Soil Science Web Resources: A Practical Guide to Search Procedures and Search Engines

M. A. Tseits and B. A. Devin

Faculty of Soil Science, Moscow State University, Leninskie gory, Moscow, 119992 Russia
Received December 5, 2002

Abstract—The Internet resources on soil science are examined. Information on soil and soil science available from different search engines is analyzed.

INTRODUCTION

The Internet long ago became a tool for providing various kinds of on-line information in all fields of human activity, including science. Electronic mail, Web pages, and videoconferences allow one to keep up to date on the latest developments in any science and significantly accelerate the exchange of opinions and sharing of knowledge among scientists. Computer and information technologies have arrived in Russia later compared to western countries; therefore, many people do not know how to find necessary information on the World Wide Web. The immensity of the Internet still complicates the situation. The aim of this paper is to systematize the possibilities of searching for information on the Internet and to analyze the information on soil science available from different Web resources.

POSSIBILITIES OF SEARCHING FOR INFORMATION ON THE WORLD WIDE WEB: GENERAL CONCEPTS

We first explain some of the terms most frequently used when searching via the Internet:

(1) The World Wide Web or Web (WWW) is a hypertext and hypermedia information system. It is an Internet service, along with telnet, news, e-mail, and ftp. The World Wide Web functions as an electronic publisher and a library.

(2) A Web site (site or host) is a server (computer) providing services for the delivery of documents and sending of electronic mail. The term "site" is frequently associated with the term "Web page" or "Web publication" (a sum of hypertexts united by common information content, design, and authorship).

(3) A URL (Universal Resource Locator) is the unique name of a document on the Internet. It has the following form: service://computer name/directory/subdirectory/@/filename (e.g., http://soil.msu.ru/~phys/news.htm).

(4) A Web browser (browser or Web client) is a program used to browse the world wide web, which runs on the client computer. The browser processes the hypertext documents received from the server and displays them [1,2]. The most popular browsers are Internet Explorer (Microsoft), Netscape Navigator (Netscape), and Opera.

SEARCH OPTIONS ON THE INTERNET

Different search mechanisms are used for the Internet and searchable directories, but the interfaces look identical for different systems. Any search system (searchable directory, search engine, or site search tool) presents a window for keywords and a set of tools for regulating the request parameters. Boxes for adjusting the request parameters are usually above the keyword window:

(1) "Search this site";

(2) "Matches on all words" (operator "AND");

(3) "Matches on any word" (operator "OR");

(4) "An exact phrase match";

(5) "Search Web sites"; and

(6) "Search categories" or "All" for engines with proper searchable directories.

In the section "Advanced search," the search can be restricted by the date, site (file) type, country or region, subject, the number of the results displayed, etc.

In searchable directories, the search program looks for the query terms in names, descriptions, and site keywords. The search mechanisms for the WWW are very diverse, but it is essential that the keyword search is usually carried out over the whole site. Therefore, it is not surprising when the response to a search for "dark gray forest soils" contains a page from a bucolic piece, because the latter includes a phrase like this: *"dark gray* heavens turned azure; peasants' faces radiated happiness; rich *soils* provided high crop yields; *forest* animals came to the very village fence." The response for "soil" includes, along with scientific sites, pages of land-management and landscape-design firms, Web

sites of popular musical groups whose names includes this word (e.g., "And The Soil Bleeds Black"), etc.

TAKING INTO ACCOUNT THE QUERY FORM (MORPHOLOGY)

Search options depend on the query form. The most evident (presently rare) is the formal search, i.e., the search for a strictly determined letter combination. For example, **1531,** 1800, and 1093 sites were found by the Lycos, Hot Boat, and AltaVista search engines, respectively, in response to a request for "soil micromorphology." However, one should not delude oneself: the searched words can be simply scattered over the text of the document found. In order to obtain the sought result, the query should be specified, e.g., by enclosing the target phrase in quotation marks.

SEARCH ON THE INTERNET

To facilitate orientation in the immense world of the Internet, there are a whole series of aides that facilitate the search. The most important searching tools are as follows:

(1) search engines,

(2) multiple search engines,

(3) searchable directories,

(4) search engines with proper searchable directories,

(5) libraries and library directories,

(6) link sections of large topical sites,

(7) bookmark collections on personal pages, and

(8) Web rings.

It is noteworthy that the most popular Web browsers (Internet Explorer, Netscape Navigator, Opera, and others) have built-in search options. Internet Explorer and Netscape Navigator address to their search resources (button "Search" or option "Search Internet"), and the Opera browser allows both the use of so-called "hot keys" (e.g., in order to find a world in the Google search system, the combination "g [target term]" can be typed in the command line) and special search tools in some search systems (Google, AltaVista, AllTheWeb, and others).

(1) Search engines include Web servers, which regularly read ("index") the contents of Web pages and place them in a general database. Search in this database is performed using keywords related to the topic of interest [1]. These are AllTheWeb (www.alltheweb.com), AltaVista (www.altavista.com), Microsoft Network (MSN) (www.msn.com), Google (www.google.com), and Hot Bot (www.hotbot.com), as well as Direct Hit, Excite, Fast Search, Go (Infoseek), Inktomi, iWon, Lycos, NBCi, Northern Light, Teoma, WiseNut, etc. One of the best known Russian search engines is Punto (www.punto.ru).

(2) **Multiple search engines** are search engines that distribute the query among the specified search engines rather than perform the search on the WWW. There are two types of multiple search engines: (i) online multisearch engines that are Web sites with attached databases of many search engines and directories (All4One (all4one.com), AskJeeves (www.askjeeves.com), Dogpile (www.dogpile.com), Findit! (www.iTools.com/find-it/find-it.html), Hotsheet (www.hotsheet.com), Ixquick (www.ixquick.com), Mamma (www.mamma.com), Metacrawler (www.metacrawler.com), QueryServer (www.queryserver.com)) and (ii) desktop multisearch engines that are software packages, which can be installed on every computer (BullsEye, Copernic, WebFerret, Express (Infoseek), EasySeeker, WebSeeker, X-Portal Findware, WebCompass, and Internet FastFind).

(3) **Searchable directories** are bases of Web pages divided according to subject headings. The directory database is usually filled by hand. Search is performed with keywords and category names, as well as in related sections. Searchable directories are Yahoo! (yahoo.com), Open Directory Project (dmoz.org) and its "clone" in the Google system (directory.google.com), MavicaNET (www.mavicanet.com), the Argus Clearinghouse (www.clearinghouse.net), Refer.Ru (refer.ru), and others.

At least several hundreds of thousands of searchable directories exist presently, and new directories appear constantly. Information on searchable directories is available on the Search Engine Colossus site (www.searchenginecolossus.com).

Most searchable directories are sets of references to popular topics: Auto and Moto, Sports, Acquaintances, MRZ, Shopping, Jobs, etc. Scientific resources are not numerous and can be found in sections like "Science and Education," where references to special problems of soil science and related disciplines can be found only by chance.

The following types of search systems (search engines and searchable directories) can be distinguished:

(a) worldwide systems, which refer to sites of different topics from different countries in different languages;

(b) national or regional systems, which survey the Web resources of a country or a region, e.g., Iceland on the Web (Iceland.vefur.is); Siftthru, the search engine of Africa (www.siftthru.com); and Sergiev Posad: Katalog resursov (sposad.al.ru); and

(c) thematic systems, e.g., sites of general scientific information, where every scientist can find useful data: SciNet (www.scinet.ee), Scirus (www.scirus.com), Russian scientific club (ruscience.newmail.ru), SciSeek (www.sciseek.com), BioChemLinks (biochemlinks.com), ScienceNet (www.sciencenet.org.ru), ChemIndustry (www.chemindustry.com), and Science-Internet-Russia (www.nir.ru).

(4) **Search engines with proper searchable directories** provide the possibility to search both in the base directories (option "Search in the directory") and in the whole Internet (option "Search in the Web): Yahoo! (www.yahoo.com), Lycos (www.lycos.com), Hot Bot (www.hotbot.com), MSN Search (www.msn.com), Excite (www.excite.com), LookSmart (www.looksmart.com), NBCi (www.nbci.com), Yandex (www.yandex.ru), Rambler (www.rambler.ru), etc.

(5) **Libraries and library directories.** Recently, large libraries have not only acquired virtual addresses, but they have also converted their catalogues into the virtual form. In the site of a library (e.g., www.rgb.ru for the Russian State Library or www.cnshb.ru for the Central Scientific Agricultural Library), the user can find the section "Electronic Catalogue" (or "Digital Catalogue"), where bibliographies and frequently requested texts of periodicals or monographs are presented. It is notable that, in other countries, electronic library services were developed long ago and not only in Web sites: information about a publication and its abstract are also available via telnet, e.g., in the National Agricultural Library of the US Department of Agriculture (www.nal.usda.gov). In Russia, electronic library services are still at an early stage of development. For example, the site of the Central Scientific Agricultural Library presents only abstracts and not full texts of scientific articles.

Purely electronic libraries should also be noted; these are large portals giving access to articles and monographs as simple text or pdf (portable document format) files. The latter are replicas of printed articles with figures and tables. The best known Russian electronic libraries are the Maksim Moshkow library (Iib.ru) and the project of e-library.ru supported by the Russian Foundation for Basic Research. A kind of electronic library is the AGRIS section of the FAO head site (www.fao.org/agris). AGRIS is the international information system for agricultural sciences and technology. It was created by the Food and Agriculture Organization of the United Nations (FAO) in 1974 as an agricultural database. To date, 240 national, international, and intergovernmental centers participate in this system. The AGRIS site includes literature databases, AGRIS/CARIS periodicals, and documents.

(6) **The links of large thematic sites** are convenient for users because they are divided according to subject headings, which facilitate the search for information. An example is provided by Links for Mineralogists (www.uni-wuerzburg.de/mineralogie/links.html).

(7) **Bookmark collections on personal pages.** These personal collections of Internet users on their sites initiated the development of searchable directories.

(8) **Web rings** are interlinked groups of Web sites devoted to common subjects.

Additional possibilities for data search are presented by discussion groups, chats, and forums (portals of interactive contacts with other Internet users), as well as news and news delivery services.

METHODS OF STORING REFERENCE COLLECTIONS

While searching the Internet, the need arises for storing references to the resources found. These references should be stored as links (Web directories), which open the page in question by clicking on the mouse button. This option exists in any browser (Zakladka (Bookmark) in the Russian versions of Opera and Netscape, Favorites in Internet Explorer and NetCaptor, and Izbrannoe in the Russian version of Internet Explorer). The links can be sorted into different folders at several levels. Several tens of links can be thus stored. Along with the low volume of links, an inconvenience is that they are available only from a specific computer. They should be available from different places (home and other computers, cities, and countries). A section "Links" can be created on a personal Web page. In this case, the links will be accessible from any computer connected to the Internet; moreover, this link collection becomes available for all Internet users. Another more progressive method is the storage of links in volunteer-edited link bases.[1] In this case, a user may be registered as a voluntary editor of a section; some sections may have several voluntary editors. The editor may rearrange the links, add new ones, etc. Another editor of the same section may also know some useful links, which can be added to the collection available to all the Internet users. The use of large directories is much higher than that of personal pages, and their network supply is better; therefore, directories have less access problems. They also have special forums for editors, where problems can be discussed. Among the volunteer-edited directories, Netscape's Open Directory Project (dmoz.org) and MavicaNET (www.mavicanet.com) are primarily recommended. The short falls of the former directory are the difficult registration for volunteer editors and the location of the servers beyond the limits of the Russian Federation, which can pose problem for shareware access (educational and other state institutions, e.g., the Faculty of Soil Science, Moscow State University) when foreign traffic is artificially limited by the provider.

RECOMMENDED RESOURCES

There are many guides for searching for subject information on the Internet. These guides, which are issued as books and brochures, become completely out of date in one or two years, because of the highly dynamic development of the World Wide Web. Permanently updated guides exposed on large analytical Web

---

[1] An URL (link) can be added to almost any directory of Internet resources, but il will be lodged automatically by a program or a staff directory editor who is not likely to be a specialist in any field of soil science.

sites are much more useful. Large directories of Internet resources have corresponding categories (directories) with collections of links to these resources. The section "How To Search" at MavicaNET can be cited as an example.

Large search systems include sections related to the potentialities and features of their search (Preferences, Help, About Us, Advertise with Us, etc.).

In conclusion of this paragraph, we cite some useful links: encyclopedia of search systems (www.search-engines.ru), what and how to search on the Russian Internet (www.comptek.ru/ya/altdictl.html), and the ABCs of Web searching for chemists (www.abc.chemistry.bsu.by)

REVIEW OF WEB RESOURCES ON SOIL SCIENCE

Language and Subjects of Web Resources on Soil and Soil Science

The number of Web resources devoted to soil and soil science is much lower than that of sites on mathematics, physics, and other natural sciences. This can be explained by the relatively small size of the community of soil scientists and by the insufficient knowledge of the wider Internet audience. English, the language of international scientific communication, is evidently the major language of the Web resources on soil. Most large Web pages created in nonanglophone countries have English versions. The Web resources in English are also characterized by the widest range of topics related to soil science. These are the sites of soil science societies, including the largest site of the Soil Science Society of America (www.soils.org), scientific and educational institutions (e.g., the site of the USDA Soil Salinity Laboratory, www.ussl.ars.usda.gov), databases on soil and soil cover (e.g., the Canadian Soil Information System (CanSIS), sis.agr.gc.ca/cansis), the International Soil Reference and Information Center in the Netherlands (ISRIS, www.isric.nl), on-line courses and educational materials (e.g., those available on the site of the Department of Soil Physics, New Mexico State University (soil-physics.nmsu.edu)), the on-line soil loss calculator using the USLE equation, etc. Most of these sites present links useful for soil scientists, lists of lines of inquiry, and publication texts. The CanSIS information system also presents information on the classification of Canadian soils and their properties, soil maps, etc. The ISRIC Web site is of more general character. It presents not only databases on soil properties, but also supporting materials for students, a virtual soil museum, soil maps, a keyword list (a terminological glossary like that on the SSSA site as a text and an MS Word document), and the World Reference Base for Soil Resources. In addition, the ISRIC site presents journal articles, news of soil science, bibliographies, and soil links. Soil and soil science are discussed in some groups and forum, including the World of Soil Forum (server3.ezboard.com/ffredmoorforum-stheworldofsoilforum).

Resources in other European languages (German, French, Spanish, and others) are less numerous. The sites in these languages are mainly presented by laboratories, departments, and task groups of large universities. The Spanish resource "Sistema Español de Información de suelos" (leu.irnase.csic.es/mimam/seisnet.htm), which presents soil maps of Spain and attendant data on the topography and geology of the country, is an analogue of the Canadian CanSIS.

Russian-language sites on soil science should also be mentioned. As was noted, information technologies have arrived to Russia relatively late, and the development of scientific Web resources (including those devoted to soil science) is still at the early stage of development. Nonetheless, some unique sites have been created. The site of the Faculty of Soil Science, Moscow State University (soil.msu.ru), is the largest of them. This resource tells about the scientific and student life of the faculty; it posts articles published in *the Moscow University Soil Bulletin* and abstracts of the Lomonosov Conference (Section of Soil Science). Informative sites are also presented by the Institute of Soil Science, Moscow State University, Russian Academy of Sciences (soilinst.msu.ru); the Department of Soil Science and Soil Ecology, St. Petersburg State University (www.bio.pu.ru/soil); and the Dokuchaev Central Soil Museum (www.soilmuzeum.by.ru). The latter is the single developed site of a soil museum, which presents not only its exposition, but also materials on the history of soil science in Russia and its current state and development. The Dokuchaev Soil Science Institute, Russian Academy of Agricultural Sciences, also has its own site (agro.geonet.ru), on which a set of electronic maps of the soil cover of Russia is presented, along with general information about the institute, its divisions, and staff members. The first electronic edition of *Interactive Soil Science Reports* appeared recently (soilinst.msu.ru/publ). Unfortunately, the Dokuchaev Society of Soil Science and the journal *Eurasian Soil Science (Pochvovedenie)* have no Web resources of their own: only short pages with little content about them can be found on the Internet.

Along with specialized resources, sites of more general character covering agricultural and natural sciences on the whole can be useful for soil scientists. They include the FAO site (www.fao.org), which publishes documents, statistical data, and multimedia resources on global agriculture and its branches. Along with the English version, this resource has versions in Arabic, Chinese, French, and Spanish; thus, its information is available to most people in the world. The sites of the United Nations Environment Program (www.unep.org) and its agency in the Norwegian city of Arendal (GRID-Arendal, www.grida.no), which publish state reports, maps, GIS data and other materi-

als on the state of and changes in the environment, are also important for soil scientists.

The site of the International Institute for Applied Systems Analysis (IIASA) (www.iiasa.ac.at) is also a meeting place for environmentalists. The institute is located in Austria and performs interdisciplinary studies related to the anthropogenic impact on the environment. Journal papers, reports, software products (databases, GIS, etc.) related to silviculture, land tenure, and environmental contamination are available at the institute's site.

TOPICS OF SOIL SCIENCE SITES IN INTERNET DIRECTORIES

In this paragraph, we dwell on the soil science topics presented in searchable directories. Note that sections on soil science are not present in all directories: soil sites are usually lodged in sections like "Earth Sciences," "Agriculture," "Agronomy," and "Biology." We shall consider three directories having sections entitled "Soil Science" (or "Soils"): ODP, Yahoo!, and Mavica-NET. In the former two directories, material is presented in categories related to the urgent problems of soil science as a whole (salinization and erosion of soil, primarily wind erosion) rather than in categories related to the subsections of soil science. In addition, the categories devoted to soil and soil science are related only to agricultural sciences, which does not show the role of soil science, which is closely related to other naturalistic sciences (biology, geology, geography, and others). This problem is partly solved in MavicaNET: the section "Soil Science" is related to the categories "Agroindustrial Sciences," "Earth Sciences," and "Biology of Ecosystems." The categories in this directory are divided by fields (the genesis and classification of soils, biology of soils, chemistry of soils, etc.) and specific problems of soil degradation (salinization, erosion, and contamination); resources on soil scientists and materials on soil surveys in different countries are presented; subsections on soil taxonomy are developed in more detail. Edaphology is separated as a science studying the effect of soil and other environments on the growth and development of plants. The difference between Russian and foreign soil scientists in the interpretation of this term is evident in this case: in Russia, edaphology is an equivalent of pedology and, hence, of soil science; abroad, the concept of edaphology has acquired an applied sense close to agrophysics and agrochemistry.

INTERNET JOURNALS ON SOIL SCIENCE

Presently, large scientific publishers—Wiley and Sons (www.wiley.com), Elsevier (www.elsevier.nl),

and others—have begun to publish the texts of articles and monographs on their sites and offer electronic versions of their publications. The possibility has appeared not only to read a paper of interest on the computer screen but also to download the file with its text to any medium. This is partly explained by the high subscription prices and the long duration of delivery by mail. For scientific journals, the choice is to make an electronic version or to disappear [3]. Purely electronic publications also have appeared: the above mentioned *Interactive Soil Science Reports* and *Sciences of Soils* (hintze-online.com/sos/). In the opinion of Hartemink [3], soil science journals are still in the intermediate state between printed and electronic documents. He believes that the electronization of journals will improve the availability of soil information and reduce its cost.

At the same time, publishers demonstrate a purely commercial approach to the publication of journals on the Internet. The reading of a paper or its recording on a medium is paid for; only access to the paper abstract is free of charge.

CONCLUSIONS

In conclusion, we would like to consider the problems that arise during the acquisition of information via the Internet. In many cases, the absence of information on the World Wide Web is explained by the unwillingness of authors (especially of Russian authors) to expose their results in open sources before their appearance in press. In our opinion, this is related to the imperfection of the Russian copyright legislation: there is still a fear that the results of studies, original projects, procedures, etc., might be used without the author's permission. However, the results can go out of date by the time a scientist reserves the right to a publication. This problem is less acute for foreign scientists: there are special tools (access for authorized users only, program codes protecting texts from being copied from the site, etc.) that prevent the use of published works in selfish ways. We hope that these problems will be solved in the near future and that soil scientists will be able to exchange information and to publish their results using the latest advances in computer and information technologies.

REFERENCES

1. V. V. Berkgaut and I. S. Chardin, *Internet: First Steps* (Moscow, 1999) [in Russian].
2. E. V. Yakushev, *Potentials of Telecommunication in Russia* (FADR, Moscow, 1995) [in Russian].
3. A. E. Hartemink, "Publish or Perish: 4. Electronic Publishing," Bull. Int. Union Soil. ScL, No. 98, 56-64 (2002).