information services and technology

IS&T

News about Information Services and technology throughout MIT

Volume 20 • Number 2 November / December 2004

Inside! Schedules of Hands-on Training Courses and Free Computer Events

A Decade of W3C Transforms the World's Webbing

• Lee Ridgway

We tend to think of the Web as being everywhere, and accessible from just about any device that can connect to the Internet. What we may not think about is how this is possible. Thanks are due to the World Wide Web Consortium, or W3C, for its creation of standards and technologies that ensure the Web's interoperability.

This year marks the 10th anniversary of the W3C and the establishment of its headquarters here at MIT. A symposium on December 1 remembered the W3C's origins, and considered the future of the Web and W3C's role in it. Associated with this celebratory event, Massachusetts Governor Mitt Romney declared December 2004 "World Wide Web Consortium Month."

This year also marks the 10th anniversary of MIT's official home page, http://web.mit.edu. So a look back and forward seems appropriate, and who better to go to than the Web's founder, Tim Berners-Lee, whom *is&t* interviewed by e-mail.

At the Start

The Web's beginning is pegged at March 1989, when Berners-Lee presented a proposal for a global hypertext system to his employers at CERN, the European laboratory for particle physics research. Berners-Lee recounts that in 1990, he wrote the software for the server, an editor, and a browser, all christened "WorldWideWeb." He also set up the first web server, on his NeXT machine, to serve the CERN community.

The next step – expanding the Web beyond CERN – came in August 1991 when, for a fee, CERN made the program files available via file transfer. The Web remained mainly within a small realm of research labs, universities, and techies, with about 50 known servers operating by January 1993. That year seemed to be significant to the Web's growth, with three developments standing out:

- CERN declared that the Web technology would be free to anyone.
- The Mosaic browser was released, for free, for all common platforms: Unix, Windows, and Macintosh.
- Articles on the Web appeared in *The New York Times*, *The Guardian*, and *The Economist*.

By the way, in 1993, the first web site at MIT made its debut, created by the Student Information Processing Board (SIPB). It's URL was www.mit.edu, which is why MIT's official site is web.mit.edu (that's another story).

The W3C Takes Shape

Planning for the W3C began in 1994, with the late Michael Dertouzos of the Lab for Computer Science playing a key role in bringing the Consortium and Berners-Lee to MIT. The first W3C meeting was held in December of that year.

continued on page 2 ▶

A Decade of W3C

continued from page 1

When asked about challenges to the W3C in its first couple of years, Berners-Lee relates that HyperText Markup Language (HTML), the coding behind web pages, "was in dire danger of fragmenting." Browser developers were adding functionality in an effort to look the best, but doing so with incompatible extensions to HTML. The W3C was left playing catch-up in this area, a situation that could have been lessened if the Consortium had started a year earlier. In other areas, though, such as style sheets and graphics, the W3C led the way with sample code and demonstrations to show what was possible.

Although less of a problem in recent years, some web pages only work properly with certain browsers. This seems to be the case with some financial services, retail purchasing, and graphics-intensive sites. Berners-Lee believes these sites need to be redesigned quickly, due to the growing diversity of browsers and devices on which web pages can be displayed.

The W3C is considering a Mobile Web initiative, which would develop standards for coding content so that it works equally well on mobile devices and conventional computers. These standards would champion the concept of separating content from the form in which it is presented.

The Semantic Web

Berners-Lee himself is heavily involved in the Semantic Web. He sees it as an extension of the current Web, "in which information, or data, is given well-defined meaning, better enabling computers and people to work in cooperation." In short, it is about data, not documents.

Vast amounts of data are in XML or relational databases, application data files in proprietary formats, and spreadsheets. Some of these data do generate documents on human-centric web sites. Even so, the data can't be treated as such: it can't be queried, joined, or combined into new relationships across the Web.

The Semantic Web is about getting data exposed to the Web in a standard format, Resource Description Framework (RDF). RDF is a set of rules for building a description of the real-world things the data is about, then describing to the computer exactly how the data are related to those things. These descriptions use the Web Ontology Language (OWL). OWL is a set of subject or domain-specific vocabularies that represent areas of knowledge (such as medicine, music, real estate). These descriptions make it possible for a computer to connect to data about the same thing, even if the data is scattered on different systems around the world.

RDF and OWL, the foundation standards of the Semantic Web, were developed by the W3C and are now well established, leading to new directions of development. To follow these developments, start at the W3C's Semantic Web site at

`http://www.w3.org/2001/sw/`

Related Sites

World Wide Web Consortium
`http://www.w3.org/`

A Little History of the World Wide Web
`http://www.w3.org/History.html`

Tim Berners-Lee's original proposal to CERN
`http://www.w3.org/History/1989/proposal.html`

SemWebCentral
`http://www.semwebcentral.org/`

SchemaWeb – RDF Schemas Directory
`http://www.schemaweb.info/`

IS&T Doubles MIT's External Network Connections

Information Services and Technology (IS&T) recently reconfigured its external Internet connections, lowering costs and effectively doubling MIT's commodity Internet bandwidth. Three communications vendors – Cogent, Level 3, and Sprint – now provide MIT's external connectivity, so if one link goes down, the Institute will still have Internet access through the other two vendors. The increased bandwidth includes new one-gigabit Ethernet connections to Level 3 and Sprint.

Fiber Ring

These new connections are routed via the Boston Area Metro Fiber Ring, which MIT, Harvard, Boston University, and Northeastern University recently acquired. This ring goes down Vassar Street, then to various "carrier hotels" – places where many carriers

have facilities – including the Level 3 point of presence (POP) at 300 Bent Street in Cambridge. This is where MITnet connects to Cogent, Level 3, and Sprint.

This ring provides MIT with "dark fiber" to selected locations in the Boston area at very little cost. "Dark fiber" refers to the whole network fiber: MIT literally controls the ends of the glass fiber and can choose to "multiplex" it into many individual circuits.

Continuing Connections

In addition, IS&T continues to maintain its existing network connections to Comcast, Northern CrossRoads (Boston area Gigpop), Abilene (Internet2), and the Energy Sciences Network (ESnet).

IS&T Information Services & Technology

Managing Editor
Robyn Fizz

Writer/Editor
Lee Ridgway

is&t is published six times a year. MIT faculty and staff receive copies through campus mail; *is&t* is also available in lobbies around campus. Individuals at MIT may subscribe by contacting the managing editor.

Send comments or subscription requests to:
MIT Room N42-290B, 77 Massachusetts Avenue, Cambridge, MA 02139-4307
Phone: (617) 253-0540
Electronic mail: <fizz@mit.edu>

is&t is published online at
`http://web.mit.edu/ist/istnews/`

The IS&T web site also offers frequent news updates on its home page or at
`http://web.mit.edu/ist/news/`

Back Up Your Data with Tivoli Storage Manager (TSM) 5.2.3

• Dave Kalenderian, Patrick Whitney, and Esther Yanow

TSM is MIT's enterprise data backup and restoral software. Backing up the contents of your computer to a secure network server lets you restore data in case of error or computer failure. For more information on IS&T's fee-based backup service, see

http://web.mit.edu/ist/topics/backup/

IS&T recently released TSM 5.2.3 for Windows and Macintosh. TSM 5.2.3 for Linux is also available and supported. This latest release of TSM includes minor bug fixes and additional features. Notable features of the 5.2.3 client on the Macintosh are scheduled backups and encryption.

Scheduled Backups

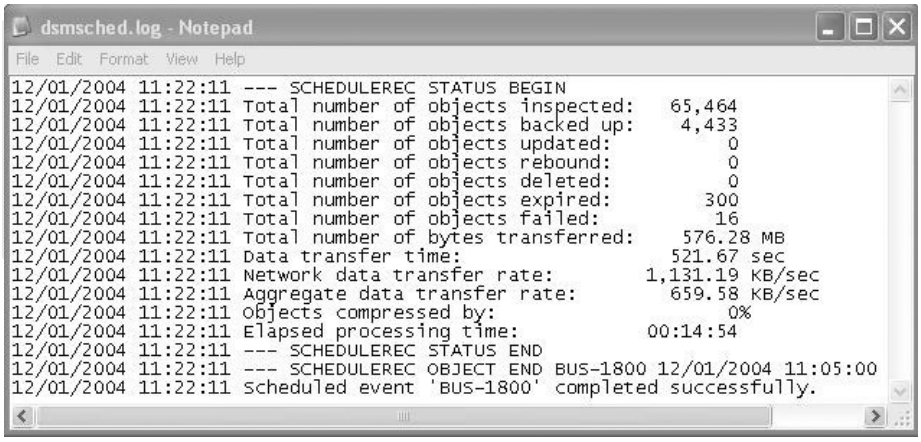
Before this release, Macintosh users had to be logged in to run a scheduled backup. With version 5.2.3, the scheduler can be installed and the scheduled backup will run regardless of whether the user is logged in.

Note: Users on all platforms who opt for scheduled backups should check the Schedule Log frequently to verify that scheduled backups are successful (see screenshot above).

Encryption

TSM 5.2.3 for all platforms now supports encryption of backup data. With encryption, your data is secure while being transmitted over the network. Often encryption is not necessary, but in certain circumstances it is required – for example, when you are legally bound to protect data.

While the option to encrypt data addresses a need, care must be taken when using it. If you decide to turn on encryption, you will be prompted to create a unique key. Without this key, you won't be able to restore your data. It is very important that you keep a copy of this key someplace other than the computer that is being backed up. If you lose this key, the TSM Service Team will not be able to help you. To save the key, copy it to removable media, such as a CD-R, or to another computer. To be prepared in the event of a disaster, you should keep a copy of the key in a secure offsite location.



You can check the status of your last backup by finding SCHEDULEREC STATUS BEGIN near the bottom of your schedule log.

For more information about encryption, go to

http://itinfo.mit.edu/article?id=7444

If you still have questions, send e-mail to <tsm-systems@mit.edu>.

You can download TSM from the MIT IS&T Software Site at

http://web.mit.edu/software/

Note that you need a current personal certificate to download this software.

Backup Reports

The TSM Service Team offers, at no additional charge, backup reports for departments and organizations with five or more TSM accounts. These reports include a list of TSM accounts and the date of the last backup. The purpose of these reports is to facilitate the management of backups, helping a local administrator to see whether or not the department computers are being backed up.

Reports are sent twice a month. Administrators interested in this service should send e-mail to <tsm-systems@mit.edu> and include a list of node names to be monitored, along with the e-mail address(es) the report(s) should be sent to.

Information and Support

For instructions about obtaining, installing, and using TSM, see

http://itinfo.mit.edu/product?name=tsm/

If you have questions about the software or service, contact the Computing Help Desk at <computing-help@mit.edu> or 253-1101.

In addition, the TSM Service Team will offer a free IAP session, "TSM: Why Back Up Anyway?" on January 25 from 2 to 3pm in the N42 Demo Center.

Where Does My Backup Go?

You may be wondering what happens to your data when you perform a backup of your computer using TSM (or when a scheduled backup occurs). The answer is fairly straightforward: The data is sent, via MITnet, to one of four dedicated backup servers in IS&T's secured Data Center in W91. Initially stored in a compressed format on the backup server's RAID (Redundant Array of Inexpensive Disks) protected disk, the data is eventually migrated to highly reliable, robust tape cartridges stored in one of three automated tape libraries. These tape libraries are a great addition to the TSM service, allowing restores to happen without human intervention in the Data Center.

The data remains online and available for restorals 24 hours a day, 7 days a week, 365 days a year. For details about TSM policies regarding data retention, see http://itinfo.mit.edu/article?id=6912.

For more information about the automated tape libraries, see http://www-1.ibm.com/servers/storage/tape/3494/ and http://www.storagetek.com/products/product_page30.html.

Update on Web Browser Support at MIT

• Lee Ridgway

Increasingly, MIT relies on secure web sites to deliver critical information, services, and applications to members of the community. Access to these secure sites, such as Employee Self-Service or WebSIS, requires authentication of the user and an encrypted connection. This secure access is controlled through x.509 digital certificates, otherwise known here as the MIT CA and your MIT personal certificate.

For many years, MIT's implementation of the x.509 certificates required the Netscape browser on Macintosh and Windows platforms, or Mozilla on Athena. Over the past couple of years, the situation has changed, so that other browsers, including those native to an operating system, are now compatible with MIT's x.509 certificates.

Read on for a summary of current supported browsers at MIT. These browsers come with their respective operating system, and are not provided through IS&T's software download site. For more information about a browser or to download later versions, go to the vendor's web site.

Athena

Mozilla 1.7.x is the current supported version available in the Athena environment on Linux and Solaris. Major updates to the default Mozilla version are made when a new release of the Athena system is installed, usually during the summer. Minor updates (e.g., to fix security holes) are installed, as needed, in Athena patch releases.

`http://www.mozilla.org/products/mozilla1.x/`

For a summary of the differences between Mozilla on Athena and Mozilla in Red Hat Enterprise Linux, see the IS&T stock answer at

`http://itinfo.mit.edu/answer?id=7166`

Linux

Mozilla 1.4.x is the supported version for stock (non-Athena) Linux. It is fully integrated into Red Hat Enterprise 3 for Linux. Mozilla 1.4.x is installed with the operating system.

`http://www.mozilla.org/products/mozilla1.x/`

Macintosh

On Mac OS X 10.3.4 or later, the recommended browser is Safari 1.2.2 or later. Safari 1.2.2 supports x.509 certificates, and IS&T anticipates that new releases of Safari will continue to operate within MIT's web and certificate environment.

`http://www.apple.com/safari/`

Versions of Safari prior to 1.2.2 do not support x.509 certificates. If you are running a version of Mac OS X previous to 10.3.4, IS&T recommends that you upgrade. The Mac OS X 10.3.x upgrades include the later versions of Safari.

Windows

Internet Explorer (IE) 6.0 with SP2 is the supported browser on Windows. It is important that the IE browser version be current with the operating system, and that all browser patches be up to date. IS&T anticipates that new releases of IE will continue to operate within MIT's web and certificate environment.

`http://www.microsoft.com/windows/ie/default.mspx`

Netscape

MIT support for Netscape will be limited after June 30, 2005. The Mozilla organization, which produced Netscape, stopped work on this browser in 2003. If you are still using Netscape on Windows or Macintosh to access secure web sites at MIT, IS&T urges you to transfer to the IS&T supported operating system and browser appropriate for your computer platform.

Firefox

Mozilla recently released Firefox 1.0. Although IS&T has not begun a release effort for Firefox, and it is not supported by IS&T, Firefox is recommended as an alternative browser for users who may still be running Mac OS X 10.2.8 or earlier. Mac OS X 10.2 users, however, are strongly encouraged to upgrade to Mac OS X 10.3.x.

`http://www.mozilla.org/products/firefox/`

Support

If you need help installing or using any of these browsers, contact the Computing Help Desk at <computing-help@mit.edu> or 253-1101.

This column presents announcements about IS&T-supported software. For more information about recent releases, see `http://web.mit.edu/swrt/`

Microsoft Campus Agreement Benefits MIT Windows Users

Information Services and Technology (IS&T) has negotiated and funded a Microsoft Campus Agreement for Windows operating systems and Client Access Licenses (CALs). This agreement grants MIT faculty, staff, and undergraduate students the right to use Windows XP Professional and subsequent operating system releases from Microsoft.

To learn more about or to participate in this program, contact the Software Release Team at <swrt@mit.edu>. Also visit the Microsoft Campus Agreement at MIT page at

`http://web.mit.edu/ist/products/vsls/mitwin/msca.html`

for the latest information regarding these agreements.

On a related note, IS&T has extended support for Windows 2000 through March 31, 2005. This extension parallels that of Microsoft, and was made in recognition of the planning it takes to upgrade an operating system across a department or lab. (Originally, IS&T had announced that support for Windows 2000 would be retired on December 31, 2004.) Note that IS&T will not be testing new software releases on Windows 2000 after January 1, 2005.

Less Support for Selected Software

To maintain high-quality support for key applications and improve responsiveness to new vendor releases, IS&T periodically downgrades its support for older versions of software. Starting December 31, IS&T will provide only limited support for Eudora 5.1 (Windows), Eudora 5.2.3 (Macintosh), FileMaker Pro 5.5, and Office 2001. On June 30, 2005, IS&T will also reduce its support for Netscape 7.02, Office XP (Windows), and Office X (Macintosh).

IS&T will not test these products in association with future releases, and won't produce new documentation or training for retired software. To learn more about these decisions, see

`http://web.mit.edu/swrt/announcements/de-support-fy05.html`

Security Principles: Identity, Authentication, Authorization

• Christopher Logan

Most of us use some form of authentication daily – for example, when we get money from an ATM machine, log into e-mail, or access the gym with a photo ID. In a world full of passwords and "prox" cards, it's useful to get a handle on the key principles on which security is based. Identity, authentication, and authorization are closely linked.

Identity distinguishes who someone is or what something is. Identity can refer to a person, program, computer, or data. Identification is the process of establishing who someone or something claims to be.

Authentication is the process of confirming a claimed identity. For example, motorists identify themselves to police by presenting a driver's license. Police compare the photo and description on the driver's license with the motorist to authenticate identity. All forms of authentication are based on something you know, something you have, or something you are.

- *Something you know* is some form of information that you can recognize and keep to yourself. This could be a personal identification number (PIN) for your bank account or a password. Within the information technology realm, a password is the most common form of authentication.
- *Something you have* is a physical item you possess. This could be an item like a photo ID or a security token. A security token is a small hardware device such as a proximity or magnetic strip card you carry to authorize access to a service or building.
- *Something you are* is a human characteristic considered to be unique, like fingerprints, voice tones, and retinal patterns. These are also referred to as biometrics.

Once identity has been confirmed, authorization may come into play. *Authorization* is the act of granting permission for someone or something to conduct an act. Even when identity and authentication have indicated who someone is, authorization may be needed to establish what he or she is allowed to do.

By Way of Example: A Secure Entry

The best way to understand how these security principles interact is by example. Let's say you work in a secure building. At the door, you present an ID card – a form of identity – to the security guard. The guard looks at you and compares your face to the photo on the card. In doing this, the guard is authenticating you. Next, the guard checks your ID number and name against a database, and finds out that you are allowed inside the building. This look-up is a kind of authorization. You are now allowed to go inside.

It's useful to be aware of these concepts because information security plays a key role in our daily lives. Understanding these terms can help you better understand requests that come your way – online or in person. It also serves to underscore the importance of keeping your identity secure.

This IAP, Get up to Speed on Information Technology

• Jeff Pankin

Once again, IS&T is offering a flurry of IAP events. The sessions described here are just a sampling. For a complete listing, see

`http://student.mit.edu/iap/nsis.html`

Home Networking

Macintosh: Jan 18, noon–1:30pm, N42 Demo Center

Windows: Jan 24, noon–1:30pm, N42 Demo Center

If you have a high-speed (broadband) connection at home, you may want to connect one or more computers to it to access the Internet and MITnet. These sessions will cover required equipment, such as routers and wireless access points, and how to configure your computer so that you can use Kerberized applications such as Eudora and Fetch from home. You'll also learn how to make your wireless networks secure.

Tools for a Less Stressful Computing Experience

Jan 27, 11:30am–2pm, 7-143

The ATIC Lab will showcase alternative keyboards, pointing devices, break software, and mouse-clicking software. The focus is for visitors to try out an extensive inventory of keyboards and pointing devices that may help prevent repetitive strain injuries (RSI) at the computer. Before and after IAP, individual keyboards and mice are available for a three-day loan period to anyone with a valid MIT ID.

Best Practices in Web Publishing

Jan 27, noon–1pm, N42 Demo Center

Learn about the basics of MIT's web publishing environment. Topics will include file naming conventions, folder structures, image formats, file transfer, using Dreamweaver MX 2004, and more. This session is recommended for those who are interested in learning the most efficient and painless ways to keep web site files up to date.

Overview of Educational Technology at MIT

Jan 27, 1–2:30pm, 3-133

Staff from Academic Computing will discuss and demonstrate a variety of educational technologies in use by MIT faculty. These include web-based simulations, math software, instructional video, geographic information systems, web-accessible databases, and electronic textbooks. Staff will be available to consult with instructors on existing or potential projects.

Protect Your Computer from Spyware

Jan 28, 11am–noon, NE49-3100

Spyware is software that surreptitiously gathers information and transmits it to interested parties. Learn about different types of spyware and how it gets installed on your computer. Understand the personal privacy, security, and system issues caused by spyware. The session will also cover strategies for removing spyware and tips for prevention.

This column presents tips about computing. If you have a question you would like to see answered here, send it via e-mail to <techtips@mit.edu>.

For more information technology Q&As, check the IS&T Stock Answers database at

`http://itinfo.mit.edu/answer/`

Backing Up When You Travel

If you are away from MIT and can connect to the Internet, you can back up files on your computer using TSM, MIT's enterprise backup system (see the Software Spotlight on page 3 for more information). The speed of the backup will depend on your Internet connection service and will probably be slower than the service on campus.

However, there may be times when TSM is not an option. For example, you may be going to a remote location and plan to use your laptop while you're away. You know that it's smart to back up your data, but also realize that you're not going to have an Internet connection and so won't be able to access TSM. Not to worry: even without an Internet connection, there are other options for backing up your data.

Backup Options

If you have a CD/DVD rewritable drive or Zip drive, you can use it to do a quick backup of your files. If you don't have either of these drives, consider using a USB flash drive/memory stick type of device. These devices are very easy to use. Just plug one into a USB port and it appears as a new drive. In addition, these devices are portable, durable, offer very good performance, and require no separate power source. Some devices include encryption software that can protect your data from being easily viewed in the event you lose the USB device.

Considerations

If you opt to use some form of USB backup device, here are some things to check out or keep in mind:

- Make sure that your laptop has an available USB port. Some flash drives require a USB 2.0 port, but many will work with the slower USB 1.1 ports (*see Note below*).
- If you decide to use the encryption software included with the USB device, be sure to store the encryption key in a safe place.
- Not all flash drives are created equal (*see Note below*).
- Flash drives are *not* a replacement for backing up to a more reliable and robust enterprise backup solution, such as TSM. They are useful as a temporary solution when you are not connected to the Internet.

You can find these devices at your local retailer or at the GovConnection web site at

`http://web.mit.edu/ecat/govconnection/`

Note: For a more in-depth review of some popular USB flash drive devices, including USB specifications, common features, performance benchmarks, and a comparison matrix, go to

`http://arstechnica.com/reviews/hardware/flash.ars`

BACKUP OPTION	COSTS	CAPACITY	PROS	CONS
USB flash drive – Memory stick – Pen drive – Jump drive – Key drive – Thumb drive	$25–300	128MB up to 4GB	Inexpensive, small, easy to connect. Some vendors include encryption software.	Small, easy to misplace or drop
CD/DVD drive	$50–100	650MB	Easy to use	Bulky
Zip drive	$75–300	750MB	Easy to use	Bulky
Diskette drive	Under $10	1.44MB	Easy to use	Small capacity, becoming obsolete

IS&T does not specifically endorse any of the alternative backup storage devices described in this article.

PI/Space Registration and SARA Reporting Move to SAP

• Lisa Lucenti

MIT is committed to achieving and maintaining compliance with federal, state, and local environment, health and safety (EHS) laws and practices. To reach this goal, MIT is in the process of integrating EHS data into SAP.

The first phase of this project – moving Principal Investigator (PI)/Space Registration and SARA (Superfund Authorization and Reauthorization Act) Reporting to the SAPweb site – was implemented in November. These transactions are available via the EHS site at

`http://web.mit.edu/ehs-ms/`

or directly through

`http://web.mit.edu/sapweb/`

PI/Space Registration

PI/Space Registration focuses on the Principal Investigator and the space for which he or she is responsible. PI/Space Registration groups rooms into roomsets assigned to a particular PI or supervisor and then identifies hazards, safety equipment, and room roles – such as emergency contacts, EHS representatives, and SARA reporters. Based on authorizations, the PI/Space Registration transactions let appropriate personnel view and update room and roomset information.

SARA Reporting

SARA reporting is a requirement of the Environmental Protection Agency (EPA). Previously a paper-based process, SARA reporting in SAPweb includes a custom worksheet and online submission transaction. During the annual SARA submission period, listings of SARA substances can be copied from the previous year, updated, and sent to MIT's Data Warehouse. EHS then sends a compiled report to the EPA.

Data Warehouse

The Data Warehouse provides a custom dashboard for PI/Space Registration and SARA Reporting. Over a dozen possible reports provide handy data analysis for EHS lead contacts and coordinators.

Questions?

If you have questions about using the EHS tab in SAPweb, send e-mail to <environment@mit.edu> or call 252-3477.

Better Web Searches: IS&T Brings Google to MIT

• Suzana Lisanti

Information Services and Technology (IS&T) has launched a phased roll-out of Google at MIT that will seamlessly replace the Google search on the MIT home page, run by Google, with a licensed version of Google, run by IS&T. By running its own installation of Google, MIT can increase the frequency of indexing of the Institute's web sites and better control its completeness. In addition, search results can be returned in a look and feel that is customizable by each department's webmaster.

IS&T will continue to support the Inktomi search engine until departments have migrated their web site search forms from Inktomi to Google.

Why Google?

Participants in an IS&T Discovery project did "masked" tests of Google and Inktomi. The testers performed various searches and ranked their perception of results. Google results were preferred twice as often as those of the Inktomi search engine. Since MIT has over one million web pages on more than 1000 servers, it made sense to license the search engine that could provide the best results.

The Google search appliance also provides

- Control over the schedule of Google's indexing crawler: the on-site search appliance crawls MIT web sites twice a week, while the commercial Google service only updates its university index about once a month
- Searching of subcollections
- Advanced features, including the ability to search for PDF, .doc and .ppt files and display them as HTML
- Through authentication, the ability to crawl and index restricted content (coming this spring)

This implementation of Google does not index images. However, you can search for MIT images using Google's Advanced Image Search at

`http://www.google.com/advanced_image_search`

To limit your search to MIT, type "mit.edu" in the Domain field.

How to Use Google at MIT

You can search MIT web sites via the MIT Home Page or by using the Advanced Search page at

`http://web.mit.edu/search.html`

For tips on effective searches, see

`http://www.google.com/help/basics.html`

For developer documentation (certificates required), go to

`https://mit.edu/ist/google/internal-ref/`

By mid-January, IS&T will post examples of Google search forms, as well as instructions or tips on

- Adding a Google search form to your site
- Customizing the look and feel of search results
- Maximizing the search engine ranking for your site
- Keeping web pages out of search engine indexes

Getting Pages into the Index

Google's indexing crawler starts at the top of the MIT web site and follows links to find all indexable pages in the MIT web environment. All you need to do to get your web pages into the MIT/Google index is

- Post the pages in a public web space
- Ensure that your pages can be reached by clicking links from one of the top-level pages in MIT's web environment

There's no need to submit pages to the index: the Google crawler will pick up changed, new, and removed pages automatically during its crawls.

Note: Currently MIT does not index dynamic pages with "?" in their URLs, except by special request. These pages are usually generated by databases with very large page counts, which would use up MIT's document limit under the license with Google.

Improving Your Site's Visibility

In most cases, if you've taken the time to create web pages, you want search engines to find them. Google has over 100 criteria for assigning a numerical "relevancy" rank to pages, but they don't publish these criteria. Some are well-known, such as putting keywords in your title, description, and keywords meta tags. For in-depth advice on improving your web site's visibility, see "Search Engine Submission Tips" at

`http://searchenginewatch.com/webmasters/`

and "Using XHTML/CSS for an Effective Search Engine Optimization Campaign" at

`http://www.alistapart.com/articles/seo/`

Keeping Pages Out of the Index

If you don't want a page to be indexed, insert this meta tag within the page's <head> tag:

```
<head>
<meta name="robots" content="noindex, nofollow">
</head>
```

This code will prevent crawlers from indexing the page and from following any links from the page. If the page has already been indexed, it will be removed from the index the next time Google crawls the page.

If you have pages that need to be removed from the Google index immediately, send mail to <google@mit.edu>.

Support

To learn more about Google at MIT, attend the Web Publishers User Group meeting on Thursday, January 27 at noon in the N42 Demo Center. The presenters will discuss improving search results, optimizing page visibility, and creating a search form for your site.

If you have comments or questions about Google at MIT, send them to <google@mit.edu>.

MIT Search Engine Consultation

The IS&T Web Communications Services (WCS) Team can help your department, lab, center, or program use MIT's search engine more effectively within your web site, and optimize your pages for more accurate search engine indexing. WCS web consultants can

- Add a custom Google search form to your site
- Customize the MIT Google search results in the look and feel of your site
- Audit your site and provide a report detailing how to optimize the pages for better search indexing
- Implement optimization recommendations discovered in a site audit

These are fee-based services. To find out more, call WCS at 253-3500 or send mail to <web-consult@mit.edu>.

Getting Help

If you don't know where to get help for your computer, network, or telephone problems, dial one of the help lines listed to the right.

If you prefer to use e-mail, you can send your questions to the corresponding e-mail addresses on the far right. (When logged into Athena, you can also use the `olc` command to send questions to Athena's online consultants.)

You can also submit a question online via Casetracker at

`http://casetracker.mit.edu/`

For help with...	Dial...	Or send a message to...
General computing questions (Macintosh, Windows, and network/connectivity)	253-1101	computing-help@mit.edu
Academic computing	253-0115	f_l@mit.edu
Administrative applications	253-1101	computing-help@mit.edu
Athena Computing Environment	253-4435	olc@mit.edu
Computer and printer repairs	253-0815	pcservice@mit.edu
Computer presales consulting	253-7686	mcc@mit.edu
Disabilities and computing	253-7808	atic@mit.edu
Telephone and voice mail services	253-3670	telecom-csr@mit.edu
Telephone repairs	253-4357	3help@mit.edu
Unix/Linux	253-1103	unix-linux-help@mit.edu

Surf Sites: Specialized Search Engines

Google, recently customized for MIT (see page 7), has been the most popular Internet search engine for quite a while. Even so, there will still be times when a specialized search engine returns the best results to your queries.

To the right are URLs for several search engines that can help you zero in on what you are looking for. They range from an image search engine to one that canvases engineering, math, and computing databases. To keep up with developments in the world of search engines, check out Search Engine Watch at

`http://searchenginewatch.com/`

A9
`http://A9.com/`

Infomine: Scholarly Internet Resource Collections
`http://infomine.ucr.edu/`

Internet Guide to Engineering, Mathematics and Computing
`http://www.eevl.ac.uk/`

Picsearch – The Search Engine for Pictures and Images
`http://www.picsearch.com/`

Radio-Locator
`http://www.radio-locator.com/`

SpeechBot – Audio Search Using Speech Recognition
`http://speechbot.research.compaq.com/`

Technorati – Searching the World Live Web
`http://technorati.com/`

IS&T Information Services & Technology

is&t is printed with soy inks on recycled paper, and can be recycled in MIT's "mixed paper" bins.