Summary

Sure about searching?
Towards a legal framework for automated search systems searching for open content on the Internet

PART A
Introduction and framework

The amount of electronic information accessible via the Internet is increasing rapidly. Search engines and other electronic means for searching for information on the Internet (hereafter: ‘search systems’) search through hundreds of exabytes of content available on the Internet to find information relevant to an information request.

This dissertation focuses on issues specifically related to open content, i.e. content that is freely accessible to the public and not protected by technological security measures such as passwords. Examples of such open content websites are http://www.wikipedia.nl, http://www.youtube.com. Accessibility of open content for the public, and therefore also accessibility of open content to search systems, is of importance to both content providers and ‘content consumers’ (people who or systems that request content from content providers). The accessibility of that content to search systems is thus also of importance, as search systems must first request and process open content before they can refer to that content in their search results. Content that is not accessible to search systems is not displayed in the search results of a search system and is therefore very difficult for customers to find on their own. However, not only the provisioning of information is an issue but also its use. Providers of open content also want to protect their content against undesired use. If the technological protection of information within a system is insufficient, providers may decide to protect their content with a password or other types of credentials. That content is then ‘closed content’, i.e. content that is less accessible to the public. There is a conflict between the interests of content providers and of content consumers: On the one hand, it is through search engines that the Internet provides access to information. The Internet has become an important source of information. On the other hand, content providers need to protect their content against unwanted kinds of use. This social issue results in a legal issue: ‘How can content providers and content users determine which actions of search systems with open content are allowed in a specific situation?’ This thesis examines this issue from different angles.
In answer to the core question of this dissertation, I first discuss how the Internet works. I also explain how search systems play a role in finding relevant information (chapter 2). I then discuss the extent to which the application of existing rules of copyright and database law clearly establishes the actions a search systems may perform in specific cases (chapters 3 and 4). Finally, I examine whether contractual means provide the necessary clarity on permissibility of actions with respect to open content to both content providers and content consumers (chapter 5 to 8).

The extensive discussion of the technical and functional operation of the Internet and search systems in chapter 2 identifies problems that occur in practice with search systems. The description and analysis of the Internet and the environment in which the provider of content, the end user, and the operator of the search system operate, results in a number of assumptions about the operation of the Internet and search systems. For example, the analysis shows that from a technical perspective, viewing a web site requires a content consumer to first send a request for content to a content provider. In other words, viewing a web site is only possible if a content provider has first sent content to a consumer. A content provider must therefore send content to search systems before they can refer content consumers to the content. Another important issue is that it is very difficult for content providers to ensure that content can be requested by end users, but not by search systems. Legal measures rather than technical measures are needed to regulate open content by search systems. As current search systems are unable to process natural language, it is not possible for content providers to place statements in natural language on their web sites to this purpose. In chapter 2, I devise a four-phase model for the interaction between content providers, (operators of) search systems and end users. In phase 0 of the model a content provider makes content available. In phase 1, a search system collects content. The search system requests the computer of the content provider to send content. This is accomplished using so-called search bots. This content is then internally processed (‘indexed’) by the search system (phase 2). Only then can the search system be made available to the public (phase 3). End-users enter search terms to use the search system, for example at http://www.google.nl. The search system then displays search results. In phase 4 end users click on a search result to view the content on their computer screen. This dissertation focuses on the relationship between content providers and (operators of) search systems.

PART B
Search systems and protection by copyright and database right

For the public to effectively use a search system, search systems may have to perform actions that are reserved to the copyright holder or the rightful claimant to a database right. For the application of search systems it is therefore important to know in advance for which of these actions authorisation of the right holder is required. Therefore, I have researched whether the Dutch system of copyright and database rights, as laid down in the Dutch Copyright Act and Database Act, which are implementations of the EU Copyright Directive 2001/29/EC and the EU Database Directive 96/9/EG, provides a sufficient basis for the assessment of
whether consent is required for the actions of a search system in specific cases. To this purpose, I have investigated to what extent authorisation is required for requesting content, for internal processing of content and presentation of content in search results.

Application of the regulations of the Copyright Act and the Database act does not provide clear answers with regard to the actions of search systems. Despite recent case law of the EU Court of Justice, there are still many questions with regard to the application of rules of copyright and database law to the actions of search systems. More specifically, I identify a lack of clarity with regard to the following issues:

– Can authorisation of the copyright holder be required for the mere consultation of open content by search systems, simply because the content is being requested with a search system?
– Under which circumstances does requesting content from a database qualify as requesting substantial or non-substantial parts as indicated in article 2 Database Act? If the actions of the search system qualify as ‘repeated and systematically’ requesting non-substantial elements of a protected database, it must also be established whether such actions cumulate to requesting or reuse of a substantial part of the protected database. Whether this is the case, largely depends on specific facts and circumstances. This therefore remains very hard to predict in advance.

With regard to the internal processing of requested content with a search system, the following questions remain unanswered after applying the rules of copyright and database law:

– What are the requirements for (temporary) copies made by search systems for internal processing of content for them to qualify as temporary copies as determined in article 13a Copyright Act? The answer partly depends on the meaning of a number of definitions in the Copyright Directive. The precise meaning of these definitions is, as yet, unclear. Additionally, the outcome partly depends on the application of the so-called Three Step Test, which may have different result for each particular situation.
– Under which circumstances does the internal processing of content from a database qualify as reuse of a substantial part, or repeated and systematic reuse of non-substantial parts of a database? Whether authorisation is required for internal processing of content also depends on the method used by a specific search system. The outcome is therefore very difficult to predict in advance.

With regard to presentation of content by search systems, the following questions are not fully addressed by copyright and database law:

– To what extent can the display of search results in a specific case be classified as an exemption for quotations of article 15b Copyright Act? If this exemption is applicable, it remains unclear whether it only applies to short summaries and ‘ordinary’ hyperlinks, or also to thumbnails.
– Under which circumstances does the presentation of search results from a database qualify as repeated and systematic reuse of non-substantial parts
of that database? The so-called two-step test must be applied to answer this question. The result of this two-step test mainly depends on the specific circumstances. Additionally, it is still unclear which circumstances are relevant in the case of search systems. All in all, the results of such a test are very unpredictable for the operator of a search system.

The question arises to what extent making content available as open content entails that (implicit) consent is granted for certain actions with that content. Arguments that support this claim refer to the central role that search systems play in the process of finding information on the Internet. Content providers voluntarily return content upon request to all content consumers, including search systems. Neither the Dutch Copyright Act, nor the Dutch Database Act provide practical handles to evaluate the scope of implied consent.

Providers and users of content who want to know beforehand which actions of search systems are permitted, do not find sufficient answers in Dutch copyright and database law. Despite the broad scope of application of copyright and database regulation, it is, as yet, insufficiently clear which actions with content on the Internet are allowed in specific situations. This is particularly the case with actions of search systems with regard to open content. The application of Dutch copyright and database law therefore leads to legal uncertainty.

PART C
Contract law

In part C, I investigate whether uncertainties with regard to the actions of search systems are better addressed by a contractual approach. In particular, I examine whether it is possible within the existing framework of regulations to make arrangements for open content that are more tailored to the needs of content providers and content consumers.

In chapter 5, I provide reasons for a contractual approach. I also introduce several ways of attaching conditions of use of open content that are currently used by content providers. A contractual approach may provide more certainty with regard to the actions that are allowed with open content in a specific case. Contracts provide a flexible tailored solution for specific situations that are complex and diverse. For operators of search systems it is particularly important to know in advance which restrictions content providers impose on requesting, internal processing and the presentation of content. Otherwise, it would be impossible to refer end users to relevant content without going against the content providers’ rights and wishes.

Contracts between parties can be useful in addition to the execution of exclusive rights. Parties may also remove uncertainties regarding the scope of copyright and database law by concluding a contract with provisions on what actions of content consumers, in particular search systems, are allowed with the content. In my opinion, a contractual approach is also useful if it is not evident that copyright or database law apply.
Chapter 6 provides the legal framework for concluding contracts by electronic means. For every situation, the actions of each of the parties must be reviewed to determine whether their conduct can be regarded as an offer or an acceptance. Of course, the normal rules for entering into a contract are also applicable in an electronic setting. Nevertheless, some of the criteria in relation to contract formation must be explained slightly differently in the context of contracting by electronic means. For example, the application of the Dutch 'nuanced recipient principle' (art. 3:37 lid 3 Dutch Civil Code) in an electronic setting determines that the moment a message is received coincides with the moment that the content arrives on the computer system of the content consumer. As a result, there is no room for an exception to the 'nuanced recipient principle' if (part of the) statement is displayed on a webpage in a very small font or is otherwise almost invisible. In my opinion, an exception is possible if a statement has not reached the content consumer because he made use of a search bot of a search system and the communication was not suitable for search bots, while the content provider knew or should have known that the content consumer would use a search bot. In that case, a statement does not reach the content consumer at the same time that it enters his computer system.

In my opinion, articles 6:227b and 6:227c of the Dutch Civil Code do not apply to agreements which relate only to actions with regard to the use of content and for which no financial contribution is required.

With regard to question whether an agreement is reached between the parties in a specific situation, the reasonable expectations of parties are decisive. I have therefore made an inventory of viewpoints that may be relevant for a normative assessment of the measure of 'the reasonable expectations of parties'. As the circumstances of the case are decisive, this inventory is of an indicative nature.

In chapter 7 and 8, I examine whether a valid contract with regard to the actions of search systems with open content is possible in each of the situations mentioned in chapter 5.

Chapter 7 deals with the question to what extent it is possible to conclude an implicit agreement with regard to the use of open content. If a content provider does not explicitly state the conditions of use, in principle, no agreement can come into being. The mere fact that open content is made available via the Internet does not imply that the content provider wants to make an offer. It is not clear which conditions are part of such an offer.

In chapter 8, I examine whether the identified ways of attaching conditions to open content in chapter five can result in the conclusion of a valid contract. In addition to the offer to enter into a click-wrap agreement, which relates to closed content, the following examples of attaching conditions to open content are examined:

- the placing of a clickable notice on a web site that use of content is subject to conditions;
- the placing terms of use on all pages of a web site with open content;
- the reference on every page to conditions of use that are displayed on one page of an open content web site;
– the addition of a no robots clause according to the Robots Exclusion Standard on the website;
– the addition of an ACAP.txt-file to the website; and
– a combination of no robots clause and conditions in natural language.

For each of these situations, I investigate whether a valid offer can be constructed that is implicitly accepted by the user, whether the user can implicitly accept an 'offer' by simply continuing to request web pages. For this investigation, the legal qualification of the conduct of parties during the process of interaction needs to be understood. I have determined this qualification both for the situation that the content consumer is an end user and for the situation that the content is retrieved using search bots of search systems. The situations included in the analysis concerned not only terms of use in natural language (so-called browse-wrap agreements), but also terms in machine interpretable format.

Many ways in which content providers currently add terms of use to their content, do not result in a valid agreement between content provider and web site user. Only in a limited number of cases an agreement is possible. For the conclusion of an agreement with regard to the actions of end-users with open content, a clickable notice is the best way. However, this method is unattractive for content consumers, because it makes the content less accessible. Therefore, this method is unlikely to be used very often.

A second possibility to get terms of use across to users, is to add short notices on each webpage of a web site. If the notice is extensive enough to constitute a determinable offer, such notices can be offers in the sense of art. 6:217 Dutch Civil Code. For example, the following notice can constitute an offer: ‘by using this web site, you agree that X is not liable for any damages related to the use of this web site’. These notices are often presented in a rather inconspicuous manner, for example at the bottom of a web page. If a statement is difficult to find for users, content providers may not easily assume that the continuing use of a website constitutes the acceptance of the offer by the content consumer. The same holds for the third possibility I investigate: adding a hyperlink with a reference to a page with the actual terms of use to each web page. Here, the lack of visibility and therefore lack of ‘knowability’ can also hinder the providers’ justified reliance in the acceptance of the terms by the user as meant in article 3:35 of the Dutch Civil Code. The provider must ensure that terms and reference are sufficiently recognisable to content consumers. A link must make the provider’s conditions clear.

Within certain boundaries, a content provider can conclude contracts with end users with regard to the use of open content. Although the same rules apply to actions of search systems, the outcome of the application of the rules is different for search engines. When a provider attaches terms to the use of the content or of the infrastructure with regard to the actions of search bots, the measures mentioned above are not as adequate. Most ‘clickable notices’ cannot be processed by search bots. Therefore, search bots can either not access web sites with a clickable notice, or they are unaware of such notices and therefore ignore them. Other expressions of a legal nature in natural language cannot be processed by search bots either. A user of a search bot can not recognise that the
expressions of the content provider constitute an offer directed at the conclusion of a contract. Any intentions of content providers to conclude a contract simply do not come across to the user of a search bot. None of the statements of the content provider reach the user of search bots in the sense of article 3:37 lid 3 Dutch Civil Code. As a content provider who sets terms of use regarding behaviour of search bots should know that any messages formulated in this way will never reach the user of a search bot, the risk for non-arrival of the statement must (normally) be attributed to the content provider.

Even if a website user can be considered to have received the statement that a content provider intends to make his content available under terms of use, it is unlikely that a contract comes into being in the same way as between a human user and a content provider. If the conditions of use are presented in natural language, but also refer to actions of search bots (e.g. limitations of use of the website for search bots), any statement in natural language will generally be insufficient to assume ‘justified reliance’ in the sense of article 3:35 of the Dutch Civil Code. This means the content provider may not have ‘justified reliance’ that continuing to request the web page meant that the offer has been accepted by the website user.

Machine interpretable statements provide a means for a provider to specify the actions with which he concedes, for example, for agreements concerning practical aspects of the use of search bots. The Robots Exclusion Standard for robots. txt-files does provide machine interpretable statements, but these statements are not of a legal nature. The ACAP-protocol has been designed for agreements of a legal nature. Adding an ACAP.txt-file does qualify as an offer, but is only directed to users of search bots that are programmed to process ACAP-files. An agreement may not easily be assumed if the content provider uses a combination of natural language and no robots clauses.

PART D

Conclusion

Application of copyright and database law is not sufficient. Legal uncertainty with regard to the question what search systems are allowed to do with open content, remains. Contracts that indicate which actions content consumers are allowed to perform with open content, makes parties less dependent on ambiguities regarding the scope of copyright and database law. Within certain limits, a contract with regard to the use of open content by human users can come into being. This is not necessarily the case when content is requested with search bots of search systems. Since agreements on actions of search systems with regard to open content are important for both content providers and content consumers, alternative ways are needed to conclude legally valid contracts. The ACAP-protocol and the system of Creative Commons licenses create interesting starting points. The legal framework outlined in this dissertation provides a legal basis to start working on standardisation of (references to) conditions of use of search systems.