

Carnegie Mellon

Language Technologies Institute &
Institute for Software Research International

Master of Science in Information Technology in
Very Large Information Systems



Fall 2005

Contents

Program	1
Focus	2
Curriculum	3-5
Placement	5
Faculty	6-7
Environment ...	8
Application	9-10
Financial Aid ..	11
Cost of Study...	11
Housing	11
Contact	12
Policies	12



Program



With the drastic drop in the price of storage and the rise of the internet, many organizations have deployed very large information systems, but are not equipped to manage them successfully and efficiently. The goal of the CMU professional degree program Masters of Science in Information Technology, specialization in Very Large Information Systems is to train a new generation of premier technologists who have the skills and knowledge to manage the layers of technology involved in VLIS deployments.

Carnegie Mellon University's Master of Science in Information Technology (MSIT) graduate program is uniquely designed to provide working professionals with this solid foundation. It offers a mix of technology and management courses to provide students with an understanding of information technology from both development and operational perspectives. The MSIT degree is a good investment for both information technologists and business professionals who want to deepen their technical knowledge and develop their management skills.

Very Large Information Systems (VLIS) are large repositories of data can be found in industry, government, military, academic, and scientific settings. They take the form of internet content providers, business transactions, text, video, financial transactions, genomic data, health care management, scientific data sets, etc.

Currently, digital librarians manage the information, while responsibilities for operating a VLIS falls to various system administrators, system architects and database administrators. This diffusion of responsibility results in inconsistent interfaces, a heterogeneous collection of systems that may not interoperate effectively, and a general disjointedness and inefficiency.

Through a comprehensive curriculum, the MSIT-VLIS program trains technologists to coordinate all aspects of VLIS deployments. Graduates will have a unified vision of VLISs after being trained in the areas of Interaction, Analysis, Access, Storage, and Quality.

The MSIT-VLIS curriculum draws from internationally recognized CMU faculty in many departments: Institute for Software Research International, Language Technologies Institute, Center for Automated Learning and Discovery, Electrical and Computer Engineering Department, Human Computer Interaction Institute, Computer Science Department, and the Tepper School of Business.

Focus

The MSIT-VLIS curriculum is unique among top universities world wide. This degree is a first-of-its kind to focus on a broad range of issues essentially to the successful construction, deployment and user of VLISs. In keeping with this objective, students will have a unified vision of VLISs after exposure to the principal aspects of each of these sections.

Access & Update

The access and update layer covers organization of data into various models.

Infrastructure

The infrastructure layer covers hardware and software at the device, network and operating system level.

Interaction

The interaction layer covers the interaction between humans and computers.

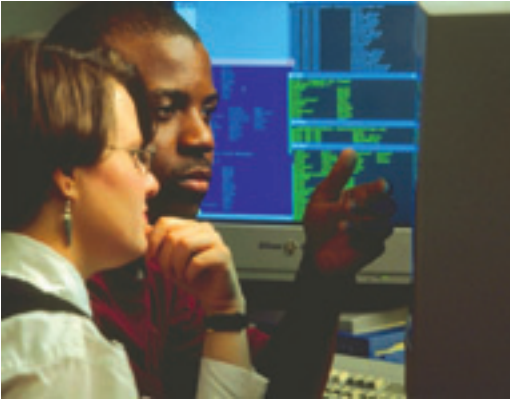
Analysis

The analysis layer covers the extraction of information from data.

Quality

The quality layer covers issues that affects all aspects of a VLIS deployment: performance, reliability, security.

Curriculum



Students will complete 10 courses and a project to earn the degree in one year. The curriculum consists of 6 core requirements, 2 courses in the student's concentration area, 2 electives, and a capstone project.

Core Courses:

Introduction to Computer Security (18-730)

Introduction to Computer Security provides an introduction to techniques for defending against hostile adversaries in modern computer systems and computer networks. Topics include operating system security; network security-cryptography and cryptographic protocols, firewalls, and network denial-of-service attacks and defenses; user authentication technologies; security for network servers; web security; and security for mobile code technologies, such as Java and Javascript. More advanced topics will additionally be covered as time permits, such as: intrusion detection; techniques to provide privacy in Internet applications; and protecting digital content (music, video, software) from unintended use.

Privacy Technology (17-702)

This course introduces students to concepts and methods for creating technologies and related policies with provable guarantees of privacy protection, while allowing society to collect and share person-specific information for many worthy purposes. Methods include those related to the identifiability of data, record linkage, data profiling, data fusion, data anonymity, de-identification, policy specification and enforcement and privacy-preserving data mining. Students get hands-on experience at being "data detectives" by building dossiers from publicly available information and identifying individuals from seemingly anonymous data. Students also learn to be "data protectors" by developing and assessing privacy protocols, algorithms and anonymity protection schemes to protect inferences in shared data. Students learn a 6-prong approach at assessing and constructing technologies that are provably fit for a stated purpose in a social-legal-organizational setting. Emerging technologies examined include: face recognition software, biometrics, surveillance systems, personal information capturing tools and position location technology (GPS, E911 telephones, IR tags). Related topics are drawn from: data mining, information retrieval, web technology, computer security, cryptography, relational databases, statistics and political philosophy.

Information Retrieval (11-741), Spring

This course studies the theory, design, and implementation of text-based information systems. The IR core components of the course include important retrieval models, including Boolean, vector space, probabilistic, inference net, language modeling; clustering algorithms; automatic text categorization; and experimental evaluation. The course covers a variety of current research topics, including cross-lingual retrieval, document summarization, machine learning, and topic detection and tracking.

Artificial Intelligence: Machine Learning (15-681), Fall/Spring

This course focuses on computer programs that automatically improve their performance through experience. This course covers the theory and practice of machine learning from a variety of perspectives. Topics include: decision trees, neural network learning, statistical learning methods, genetic algorithms, Bayesian learning methods, explanation-based learning, and reinforcement learning. We study theoretical concepts such as inductive bias, the PAC and Mistake-bound learning frameworks, minimum description length principle, and Occam’s Razor. Programming assignments include hands-on experiments with various learning algorithms.

Human Computer Interaction for Computer Scientists (05-TBA), Spring

This course introduces the skills and concepts of Human Computer Interaction (HCI) that enable computer scientists to design systems that effectively meet human needs. This course covers iterative design processes, interactive prototype construction, discount evaluation techniques, and the historical context of HCI.

Software Engineering for IT (11-791), Fall

This course covers the fundamentals of software engineering for information technology, including project management and software methodology. A basic understanding of programming is required. Students will analyze, design, and plan a specific software project. There are no programming assignments in this course, but students may implement their plan from this course in 11-792.

Very Large Information Systems (17-TBA), Spring

This course teaches principals of distributed and parallel database technology: distributed query processing, parallel query processing, distributed transaction processing, failure and recovery, federated databases, etc. and provides a forum to combine lessons learned from the core curriculum into a unified framework.

Concentration

Every student is required to complete at least two courses in their selected concentration. The following concentrations are currently available:

Databases

- 18-746 Storage Systems
- 17-801 Privacy Policy, Law, and Technology, or 17-802 Privacy and Anonymity in Data
- 18-821 Mobile Systems
- 18-842 Distributed Computing
- 18-845 Internet Services
- 15-410 Operating Systems

Language Technologies

- 11-711 Algorithms for Natural Language Processing
- 11-718 Conversational Interfaces
- 11-751 Speech Recognition
- 11-731 Machine Translation

Computational Biology

- 15-856 Computational Molecular Biology and Genomics
- 15-879 Computational Structural Biology
- 15-899 Computational Genomics

Management

This concentration is still under construction. We will add an additional course to the two mini-courses listed below.

- 45-775 Business Management
- 45-750 Managerial Economics

Elective Courses:

The following are permitted elective courses students might take:

- Any graduate course in the School of Computer Science
- Any Carnegie Mellon course, if cleared by the program director

In particular, the following mini-courses are encouraged, where appropriate:

- 95-718 Professional Speaking (Mini)
- 95-717 Professional Writing (Mini)

Capstone Project

For the capstone project, students will work at Carnegie Mellon on a research project, or on an industry-sponsored project, during the summer. The project, which consists of three months of full time work related to the student's concentration, must be approved by the program director. Students will record their experiences in a project report.

Placement

Corporations are interested in graduates with intensive technological information management and sufficient technical background to understand VLISs in depth. Graduates will have a deep understanding of the technology of VLISs and will be prepared to manage all aspects of a VLIS from requirements analysis to post-deployment analysis and maintenance. Graduates of this program are eminently qualified for industrial positions.



Faculty



Jamie Callan

Jamie Callan is an associate professor in the School of Computer Science Language Technologies Institute (LTI) and H. John Heinz III School of Public Policy and Management. While his background is in Information Retrieval (IR) and Machine Learning, his interests include a range of information access and analysis topics.

His group's current research is oriented around 4 projects. Federated Search ("Distributed IR"), Accurate Document Filtering, Large-Scale Text Analysis, and IR for Language Applications.

Callan's students initially work closely with him to study specific ideas while learning research skills and IR. As students gain expertise, they develop their own interests and have more freedom in exploring them.

<http://www.cs.cmu.edu/~callan>



Andrew Moore

Andrew Moore is a professor of Robotics and Computer Science at the School of Computer Science. His main research interest is data mining: the exciting world of algorithms for finding all the potentially useful and statistically meaningful patterns in massive sources of data. Data mining is a rewarding area in which to work because the fundamental data structures, algorithms and mathematics are beautiful. And it's a way for Computer Science to have a direct impact in the real world. His research group, The Auton Lab, works with Astrophysicists, Biologists, Marketing Groups, Bioinformaticists, Manufacturers and Chemical Engineers.

<http://www-2.cs.cmu.edu/~awm/>



Eric Nyberg

Eric Nyberg is an Associate Professor at Carnegie Mellon University, with a joint appointment in the School of Computer Science and the Heinz School of Public Policy and Management. He is co-director of the MSIT-VLIS program.

Eric joined the Center for Machine Translation (CMT) at Carnegie Mellon in 1986. Since then, his research projects have focused on Knowledge-Based Machine Translation (KBMT) for practical applications; most notably, the KANT System, which has been deployed for Caterpillar, Inc. When the CMT expanded into the Language Technologies Institute (LTI) in 1996, he became involved in curriculum development and teaching. He currently teaches a two-course sequence in Software Engineering for Information Technology, as well as two laboratory courses in Natural Language Processing and Machine Translation.

<http://www-2.cs.cmu.edu/~ehn>

**Michael Reiter**

Michael Reiter is a professor of Electrical and Computer Engineering and Computer Science, and Technical Director of CyLab. His research interests include computer and communications security and distributed computing.

Reiter regularly publishes and serves on conference organizing committees. He has served as the program chair for the flagship computer security conferences of the Institute of Electrical and Electronics Engineers (IEEE), the Association for Computing Machinery (ACM), and the Internet Society. He currently serves as Editor-in-Chief of ACM Transactions on Information and System Security, on the editorial board of the International Journal of Information Security, and on the Board of Visitors for the Software Engineering Institute. He previously served on the editorial boards of IEEE Transactions on Software Engineering and IEEE Transactions on Dependable and Secure Computing, and as Chair of the IEEE Technical Committee on Security and Privacy.

<http://www.ece.cmu.edu/~reiter/>

**Latanya Sweeney**

Latanya Sweeney is an Associate Professor of Computer Science, Technology and Policy in the School of Computer Science at Carnegie Mellon University. She founded and serves as the Director of the Data Privacy Lab, which works with real-world stakeholders to solve today's privacy technology problems. Her work involves creating technologies and related policies with provable guarantees of privacy protection while allowing society to collect and share person-specific information for many worthy purposes.

Sweeney's work has received awards from numerous organizations, including the American Psychiatric Association, the American Medical Informatics Association, and the Blue Cross Blue Shield Association. Her work has appeared in hundreds of news articles, numerous academic publications, and was even cited in the original publication of the HIPAA Privacy Rule. Companies have licensed and continue to use her privacy technologies.

<http://privacy.cs.cmu.edu/people/sweeney/index.html>.

**Anthony Tomasic**

Anthony Tomasic is a Senior Systems Scientist in the Institute for Software Research International at Carnegie Mellon University. He is co-director of the MSIT-VLIS program. His research interests focus on very large information systems and the application of machine learning to the desktop.

He has worked as an officer for various internet start-ups and a researcher for Dyade, a research and development consortium established by Institut National de Recherche en Informatique et en Automatique (INRIA) and the Group Bull. Tomasic was scientific director for the team of students and engineers that built Disco, a distributed heterogeneous database system.

Tomasic has worked as a researcher for the IBM Almaden Research Center, the European Computer-Industry Research Centre (ECRC), and the Database Group at Stanford University, where he completed his Ph.D. work on the performance of distributed information retrieval search engines. He earned an MS and a Ph.D. from the Department of Computer Science at Princeton University.

Environment



Carnegie Mellon

Carnegie Mellon is a private research university with a distinctive mix of programs in computer science, robotics, engineering, the sciences, business, public policy, fine arts and the humanities. More than 8,000 undergraduate and graduate students receive an education characterized by its focus on creating and implementing solutions to solve real problems, interdisciplinary collaboration and innovation. A small faculty-to-student ratio provides an opportunity for close interaction between students and professors. While technology is pervasive on its 110-acre campus, Carnegie Mellon is also distinctive among leading research universities because of conservatory-like programs in its College of Fine Arts.

For more information, visit www.cmu.edu, www.lti.cmu.edu, www.isri.cmu.edu.

Pittsburgh

Boasting safe neighborhoods, a low cost of living, and an abundance of educational and cultural activities, Pittsburgh has consistently been ranked near the top of Rand McNally's Most Livable Cities index. The downtown district at the confluence of the Allegheny and Monongahela Rivers encompasses fine stores, restaurants, and theaters for the performing arts. Pittsburgh is also the home of major league teams like the Steelers, Pirates, and Penguins. The Oakland neighborhood around Carnegie Mellon juxtaposes international restaurants, coffee houses, shops, alternative cinema, and the Carnegie Museums of Art and Natural History.

For more information, visit www.pittsburgh.net or www.realpittsburgh.com.



Application

Who should apply?

The MSIT-VLIS program is geared toward individuals with a degree in Computer Science, Computer Engineering, or a related degree from a “top-twenty” university, or foreign equivalent. The typical applicant has a minimum of one year of experience in industry and seeks a new job at the bottom tier of the technical management hierarchy. This degree will provide the “leg-up” to management tier.

How to apply

Read the following instructions carefully and make certain that you have met all requirements before you submit your application. All materials must be received by the Admissions Committee no later than July 1, 2005 in order to be considered for admission for the coming Fall semester. The MSIT-VLIS does not accept Spring admissions.

1. Complete a cover sheet that lists the following information: full name, address, telephone number, e-mail address, undergraduate and graduate degrees and cumulative grade point average, and industrial experience.
2. Prepare a Statement of Purpose. Type or print neatly a concise one- or two-page statement in the format below.

Part I: Briefly state your objective in pursuing a professional graduate degree in MSIT-VLIS. Tell us if you have a particular reason for applying to this degree.

Part II: Describe your background in fields particularly relevant to your objective. List here any relevant industrial or commercial experience.

Part III: Include any additional information you wish to supply to the Admissions Committee.

3. Submit transcripts from all undergraduate and graduate institutions attended, even if no degree was granted. Transcripts should be in an official sealed envelope and mailed with your application.
4. Enclose your current resume, including a summary of research & industrial experience and a list of publications (if any). Include copies of any publications (in English only) that you may have.
6. Take the Graduate Record Examination (GRE) and have scores sent directly to Carnegie Mellon University (Institutional Code 2074, Departmental Code 0402). All applicants must take the General Test. It is recommended that applicants to the MSIT-VLIS program submit a subject test. Please refer to the GRE testing schedule (<http://www.gre.org/testdate.html>) to determine test dates. No application will be considered complete until we have received these scores. GRE scores will not be accepted if more than five years old.

7. Take the Test of English as a Foreign Language (TOEFL) and have your score sent directly to Carnegie Mellon University (Institutional Code 2074, Departmental Code 78). All students whose native language is not English must either take this examination, or have earned an undergraduate degree from a U.S. educational institution.

8. Submit the application. Mail all required materials, including the application fee, statement of purpose, sealed transcript(s), and three sealed letters of recommendation in one envelope to:

Janice Kusmierek, MSIT-VLIS Admissions Coordinator
Institute for Software Research International, Newell- Simon 3217
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213-3891
USA

It is your responsibility to make certain that all items listed above and any other materials not included with your application reach the Admissions Committee by July 1, 2005.



Financial Aid

Students find various ways of financing education at Carnegie Mellon, including external scholarships, Federal Stafford Loans, university employment, or payment in full by the student.

Many organizations award external scholarships for Carnegie Mellon students. Students may also finance the cost of education through government secured Stafford Loans. Two other methods involve working for the university in some capacity. Employment is generally available only to students already on campus and these methods usually extend the time required to complete the degree. Some students are able to pay their tuition in full. This method is the most cost effective from a long term point of view.

For more information on graduate student financial aid please visit <http://www.cmu.edu/hub/gradfinaid.html>.

Cost of Study

Full time graduate student cost of attendance is approximately \$40,704 for the current year and approximately \$17,867 for the summer. These figures will rise in the 2005/2006 academic year and 2006 summer. See http://www.cmu.edu/hub/grad_tuition.html for a cost estimate for the current year.

Housing

The Carnegie Mellon Housing Office helps new students find apartments within walking distance of campus or near the Pittsburgh bus or Carnegie Mellon shuttle bus routes.

For more information, visit <http://www.housing.cmu.edu/>.

Contact

Anthony Tomasic is the co-director of the MSIT-VLIS program. He is available to discuss the program with you.

phone: (412) 268 3242
email: tomasic @ cs.cmu.edu

Policies

Program Disclaimer

This document reflects the draft curriculum design of the MSIT-VLIS degree as of March 20th, 2004. The curriculum is under revision so please check back with us for changes.

Equal Opportunity Assurance

Carnegie Mellon University does not discriminate and Carnegie Mellon is required not to discriminate in admission, employment, or administration of its programs or activities on the basis of race, color, national origin, sex or handicap in violation of Title VI of the Civil Rights Act of 1964, Title IX of the Educational Amendments of 1972 and Section 504 of the Rehabilitation Act of 1973 or other federal, state, or local laws or executive orders.

In addition, Carnegie Mellon does not discriminate in admission, employment, or administration of its programs on the basis of religion, creed, ancestry, belief, age, veteran status, sexual orientation or gender identity. Carnegie Mellon does not discriminate in violation of federal, state, or local laws or executive orders. However, in the judgment of the Carnegie Mellon Human Relations Commission, the Presidential Executive Order directing the Department of Defense to follow a policy of “Don’t ask, don’t tell, don’t pursue” excludes openly gay, lesbian and bisexual students from receiving ROTC scholarships or serving in the military. Nevertheless, all ROTC classes at Carnegie Mellon are available to all students.

Inquiries concerning application of these statements should be directed to the provost, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, telephone 412-268-6684 or the vice president for enrollment, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, telephone 412-268-2056.