of all the above factors.

240 (Andersen, 2005), p. 280.

241 Ot.prp.no.4 (2003-2004) page 22.

242 § 106 regulates "Exclusive rights in copyrighted works" and § 106A regulates "Rights of certain authors to attribution and integrity".

The DMCA does not contain a definition of fair use, but four factors that must be considered when deciding whether or not a use of a work can be considered as fair use. The DMCA also contains examples of fair use, for instance in relation to teaching or research. In this thesis I will not consider or interpret the factors thoroughly, mainly because U.S. Law is not regard in this thesis as a part of the de lege lata-analyze. There are however case law on this subject; more specifically there is especially one case regarding search engines and fair use. The case of *Kelly v. Arriba Soft Corp.* [243] ("Arriba") dealt with the fact that Arriba's search engine indexed thumbnail images.[244] The plaintiff, Leslie A. Kelly, was a photographer that claimed Arriba violated his copyright to the images when they did this. The court established that there was a copyright infringement before analyzing whether or not the fair use doctrine applied. All four factors concerning fair use were discussed before making a verdict. In the following I will look at these four factors, and briefly recite how the court viewed them.

The first factor considers the *purpose and character of the use,* cf. § 107 (2) (1) . It should especially be considered if the use is for commercial purposes. The court states that *purpose and character* also involves an assessment of whether or not a transformation of the work has happened. If the work is used anew in the exact same form and way as the original, then this will weigh against fair use.[245] The more transformative the new work is, the less other factors, like commercialism, will weight against fair use. The court found that since the images were used by Arriba in a different setting and for a different purpose, the fact that they were used for commercial purposes did not weigh as much. Therefore the first factor was considered to weigh in favor of fair use.

The second factor is the *nature of the work*, cf. § 107 (2) (2). The court states *"(…) that some works are closer to the core of intended copyright protection than others, with the consequence that fair use is more difficult to establish when the former works are copied (…) [Campell, supra 510 U.S. at 586]"*.[246] Because artistic works are part of the core mentioned in the citation, the court found that the second factor weighted against fair use.

The third factor concerns the amount copied, cf. § 107 (2) (3). This factor is considered by the court to be an assessment of *"(…) whether the amount copied was reasonable in relation to the purpose of the copying (…)"*.[247] The

243 Case No. SA CV 99-560 GLT[JW]

244 The court states that Arriba's search engines functioned like any other Internet search engine, apart from that instead of retrieving texts to match a query, it retrieved images.

245 "II Discussion"; "1. Purpose and Character Of The Use".

246 "II Discussion"; "2. Nature of the Copyrighted Work".

247 "II Discussion"; "3. Amount And Substantiality of the Portion Used".

court found that the thumbnail images probably were reasonable in light of its purpose, but since Arriba also presented an image attributes page with a more full-size image, this factor weighted slightly against fair use.

The fourth and final factor is whether the effect of the use on the potential market or value, and considers the direct impact made on the interests of the rightholder[248], cf. § 107 (2) (4). Arriba argued that there was no negative impact because they did not compete with Kelly, but actually increased traffic to his site with their search engine. It is recognized by the court that since Arriba enabled "deep links" directly to the pages containing the images, the main page were bypassed and this could potentially harm advertisement income. However, since Arriba showed evidence that tended to show a lack of market harm, and Kelly could not prove that he had suffered any loss, the court found that the fourth factor weighed in favor of fair use.

There were two factors weighing against fair use, and two in favor. However, since these are factors and not requirements, an assessment of the whole case would be necessary. The court found that Arriba's conduct were fair use of Kelly's images.

Conclusion

Whether it is possible or not to implement the fair use-doctrine in European and Norwegian legislation, will require a more thorough analysis of the doctrine in regard to European Copyright traditions. However, as shown in Kelly v. Arriba Soft Corp., this doctrine might grant the search engines the formal statues they need for their indexing.

4.3 Technical solutions to a legal problem

4.3.1 Inclusion protocols instead of exclusion protocols?

As explained above in ch. 2.4 the robot.txt exclusion protocol is widely known among search engine administrators, and will be recognized and respected by every search engine that wishes to be perceived as serious. The most important objection against relying on the robot.txt protocol for avoiding web crawlers, is the Berne Convention Art. 5 (2) which states that protection shall not be subject to any formality.

248 "II Discussion"; "4. Effect of the Use On The Potential Market or Value".

A though could therefore be to turn for instance the principle of the robot.txt protocol around. Instead of crawling the web with the point of view that if there is not *protective measures*, then it is okay to index the pages, the attitude could be that if there are no *explicit records of approval*; i.e. records of registration or inclusion protocols, then the pages shall not be indexed.

To obtain such a turn-around would need legislative measures, because today's legal situation does not require explicit consent in regard to search engines and indexing, cf. ch. 2.4 above.

Furthermore, a requirement of such explicit consent would cause a very pronounced drop in the number of indexed pages, and effectively crippling the search engines; at least the major ones.

It is this author's opinion that the solution with explicit consent would be a poor one in regard to indexing. Although, in regard to the cache-function it could work well. However, if this type of legal solutions were to be respected by the search engines, something would probably have to be done with the exemptions from liability found in the electronic commerce legislation. As long as they do not risk anything, most search engines would probably continue today's practice.

4.3.2 Creative Commons

Creative Commons[249] ("CC") is an international network which originated in the USA, and provides licenses for the use of works on the Internet.[250] By applying a CC-license to a work, the rightholder allows the public to use the work in part or in full. There are several CC-licenses, and the scope of rights relinquished to the public depends on which license is applied. The licenses makes for the most part use of four different rights; *"Attribution", "Non-commercial", "No-derivatives"* and *"Share Alike"*.[251] These rights are combined differently in the various licenses, and an example is the *"Attribution Non-commercial No Derivatives (by-nc-nd)"*. [252] If a rightholder applies this license to a work it will allow others to download the work and share it with others, but they must mention and link back to the rightholder, and in addition they must not change the work in any way or use it commercially. This is the most restrictive license, and stands in contrast to the *"Attribution (by)"*-license which allows others to do what they wish with the work, even change it and/

249 http://creativecommons.org [16.04.07].
250 http://www.creativecommons.no [16.04.07].
251 http://creativecommons.org/about/licenses/meet-the-licenses [16.04.07].
252 There are six main licenses.

or use it commercially, as long as they credit the rightholder for the original creation.

The licenses are furthermore not revocable; i.e. that one can only prevent more people from making use of the work, but not stop someone that has already used the work in accordance with the license.[253]

There are no CC-licenses that directly deal with whether search engines may make use of the work. The closest is the license which relinquishes the work in its entirety to the public domain; this would include indexing and caching by search engines. To create a license allowing search engines to operate would probably not be a very fruitful concept in regard to the indexing. That is the case much for the same reasons as with the possibility for an "inclusion protocol", as mentioned above.

For the cache however, it would probably be a very good idea. The CC-licenses already exists, and are searchable.[254] In order for a standard for protocols enabling explicit consent to work, it would have to be international. And developing international standards is not always an easy task. Therefore, the further development of the CC-licenses in regard to the search engines' cache would probably be fruitful.

However, as stated above in 4.2.1, in order to get the search engines' to follow such a standard, something would most likely have to be done with the E-commerce legislation. As long as they have some form of "immunity" against liability, they probably will not undertake projects that will cost a great deal of money and time, in addition to leaving them with a strikingly smaller amount of pages cached.

253 http://creativecommons.org/about/licenses/fullrights [16.04.07].
254 http://search.creativecommons.org/ [16.04.07].

III – REFERENCE MATERIAL

BIBLIOGRAPHY

1) **Andersen Mads Bryde** *IT-retten* [Book]. - København : Gjellerup, 2005. - 2nd Edition.

2) **Burnham William** *Introduction to the law and legal system of the United States* [Book]. - St.Paul : West Group, 2002. - 3rd edition.

3) **Eckhoff Torstein & Jan E. Helgesen** *Rettskildelære* [Book]. - Oslo : Universitetsforlaget, 2001. - 5th Edition.

4) **Fielden Ned L. and Kuntz Lucy** *Search Engines Handbook* [Book]. - Jefferson, North Carolina and London : McFarland & Company, Inc., 2002.

5) **Harris DJ** *Cases and Materials on International Law* [Book]. - London : Thomson, Sweet & Maxwell, 2004. - 6th edition.

6) **Kielland Geir** *Teorien om implisitt lisens - Det rettslige grunnlaget og lisensens innhold.* - [s.l.] : The ALFEBIITE project (IRI), 2002.

7) **Koktvedgaard Mogens and Schovsbo Jens** *Lærebog i Immaterialret* [Book]. - København : Jurist- og Økonomforbundets Forlag, 2005.

8) **Mathijsen P.S.R.F.** *A Guide to European Union Law* [Book]. - London : Thomson, Sweet & Maxwell, 2004. - 8th Edition.

9) **Nielsen Ruth** *E-handelsret* [Book]. - København : Jurist- og Økonomforbundets Forlag, 2004. - 2nd Edition.

10) **Nygaard Nils** *Rettsgrunnlag og standpunkt* [Book]. - Bergen : Universitetsforlaget, 2004. - 2nd Edition.

11) **Rognstad Ole-Andreas** *Fragmenter til en lærebok i opphavsrett* [Book]. - Oslo : Institutt for privatrett, 2004.

12) **Sejersted Fredrik [et al.]** *EØS-rett* [Book]. - Oslo : Universitetsforlaget, 2004. - 2nd Edition.

13) **Sterling J.A.L** *World Copyright Law* [Book]. - London : Thomson, Sweet & Maxwell, 2003. - 2nd Edition.

14) **Tanenbaum Andrew S.** *Computer Networks* [Book]. - Amsterdam : Pearson Education, Inc., 2003. - 4th edition.

15) **Udsen Henrik and Schovsbo Jens** *Ophavsrettens missing link?* [Journal] // Nordiskt Immateriellt Rättsskydd. - 2006. - pp. 47-65.

TIDLIGERE UTGITT I COMPLEX-SERIEN

CompLex er Senter for rettsinformatikks skriftserie. Serien startet i 1981, og det har blitt utgitt mer enn hundre titler. Bøkene i CompLex-serien kan bestilles fra Akademika (se bestillingsskjema bak i boken eller www.akademika.no).

2007

5/07 **Gjennomgang av arkivretten**
Martin Rødland. ..... *NOK 129,-*

4/07 **Privacy & Identity Management**
Thomas Olsen, Tobias Mahler, et al. ..... *NOK 234,-*

3/07 **Personvern og transportsikkerhet**
Dag Wiese Schartum ..... *NOK 306,-*

2/07 **ZEBLEX 06 - Tre avhandlinger om fildeling, IT-sikkerhet og e-postkontroll**
Ida Otterstad, René Stub-Christiansen & Cecilie Wille Søvik ..... *NOK 348,-*

1/07 **Kontraktsregulering av domstolens kompetanse ved elektronisk handel**
Vebjørn Krag Iversen ..... *NOK 186,-*

2006

6/06 **Lover - fra kunngjøring til hyperstrukturer: to avhandlinger**
Per Marius Slagsvold & Kirill Miazine ..... *NOK 222,-*

5/06 **Retten til eget bilde**
Maria Jongers ..... *NOK 198,-*

4/06 **Legal, Privacy, and Security Issues in Information Technology Vol. 2**
Kierkegaard, Sylvia Mercado (editor) ..... *NOK 783,-*

3/06 **Legal, Privacy, and Security Issues in Information Technology Vol. 1**
Kierkegaard, Sylvia Mercado (editor) ..... *NOK 918,-*

2/06 **Rettslige reguleringer av informasjonssikkerhet**
Are Vegard Haug ..... *NOK 420,-*

1/06 **Anti-spam Legislation Between Privacy And Commercial Interest. An overview of the European Union legislation regarding the e-mail spam**
Dana Irina Cojocarasu ..... *NOK 155,-*

2005

1/05 **Renessansen som unnfanget Corpus Iuris Civilis. Keiser Justinians gjenerobring av Romerriket**
Halvor Manshaus ..... *NOK 249,-*

2/05 **Personvern og ytringsfrihet. Fotografering av siktede i straffesaker – et vern for ytringsfrihet?**
Anette Engum ..... *NOK 132,-*

3/05 **Rettigheter til geografisk informasjon. Opphavsrett, databasevern og avtalepraksis.**
Steinar Taubøll ..... *NOK 206,-*

4/05 **"The Answer to the Machine is in the Machine» and Other Collected Writings**
Charles Clark ..... *NOK 401,-*

5/05 **Digital Rights Management - Promises, Problems and Alternative Solutions**
Kristian Syversen ..... *NOK 201,-*

6/05 **DRM og Demokrati. Argumentasjoner, rettferdiggjøringer og strategier bak endringen av åndsverksloven 2003-2005**
Jan Frode Haugseth ..... *NOK 224,-*

2004

1/04 **Opphavsrettslige problemstillinger ved universitetene og høyskolene. Innstilling fra immaterialrettsutvalget, oppnevnt av Universitets- og Høgskolerådet 31. januar 2000. Avgitt til universitets- og høgskolerådet 8. oktober 2003**
Immaterialrettsutvalget ..... *NOK 341,-*

2/04 **Ansvarsfrihet for formidler ved formidling av informasjonssamfunnstjenester**
Bård Standal ..... *NOK 311,-*

3/04 **Arbeidsgivers adgang til å kontrollere og overvåke sine ansatte med hovedvekt på grunnvilkårene for behandling av personopplysninger i arbeidslivet**
Stefan Jørstad ..... *NOK 191,-*

4/04 **Elektroniske spor fra mobiltelefoner – om politiets bruk og teleoperatørenes lagring av trafikkdata.**
Christian Dahlgren ..... *NOK 117,-*

5/04 **International Jurisdiction and Consumers Contracts – Section 4 of the Brussels Jurisdiction Regulation.**
Joakim S. T. Øren ..... *NOK 172,50*

6/04 **Elektronisk dokumentfalsk.**
Lars Christian Sunde ..... *NOK 60,502003*

2003

1/03 **IT i domstolene. En analyse av norske domstolers teknologianvendelse fra 1970 til 2001**
Even Nerskogen ..... *NOK 330,-*

2/03 **Hvorfor vokser Norsk Lovtidend? En empirisk analyse av veksten**
Martin Støren ..... *NOK 87,-*

3/03 **Etableringslandsprinsippet. En analyse av e-handelsdirektivet art 3 og prinsippet om fri bevegelighet av tjenester ved elektronisk handel**
Jon Christian Thaulow ..... *NOK 213,-*

4/03 **The Law of Electronic Agents. Legal contributions to ALFEBIITE – A Logical Framework for Ethical Behaviour between Infohabitants in the Information Trading Economy of the Universal Information Ecosystem, IST-1999-10298**
Jon Bing and Giovanni Sartor (eds) ..... *NOK 351,-*

5/03 **LEA 2003: The Law and Electronic Agents Proceedings of the Second LEA Workshop, 24th June 2003, in connection with the**

ICAIL 2003 Conference (Ninth International Conference on Artificial Intelligence and Law), Edinburgh, Scotland, UK
Seminarrapport).....*NOK 228,-*

6/03 **Opphavsrettslige aspekter ved nettbasert formidling av musikk**
Stig Walle .....*NOK 153,-*

7/03 **Sceneinstruktørens opphavsrettslige stilling**
Edle Endresen.....*NOK 119,-*

8/03 **User-Centred Privacy Aspects In Connection With Location Based Services**
Christian B. Hauknes.....*NOK 203,-*

2002

1/02 **Koblingshandel og forholdet til fysisk og teknologisk integrasjon i relasjon til EØS-avtalens art.54(d)**
Ole Jacob Garder.....*NOK 180,-*

2/02 **To opphavsrettslige arbeider:**
Bjarte Aambø – Opphavsrettslige rettsmangler
Erlend Ringnes Efskind – Skjermbildets rettslige natur
Aambø / Ringnes Efskind.....*NOK 201,-*

3/02 **Arbeidstakeroppfinnelser ved universiteter og høyskoler. Innstilling fra et utvalg oppnevnt av universitets- og høgskolerådet 31 januar 2000. Avgitt til universitets- og høgskolerådet i oktober 2001**
.....*NOK 213,-*

4/02 **Utøvende kunstneres direkteoverføringer på Internett – med hovedvekt på kringkastingsbegrepet**
Irina Eidsvold Tøien .....*NOK 225,-*

5/02 **Administrasjon av radiofrekvensspekteret. Rettslige problemstillinger knyttet til telemyndighetenes forvaltning av frekvensressursene**
Øyvind Haugen .....*NOK 177,-*

6/02 **Overføring av personopplysninger til tredjeland. Kravet til tilstrekkelig beskyttelse etter EU-direktivet om personvern art. 25**
Mona Naomi Lintvedt og Christopher J. Helgeby.....*NOK 198,-*

7/02 Digitale mellomledds ansvar for videreformidling av ytringer. E-handelsdirektivet art. 12-14
Just Balstad.....*NOK 186,-*

8/02 Platekontrakten. Eksklusive overdragelser av utøverens rettigheter til eksemplarfremstilling og spredning
Øyvind Berge.....*NOK 237,-*

9/02 Varemerkerettslige konflikter under .no. I hvilken grad kan registrering og bruk av et domenenavn medføre inngrep i en varemerkerett? Hvordan løses konflikter under .no i dag, og hva kan være en mer hensiktsmessig tvisteløsningsmekanisme i fremtiden?
Silje Johannessen.....*NOK 192,-*

10/02 Vegard Hagen – Pekeransvar. Spørsmålet om ansvar for publisering av pekere på verdensveven (World Wide Web)
Hans Marius Graasvold – Pekeransvaret. Straffe- og erstatningsansvar for publisering av pekere til informasjon på Internett
Vegard Hagen / Martin Grasvold.....*NOK 234,-*

11/02 Personopplysningsloven § 7. En analyse av forholdet mellom personvern og ytringsfrihet slik det er uttrykt i personopplysningsloven § 7
Karen Elise Haug Aronsen.....*NOK 198,-*

12/02 Databasevern. Sui generis-vern av sammenstillinger etter gjennomføringen av databasedirektivet i åndsverkloven § 43
Lisa Vogt Lorentzen.....*NOK 210,-*

2001

1/01 Internet and Choice-of-Law – The International Sale of Digitised Products through the Internet in a European Context
Peter Lenda.....*NOK 275,-*

2/01 Internet Domain Names and Trademarks
Tonje Røste Gulliksen.....*NOK 227.-*

3/01 **Internasjonal jurisdiksjon ved elektronisk handel – med Luganokonvensjonen art 5 (5) og elektroniske agenter som eksempel**
Joakim S. T. Øren ..... *NOK 204.-*

4/01 **Legal issues of maritime virtual organisations**
Emily M. Weitzenböck ..... *NOK 164.-*

5/01 **Cyberspace jurisdiction in the U.S. – The International Dimension of Due Process**
Henrik Spang-Hanssen ..... *NOK 685.-*

6/01 **Norwegian border control in a Europe without Internal Frontiers – Implications for Data Protection and civil liberties**
Stephen Kabera Karanja ..... *NOK 252,-*

2000

1/00 **Klassikervernet i norsk åndsrett**
Anne Beth Lange ..... *NOK 268.-*

2/00 **Adgangen til å benytte personopplysninger. Med vekt på det opprinnelige behandlingsformålet som begrensningsfaktor**
Claude A. Lenth ..... *NOK 248.-*

3/00 **Innsyn i personopplysninger i elektroniske markedsplasser.**
Line Coll ..... *NOK 148.-*

1999

1/99 **International regulation and protection of Internet domain and trademarks**
Tonje Røste Gulliksen ..... *NOK 248.-*

2/99 **Betaling via Internett**
Camilla Julie Wollan ..... *NOK 268.-*

3/99 **Internett og jurisdiksjon**
Andreas Frølich Fuglesang & Georg Philip Krog ..... *NOK 198.-*

1998

1/98 **Fotografiske verk og fotografiske bilder, åndsverkloven § 1 og § 43 a**
Johan Krabbe-Knudsen ..... *NOK 198.-*

2/98 **Straffbar hacking, straffelovens § 145 annet ledd**
Guru Wanda Wanvik ..... *NOK 238.-*

3/98 **Interconnection – the obligation to interconnect telecommunications networks under EC law**
Katinka Mahieu ..... *NOK 198.-*

1997

1/97 **Eksemplarfremstilling av litterære verk til privat bruk**
Therese Steen ..... *NOK 158.-*

2/97 **Offentlige anskaffelser av informasjonsteknologi**
Camilla Sivesind Tokvam ..... *NOK 175.-*

3/97 **Rettslige spørsmål knyttet til Oppgaveregisteret**
Eiliv Berge Madsen ..... *NOK 170.-*

4/97 **Private pengespill på Internett**
Halvor Manshaus ..... *NOK 160.-*

5/97 **Normative Structures in Natural and Artificial Systems**
Christen Krogh ..... *NOK 255.-*

6/97 **Rettslige aspekter ved digital kringkasting**
Jon Bing ..... *NOK 178.-*

7/97 **Elektronisk informasjonsansvar**
Tomas Myrbostad *NOK148.-*

8/97 **Avtalelisens etter åndsverksloven § 36**
Ingrid Mauritzen ..... *NOK 120.-*

9/97 **Krav til systemer for forvaltning av immaterielle rettigheter**
Svein Engebretsen ..... *NOK 168.-*

10/97 **American Telephony: 120 Years on the Road to Full-blown Competition**
Jason A. Hoida ..... *NOK 140.-*

11/97 Rettslig vern av databaser
Harald Chr Bjelke.....*NOK 358.-*

1996

1/96 Innsynsrett i elektronisk post i offentlig forvaltning
Knut Magnar Aanestad og Tormod S. Johansen.....*NOK 218.-*

2/96 Public Policy and Legal regulation of the Information Market in the Digital Network Enviroment
Stephen John Saxby .....*NOK 238.-*

3/96 Opplysning på spill
Ellen Lange.....*NOK 218.-*

4/96 Personvern og overføring av personopplysninger til utlandet
Eva I. E. Jarbekk .....*NOK 198.-*

5/96 Fjernarbeid
Henning Jakhelln .....*NOK 235.-*

6/96 A Legal Advisory System Concerning Electronic Data Interchange within the European Community
Andreas Mitrakas.....*NOK 128.-*

7/96 Elektronisk publisering: Utvalgte rettslige aspekter
Jon Bing og Ole E. Tokvam .....*NOK 186.-*

8/96 Fjernsynsovervåking og personvern
Finn-Øyvind H. Langfjell.....*NOK 138.-*

1995

1/95 Rettslige konsekvenser av digitalisering: Rettighetsadministrasjon og redaktøransvar i digitale nett
Jon Bing.....*NOK 368.-*

2/95 Rettslige spørsmål i forbindelse med utvikling og bruk av standarder innen telekommunikasjon
Sverre Sandvik .....*NOK 178.-*

3/95 Legal Expert Systems: Discussion of Theoretical Assumptions
Tina Smith ..... *NOK 278.-*

4/95 Personvern og straffeansvar – straffelovens § 390
Ole Tokvam ..... *NOK 198.-*

5/95 Juridisk utredning om filmen «To mistenkelige personer»
Johs. Andenæs ..... *NOK 138.-*

6/95 Public Administration and Information Technology
Jon Bing and Dag Wiese Schartum ..... *NOK 348.-*

7/95 Law and Liberty in the Computer Age
Vittorio Frosini ..... *NOK 158.-*

1994

1/94 Deon'94, Second International Workshop on Deontic Logic in Computer Science
Andrew J. I. Jones & Mark Sergot (ed) ..... *NOK 358.-*

2/94 Film og videogramrett. TERESA (60)
Beate Jacobsen ..... *NOK 318.-*

3/94 Elektronisk datutveksling i tollforvaltingen – Rettslige spørsmål knyttet til TVINN
Rolf Risnæs ..... *NOK 225.-*

4/94 Sykepenger og personvern – Noen problemstillinger knyttet til behandlingen av sykepenger i Infotrygd
Mari Bø Haugestad ..... *NOK 148.-*

5/94 EØS, medier og offentlighet. TERESA (103)
Mads Andenæs, Rolf Høyer og Nils Risvand ..... *NOK 148.-*

6/94 Offentlige informasjonstjenester: Rettslige aspekter
Jon Bing *NOK148.-*

7/94 Sattelittfjernsyn og norsk rett. MERETE (3) IV
Nils Eivind Risvand ..... *NOK 138.-*

8/94 Videogram på forespørsel. MERETE (14) IV
Beate Jacobsen (red) ..... *NOK 158.-*

9/94 **«Reverse engineering» av datamaskinprogrammer. TERESA (92) IV**
Bjørn Bjerke.....*NOK 198.-*

10/94 **Skattemessig behandling av utgifter til anskaffelse av datamaskinprogrammer. TERESA (75)**
Gjert Melsom.....*NOK 198.-*

1993

1/93 **Artificial Intelligence and Law. Legal Philosophy and Legal Theory**
Giovanni Sartor.....*NOK 148.-*

2/93 **Erstatningsansvar for informationstjenester, særlig ved databaseydelser**
Connie Smidt.....*NOK 138.-*

3/93 **Personvern i digitale telenett**
Ingvild Hanssen-Bauer.....*NOK 178.-*

4/93 **Consumers Purchases through Telecommunications in Europe. – Application of private international law to cross-border contractual disputes**
Joachim Benno.....*NOK 198.-*

5/93 **Four essays on: Computers and Information Technology Law**
Morten S. Hagedal.....*NOK 218.-*

6/93 **Sendetidsfordeling i nærradio MERETE (3) III**
Marianne Rytter Evensen.....*NOK 148.-*

7/93 **Essays on Law and Artificial Intelligence**
Richard Susskind.....*NOK 158.-*

1992

1/92 **Avskriving av mikrodatamaskiner med tilbehør – en nordisk studie TERESA (87)**
Beate Hesseltvedt.....*NOK 138.-*

2/92 **Kringkastingsbegrepet TERESA (78)**
Nils Kr. Einstabland.....*NOK 208.-*

3/92 Rettskilderegistre i Helsedirektoratet NORIS (94) I & II
Maria Strøm.....*NOK 228.-*

4/92 Softwarepatent – Imaterialrettens enfant terrible. En redegjørelse for patenteringen af softwarerelaterede opfindelser i amerikansk og europæisk ret
Ditlev Schwanenfügel.....*NOK 158.-*

5/92 Abonnementskontrakter fro kabelfjernsyn TERESA (78II)
Lars Borchgrevink Grindal.....*NOK 248.-*

6/92 Implementing EDI – a proposal for regulatory form
Rolf Riisnæs.....*NOK 118.-*

7/92 Deponering av kildekode«escrow»-klausuler TERESA (79)
Morten S. Hagedal.....*NOK 128.-*

8/92 EDB i juridisk undervisning – med en reiserapport fra England og USA
Ola-Kristian Hoff.....*NOK 228.-*

9/92 Universiteters ansvar for bruk av datanett TERESA (94)
Jon Bing & Dag Elgesem.....*NOK 198.-*

10/92 Rettslige sider ved teletorg
Andreas Galtung.....*NOK 148.-*

BESTILLING

Jeg bestiller herved følgende CompLex-utgivelser:

Nummer / årgang: _____
Tittel: _____

Nummer / årgang: _____
Tittel: _____

Nummer / årgang: _____
Tittel: _____

Nummer / årgang: _____
Tittel: _____

Nummer / årgang: _____
Tittel: _____

Navn: _____
Adresse: _____
Postadresse: _____
Telefon: _____

Bestillingsskjemaet sendes pr.post eller telefaks til:

akademika
Fagbokhandelen i Oslo

Avd. juridisk litteratur Aulabygningen
Karl Johansgt. 47, 0162 Oslo
Telefon: 22 42 54 50
Telefaks: 22 41 17 08
CompLex kan også bestilles via nettbokhandelen www.akademika.no