Agent-Based Information Retrieval:
Legal and Technical Considerations in a Simple Case

Frances M.T. Brazier[1], David R.A. de Groot[1],
Anja Oskamp[2], and Niek J.E. Wijngaards[1] (in alphabetical order)

[1] IIDS Group, Faculty of Sciences, Vrije Universiteit Amsterdam,
de Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands
Phone: +31 - 20 - 444 7737; Fax: +31 - 20 - 444 7653
{frances, davidra, niek}@cs.vu.nl
http://www.iids.org/
[2] Computer/Law Institute, Faculty of Law, Vrije Universiteit Amsterdam,
de Boelelaan 1105, 1181 HV, Amsterdam, The Netherlands
Phone: +31 - 20 - 444 6215; Fax: +31 - 20 - 444 6230
a.oskamp@rechten.vu.nl
http://www.rechten.vu.nl/~CLI

Abstract. Information retrieval is currently one of the most important usages of the Internet. Scientific publishers are one of the sources of information, providing access to digital libraries on a subscription basis. Contracts between publishers and larger organizations specify how access is to be regulated. This paper explores the legal implications and technical considerations of the use of software agents to facilitate on- and off-campus access to the ScienceDirect digital library.

1 Introduction

Almost everyone requires information for work and private activities. More and more people regard the Internet as an important source of information. For example, in the scientific community many scientists use web-pages, newsgroups and E-mail to remain well-informed about their research areas. Important activities include searching for new publications, monitoring research progress and keeping track of citations to their own publications.

Publishers, such as IEEE (http://www.ieee.org/), ACM (http://www.acm.org/), Elsevier (http://www.elsevier.com/), Kluwer (http://www.wolters-kluwer.nl/), MIT Press (http://mitpress.mit.edu), and Springer (http://www.springeronline.com/), recognize these needs and provide online digital libraries. Other organizations, e.g. DBLP (http://dblp.uni-trier.de/), CiteSeer (http://citeseer.ist.psu.edu/) and Ingenta (http://www.ingenta.com), offer additional services, such as maintaining indices on a per-author basis. Some of these organizations offer public access to digital versions of publications, while others provide such access on a subscription basis.

Subscribers can be organizations themselves, such as universities and research institutions. Usually, a subscription applies to an entire organization, including for a university, for example, access for employees, students and guests. This paper explores the options provided by the contract that the Vrije Universiteit Amsterdam (VUA) has with Elsevier for access to content on the ScienceDirect[1] website (http://www.sciencedirect.com). The contract regulates access for employees and students. The access policy, however, is 'campus-based': off-campus access is limited. Agent-based software could offer new possibilities for functionality and access.

Agent technology is a promising and enabling technology [1], not only useful for e-commerce applications, but also for information retrieval applications. This paper explores the roles software agents can play in the context of this specific contract, given current technology. The cases presented here provide a setting to explore legal and technical issues in deploying software agents in the real world. In particular, this paper focuses on possibilities and consequences of the use of software agents. While other studies on electronic libraries mainly focused on users and usage statistics [2,3,4,5], this study focuses on access methods and contractual arrangements. A number of issues, related to privacy, copyright, reputation, trust, integrity of software agents, etc. are not included in this paper, for reasons of brevity and clarity. These aspects are considered within the ALIAS project [6] in the context of which this paper has been written. The ALIAS project, investigates legal implications of the use of agent systems from both a legal and technological perspective.

This paper presents the current usage of the ScienceDirect digital library, in the context of the VUA's contract with Elsevier, and analyses possibilities with respect to automated access involving software agents. First, usage scenarios involving humans, with and without software agents, are presented. Next, legal and technical issues in the context of these scenarios are addressed. Finally, a brief conclusion and recommendations are provided.

2 Science Direct Digital Library

ScienceDirect is probably the world's largest electronic collection of science, technology and medicine full text and bibliographic information. Since its launch in 1997, ScienceDirect has evolved from a web database of Elsevier journals to one of the world's largest providers of scientific, technical and medical literature. ScienceDirect is available to institutions, ranging from single sites to global organizations. ScienceDirect consists of a public part, often containing indices to scientific journals, and a paid-for part, often concerning downloadable articles.

The VUA and ScienceDirect have different needs, as the VUA strives for full access for its employees and students to stimulate education and scientific research, while Elsevier needs to enforce copyright protection and has an obvious financial interest. Additionally, being read and cited is of importance to both the authors and the publisher.

---

[1] ScienceDirect® is a registered trademark of Elsevier B.V.

2.1 Contractual arrangements

The contract between the VUA and ScienceDirect is *not* open for public inspection. However, based on information received from the VUA Library employees and general terms and conditions on ScienceDirect's website, it is clear that this contract specifies that:

- Responsibility for the administration of employees and students who can access ScienceDirect via the VUA network environment lies with the VUA;
- Specific access methods are permitted, e.g. www-browser and alternative access platforms;
- The VUA is aware of the accepted interpretation of ScienceDirect's 'fair use' policy, including consequences for transgressions;
- The VUA is aware of the rights of employees and students with respect to downloading, on-screen viewing and printing of publications for their own use, scientific activities and education;
- Reasonable security measures are to be maintained by the VUA;

Additionally, financial arrangements are in place, which are not of our current concern.

2.2 Technical considerations

From a technical point of view, the stipulations concerning access control are interesting. The campus-based subscription for institutions and organizations often implies the use of IP-safe lists. An IP-number is a globally unique address of a computer or an Intranet, usually administered by a specific organization. An organization typically uses multiple IP-numbers. The uniqueness property is an important argument to use IP-numbers for access control. The effect is that from any computer with a valid IP number a user can access information on the ScienceDirect website, e.g. to download articles. ScienceDirect cannot know whether a person using a valid computer is an employee or an outsider, this authentication policy is delegated to the subscribed institution. It should be noted that techniques exist to use fake IP-addresses, so called 'IP-spoofing', but these require significant effort and are detectable.

It is acknowledged that ScienceDirect maintains logs of actions originating from its subscribers. For example, usage statistics can be maintained, as well as the distribution of usage over time, per journal and/or for specific IP-numbers. This information may be used to determine invalid usage, such as downloading each and every downloadable article.

3 A Case for Software Agents

In this section current usage scenarios are described and possibilities for future usage with software agents are sketched. A number of scenarios are provided in which the need for off-campus access is addressed. First scenarios are presented in which only

humans partake, then scenarios in which both humans and software agents play a role. The agent-based scenarios are kept realistic by assuming software agents are deployed at the VUA; ScienceDirect need not be changed.

For each scenario a graphical representation is used, in which computer-symbols represent individual computers, grouped into domains, arrows denote authenticated communication channels, and smiley faces represent software agents. A puppet figure denotes a human user (a VUA employee or student) involved in (in)-directly accessing ScienceDirect.

In this context, agents are software programs, running on a computer as processes or thread-groups consisting of code and data. Software agents need an environment to become operational: an agent platform. Software agents are able to communicate with other agents, may or may not migrate to other agent platforms, and interact with objects (e.g. web-pages) and services on the Internet: an agent platform supports these actions, hiding complexities at operating system level from the agent developer.

3.1 Scenarios Involving Human Users

Two scenarios involving solely human users are described based on the on-campus or off-campus location of a user. These scenarios are depicted in Figure 1.
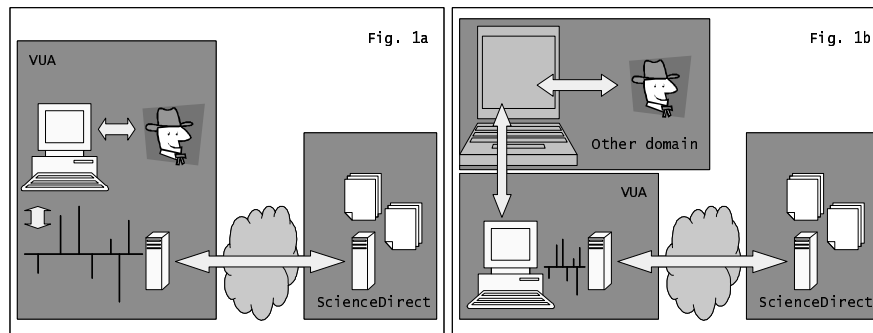

Fig. 1. Graphical representation of Scenario 1 and 2:
on-campus and off-campus access, involving only humans.

Scenario 1: Normal on-campus usage. As illustrated in Fig. 1a, a user resides on-campus, logs in to a computer on the local campus network. When using a web browser (e.g., Internet Explorer or Mozilla) the user has full access to ScienceDirect.

Scenario 2: Normal off-campus usage. Quite a number of employees have a need for off-campus access to ScienceDirect when they are at an international conference or work from home. However, computers residing in a different network, and therefore having an IP-address that is not covered in the VUA contract, cannot be used to directly access ScienceDirect. To facilitate off-campus access, ScienceDirect provides login-names and passwords, which an institution's library personnel can provide to individual employees. This is, however, not practiced at the VUA. The VUA offers a few different mechanisms, which are guaranteed to be usable only by its employees and students: the VU-ACCESS dial-up connection and the UBVU proxy server. The VU-ACCESS dial-up connection uses caller-ID checks in addition to common login and password combinations to verify the identity of an employee or student. Via this dial-up Internet connection employees and students can extensively use ScienceDirect. Actually, the computer of the user becomes part of the domain of the VU, thereby realizing normal on-campus access as illustrated in Fig. 1a. However, in many of these situations they cannot make use of the VU-ACCESS services, e.g. because their local telephone number is not registered for dial-up verification, or because their computer is connected to another (broadband) network. Via the UBVU proxy server students and staff are allowed access to e-journals and e-resources licensed by the library from a remote location by means of a VU-net-ID and password. The proxy server was recently introduced and stands between the user's computer and a remote server, e.g. ScienceDirect, providing an internet address that the server recognizes as part of the UBVU domain and sending remote data back to your browser. Using the proxy server requires one to set up your browser for the correct type of proxying.

Normal off-campus usage is depicted in Fig. 1b, a user resides at an off-campus location, e.g. at home or at an international conference. The user logs in at a local computer, located in a different network, and sets up a connection to an on-campus computer (which acts as a server), using his or her VUA login-name and password. Via this server, the user can access ScienceDirect.

Additional techniques to acquire such a connection from any remote computer, apart from the UBVU proxy server connection, include techniques to run a web-browser on a VUA computer and interact with its user interface on another computer. For example with an SSH (e.g., http://www.openssh.org) connection to establish a secure 'tunnel' via which a virtual network connection (e.g., http://realvnc.com) can be setup or an X-connection with Exceed (Hummingbird copyrighted). An alternative is to setup a SSH connection in combination with the text-only browser Lynx (http://lynx.browser.org). Note that in these cases the actual processes involved in accessing ScienceDirect (e.g., a web-browser) are run on VUA computers.

3.2 Scenarios Involving Human Users and Software Agents

Three scenarios involving both humans and software agents are described. First an on-campus scenario is presented for reference, after which two off-campus scenarios are presented, in one of which features mobile agents. Figure 2 depicts scenario 3 and 4 in respectively Fig. 2a and Fig. 2b.
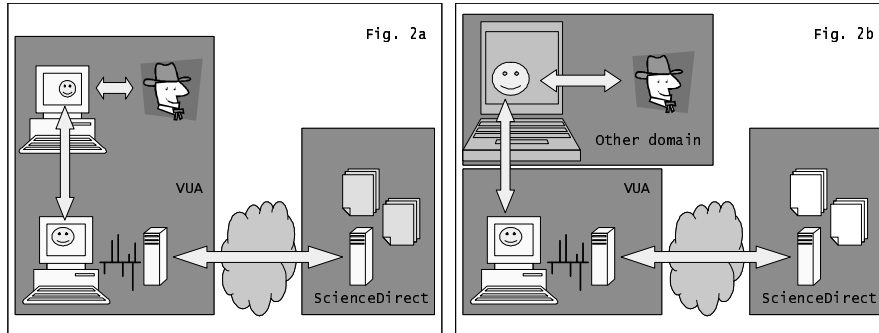
Fig. 2. Graphical representation of Scenario 3 and 4:
on-campus and off-campus access, involving both humans and software agents.

Scenario 3: Agent-based on-campus usage. A user resides on-campus and logs in to a computer. The user starts a software agent with a request, and the software queries the ScienceDirect website. This scenario is comparable to using a web-browser to access ScienceDirect. This scenario is illustrated in Fig. 2a.

Scenario 4: Communicative agent-based off-campus usage. A user resides at an off-campus location, and logs in at a local computer. The user starts a software agent with a request at this computer. This software agent contacts, via an authenticated communication channel, another software agent residing on an agent platform on an on-campus computer. This latter agent relays queries to ScienceDirect, and responses to the user's agent. See also Fig. 2b.

In this scenario, depending on the communication policies used, the user's software agent can remain online (i.e., in connection with the VUA-based agents), or go temporarily offline. When the user's agent connects with the VUA-based agents at a later time, results from its request may have arrived. In the off-line situation, extensive querying and browsing of ScienceDirect could be delegated to VUA-based agents, in the on-line situation each VUA-based agent may act as a 'proxy' (middleman).
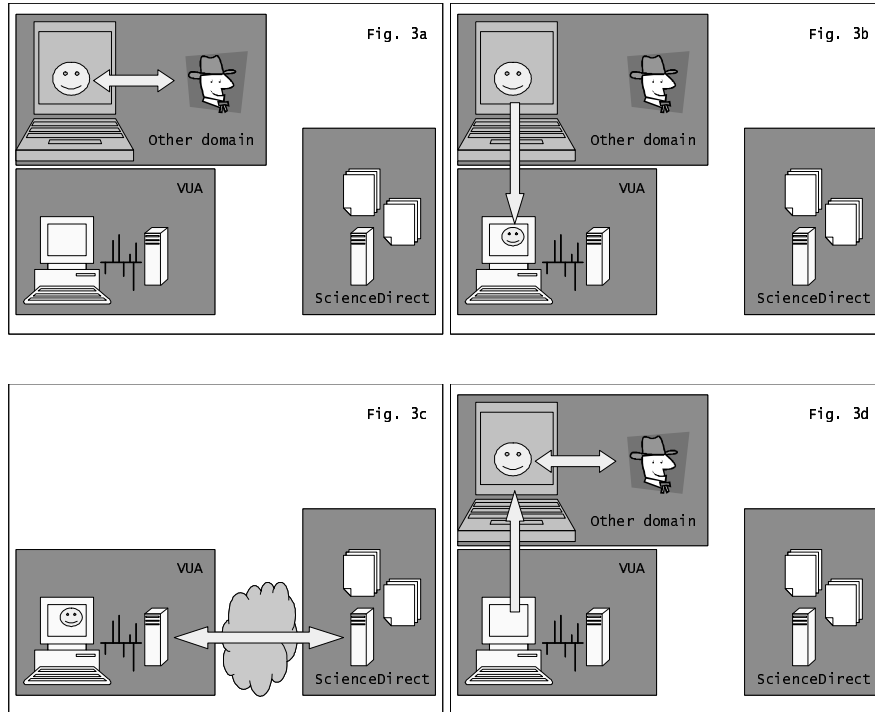
Fig. 3. Graphical representation of Scenario 5: agent-based off-campus access, involving both humans and migrating software agents.

Scenario 5: Migratory agent-based off-campus usage. Figure 3 depicts this scenario in four steps. A user resides at an off-campus location, and logs onto a local computer. The user starts a software agent with a request on this computer (Fig. 3a). This software agent contacts a VUA-based agent platform and starts a migration procedure (Fig. 3b), involving authentication of the agent (and possibly its user). When the user's software agent has arrived at the VUA-based agent platform, it has full access to ScienceDirect (Figure 3c). Meanwhile the user can log of at the computer. At a later time, the user can log on to (possibly a different) computer and the user's software agent can send a message with results or migrate back to the user's (current) computer carrying results (Figure 3d).

4   Discussion

The scenarios described in the previous section outline possible consequences of the use of software agents. In this section the scenarios are briefly analyzed, after which opportunities of agent-based approaches are sketched.

4.1 Analysis

The first two scenarios, involving only humans, include strong authentication: only employees and students can log in to on-campus computers. Scenario 1 describes normal on-campus usage according to the VUA contract. Scenario 2 describes current off-campus practices, in which techniques are employed that are reasonable secure and only usable by employees and students, and involve the use of web-browsers. Scenario 2 is thought to be in compliance with the VUA contract.

In both scenarios, the user is able to view, print and save digital publications for later use. The web-site terms and conditions prohibit long-term storage, reselling or providing access to publications for third parties. VUA policies require users to safeguard their login-names and passwords from use by third parties. Users may make heavy use of ScienceDirect by downloading large numbers of publications. The use of agents does not change these consequences.

When the first scenario, involving human on-campus access, is augmented with software agents in Scenario 3, software agents replace the role played by the human's web-browser. Is this legal use of ScienceDirect according to the VUA contract? If not, then the other agent-based scenarios are probably also in violation of the contract. Since the contract allows access from alternative platforms and even expresses the willingness of ScienceDirect to cooperate and assist in the development, it can be concluded that the legality of using software agents in this context is no longer questionable.

Scenario's 4 and 5 are agent-based extensions of Scenario 2, in which a human resides off-campus. Scenario 4 involves a non-mobile, communicative agent and resembles Scenario 2, in which other techniques are used to access ScienceDirect via on-campus computers. In case that software agents with authentication are used, and not web-browsers, we consider this form of access to be in compliance with the demands of reasonable security in the VUA-ScienceDirect contract.

Scenario 5 is a more advanced than Scenario 4, and involves mobile agents that may obtain downloaded publications for a while, and transporting them to their users. In this scenario, multiple devices, such as laptops, mobile telephones, etc., may be involved: the user is not restricted to a specific computer. A user could launch a software agent from a mobile telephone, and later receive the agent with downloaded articles on his or her laptop. Although this scenario facilitates off-campus access with possibly multiple (mobile) devices, involving agent authentication, issues concerning copyright and caching of publications may play a role.

Actually, ScienceDirect cannot determine whether humans with web-browsers or software agents are used to access ScienceDirect, as in both cases they originate from (valid) on-campus computers. In these scenarios, the software agents are to use the http-protocol, as web-browsers do. This implies that the software agents need to understand the structure of the ScienceDirect website. When ScienceDirect monitors, e.g. frequency of usage and speed of browsing in sessions, it may conclude that access is automated, which may be prohibited in their contract.

An important element of ScienceDirect's contractual policies concerns authentication. The VUA is responsible of keeping an administration of employees and students authorised to fully access ScienceDirect, by controlling access to VUA's computers.

Software agents will need to comply with these policies, which influence the use of authentication mechanisms between off-campus agents and on-campus computers, agent platforms, and agents. Whether these mechanisms need to be 'stronger' than the mechanisms for human authentication (commonly involving login-names and passwords), is, as yet, unclear. An implication may be that a closed agent system needs to be developed, in which all agents are known beforehand, and their roles and actions are strictly regulated. An open agent system, involving agents from unknown sources, may not currently be feasible given the current state of technology with respect to security mechanisms.

4.2 Agent-based opportunities

Software agents offer business and scientific opportunities for all involved parties. ScienceDirect may make its products more attractive by providing new facilities for (remote) automated access. Possible agent-based functionality includes that employees could train their own agents with a set of publications to find similar publications by searching through the full text of publications in ScienceDirect's database. Also, updates on changes to the database could be sent to agents, which filter this information for relevance for their users. ScienceDirect may even consider providing a local agent platform for agents to query their databases. Such facilities can influence the ease of finding (relevant) publications, and therefore provide more revenue to ScienceDirect as well as increase citation possibilities for authors.

The VUA may benefit from gaining more control on the use of ScienceDirect, e.g. by enforcing access policies and volume-restrictions on agent-based access. In addition, VUA may offer different services than other universities. VUA employees and students may benefit by acquiring 24/7 access to ScienceDirect together with possibilities to better search for relevant publications and be better informed about new developments.

Despite these opportunities, human-only usage of ScienceDirect already has consequences involving possible misuse, such as handing out (copies of) publications to other parties or long-term caching digital copies of publications. By involving software agents, these consequences still exist, including over-use of ScienceDirect's website.

5 Conclusion

The use of software agents in real-world applications offers opportunities for business and scientific organizations, however, are often accompanied by legal and technical consequences. The case studied in this paper involves the contract between VUA and ScienceDirect: a simple and realistic case. A central question is the extent of the dif-

ference made by introducing software agents at VUA. Five scenarios are described in which humans, with or without the aid of software agents, access the digital library. In sum, software agents do not change the human uses of ScienceDirect, but they may access ScienceDirect more frequently and provide geographical independence to human users. Software agents do offer new opportunities to all parties involved, including not only ScienceDirect and VUA but also VUA employees, students and authors of publications.

Future research may on the one hand include VUA contracts with other digital libraries and on the other hand focus on other agent-based information retrieval applications, e.g. in courthouses or hospitals.

Acknowledgements

References

1. Luck, M., McBurney, P. and Preist, C., (2003) "Agent Technology: Enabling Next Generation Computing". AgentLink, ISBN 0854 327886, http://www.agentlink.org/roadmap/
2. McKnight, C., (1997), "Electronic Journals: What do Users think of them?" Proceedings of International Symposium on Research, Development and Practice in Digital Libraries 1997, ISDL '97, Sugimoto, S. (ed), University of Library and Information Science , Tsukuba, Ibaraki, Japan, 18th November 1997, pp 23-27.
http://www.dl.u lis.ac.jp/ISDL97/proceedings/mcknight.html
3. Bishop, A.P., (1998) "Logins and Bailouts: Measuring Access, Use, and Success in Digital Libraries". In *The Journal of Electronic Publishing*, 4(2). Available from http://www.press.umich.edu/jep/04-02/index.html.
4. Luther, J., (2000), "White Paper on Electronic Journal Usage Statistics", In: *The Journal of Electronic Publishing*, 2001, Volume 6, Issue 3, ISSN 1080-2711, http://www.press.umich.edu/jep/06-03/luther.html
5. Tenopir, C., Hitchcock, B., Pillow, A. (2003) "Use and Users of Electronic Library Resources: An Overview and Analysis of Recent Research Studies", Council on Library and Information Resources, 2003, http://www.clir.org/pubs/abstract/pub120abst.html
6. Brazier, F.M.T., Oskamp, A., Prins, J.E.J., Schellekens, M.H.M., Schreuders, E., Wijngaards, N.J.E., Apistola, M., Voulon, M.B. and Kubbe, O., "ALIAS: Analysing Legal Implications and Agent Information Systems", Technical Report no. IR-CS-004, Computer Science, Faculty of Sciences, Vrije Universiteit Amsterdam (2003), available from http://www.iids.org/.