1999 Intelligent MultiMedia Study Trip

Boston and New York

Preface

At the end of February 1999, the 19 graduate students at Aalborg University studying for their MSc. in Intelligent Multimedia went on a study tour to Boston and New York. They were joined by Dr. Paul Mc Kevitt, Visiting Professor at the Centre for PersonKommunication (CPK) at Aalborg University, and Thomas Rask, an employee of MindPass A/S. MindPass A/S paid all Mr. Rask's expenses and he was not included in the budget for the tour.

The study tour ran from February 26 to March 8, during which the group visited ten of the world's leading research centres (both academic and corporate) in the Boston and New York area. This report was written by the students on the study tour and describes what we saw there.

The students were divided into three areas of responsibility: the Funding group contacted sponsors, the Travel group arranged the travel and accommodation and the Sights group arranged some events for our spare time. Each of the groups maintained homepages, which can be found at http://www.kom.auc.dk/Boston99/. Professor Mc Kevitt arranged our visits to the research centres.

This report is intended for the tour participants and the companies, foundations and institutes who sponsored a major part of the travel expenses and thus made the study tour possible. We would like to thank them all very much. They are:

KMD http://www.kmd.dk
Systematic software engineering http://www.systematic.dk
Henry & Mary Skovs Fond Bosch Telecom Danmark A/S http://www.bosch-telecom.dk
Fischer & Lorenz - European Telecommunications Consultants http://www.fl.dk
Radiometer Medical A/S http://www.radiometer.com
Tele Danmark Research http://www.tdr.dk
Ole Kirks Fond / LEGO A/S http://www.lego.com
Balslev Raagivende Ingenioerer A/S http://www.balslev.dk
Frants Richters Fond Lyngs Industri A/S http://www.lyngsoe-industri.dk
GN Danavox http://www.gn.dk
NetPass A/S http://www.netpass.dk
NKT Holding http://www.nkt.dk
Ingenioeren A/S http://www.ing.dk
MAN B&W http://www.manbw.dk
AN-Instrument Consult AS http://www.angroup.dk
Cheminova http://www.cheminova.dk
RS Radio Parts http://www.rs-radio-parts.dk
Danish Steel House

We would also like to thank the research centres who invited us to visit them:

Massachusetts Institute of Technology (MIT) AI Lab.
MIT Speech Group
MIT Media Lab.
MITRE
BBN
Harvard NLP Lab.
Lucent Technologies (Bell Labs Innovations)
Rutgers University CAIP
New York University NLP Lab.
Columbia University Dept. of Computer Science NLP Group

Each visit consisted of three parts. Firstly, Prof. Mc Kevitt introduced the group and presented the research at CPK. This was followed by a brief overview of the students' projects, after which the host institution presented some of the latest projects and innovations. The presentations of the host institutions often included a tour of their laboratories. The reports that follow describe what we saw on each visit and include links to further information available on the WWW.

Contents

MIT Artificial Intelligence Laboratory

MIT AI Laboratory
545 Technology Square
Cambridge, MA 02139, USA
+1 (617) 253 3699
http://www.ai.mit.edu

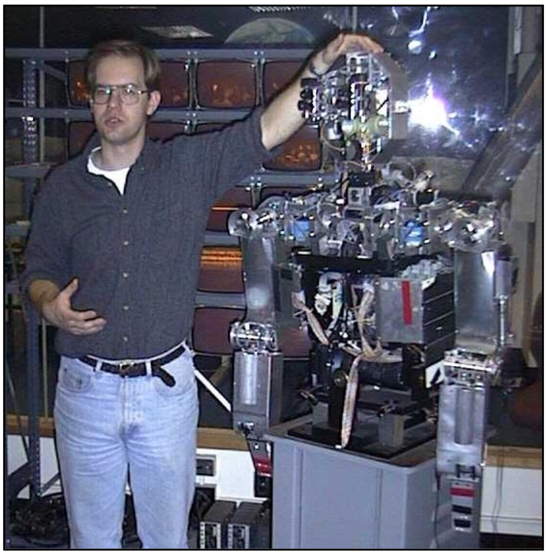
Monday 1st March @ 9:15am – 12pm

Contacts:

Anne Lawthers (ladybug@ai.mit.edu)
Mark Ottensmeyer (canuck@ai.mit.edu)

The MIT Artificial Intelligence Laboratory (AI Lab) conducts research in many aspects of intelligence. The primary goal of the research is to understand human intelligence at all levels including intellectual reasoning, perception, emotion, language, development, learning and social relations. The research is used to build useful artefacts based on human intelligence.

The main fields of research at AI Lab are robotics and machine vision, learning systems, information access, human computer interaction (intelligent environments), virtual and enhanced reality and computing systems and environments. Of these, the major focus seems to be on robotics and machine vision.



Cog, the humanoid robot

Presented projects

Our guide to the lab was Mark Ottensmeyer. He is a PhD student working in the Haptics group at AI Lab, presently looking at providing thermal feedback to users in virtual environments. He presented three projects to us: the Intelligent Room, the START system and the Anatomy Browser. These are described in the following sections.

The Intelligent Room

The Intelligent Room is a highly interactive environment, which uses embedded computation to observe the events happening in the room. The room utilises tracking, gesture recognition, enhanced reality, 3D image reconstruction and speech recognition to give a high level of human computer interaction. The idea is that the room is able to understand the intention of the people who are "using" the room, in a wide variety of applications. The specific application presented was the "Control Center" in which the room's inhabitants can get an overview of weather reports in the case of an approaching storm or really bad weather conditions. The operator was positioned in front of a wall onto which was overlayed an image of some islands in the USA. The presentation showed how commands could be given both verbally and as gestures by the operator to zoom in and out of geographically important places of the USA. One example could be the spoken command "zoom in here" while pointing to the projected image of an island. Information regarding the Intelligent Room is located at the following address: http://www.ai.mit.edu/projects/hci/hci.html

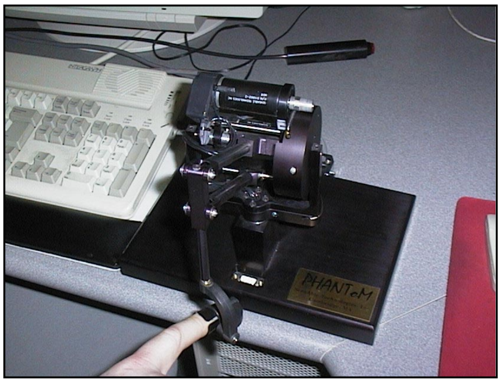
START

The START system is a Natural Language Question Answering System which is used with the world wide web. The system is developed by Boris Katz and has been available to world wide web users since 1993. The idea behind START is for the user to type in queries about information on the internet as questions in English. Hence, it is possible to ask questions such as "What is artificial intelligence?"

and “Show me the weather condition in Denmark?” after which the START server returns a generated web page containing information and links. The application for the system which was demonstrated at the presentation was the defense of the USA. The operator was able to ask questions about boats approaching the US coastline, about stolen nuclear rods and about relating the answers for these queries to the amount of nuclear material needed to produce atomic bombs. A demonstration of the START system can be found at the following web-address: http://www.ai.mit.edu/projects/infolab/.

Interactive Anatomy Browser

The Interactive Anatomy Browser is a tool for visualization and integration of various kinds of medical information. It is an applet programmed in Java which makes it possible to view anatomical structures combining 3D models, original and segmented scans in a convenient way. The browser is meant to be used both for clinical cases and for teaching. Because it uses the world wide web both applications are possible over large distances, although the applet uses a lot of bandwith to transfer the pictures and text involved in the presentation of anatomical structures. The presentation given at the AI Lab showed how it was possible to see the segment of a scanned head in different rotations, while using the mouse to identify various highlighted areas of the scan. An interactive version of the Anatomy Browser can be found at: http://www.ai.mit.edu/projects/anatomy_browser/index.html



The PHANToM haptical interface

Comments

The visit to the AI Lab proved to be one of the exciting ones. Besides demonstrating the above mentioned projects, a tour of the department was given where a number of ongoing projects were briefly displayed. This included the PHANToM haptical interface and COG the humanoid robot. The most impressive thing about the visit was the amount of resources available to the researchers at the lab. Here computer scientists and mechanical engineers work closely together to be able to develop new and interesting haptical interfaces and humanoid robots.

MIT Spoken Language Systems Group

Spoken Language Systems Group
MIT Laboratory for Computer Science
545 Technology Square
Cambridge, MA 02139, USA
(+1) 617.253.8924
http://www.sls.lcs.mit.edu/

Monday 1st March @ 2 – 4pm.

Contact:

Jim Glass (glass@mit.edu)

Introduction

The MIT Spoken Language Systems (SLS) Group is one of the approximately 20 research groups, associated with the MIT Laboratory for Computer Science (LCS). LCS has been or is involved in developing many of the tools and techniques used today, such as RSA, TCP/IP, Time-Sharing, Spreadsheets and Object Oriented Programming and many more. Currently, LCS is focusing its research on the architectures of tomorrow's information infrastructures.

The SLS Group is dedicated to exploring and implementing technologies that allow computers to communicate the way people do, i.e. by speaking and listening. This is also denoted as Conversational Interfaces. The SLS Group has approximately 35 members and is frequently cooperating with the industry partners, such as MITRE, Lockheed or Intel, on research and on setting standards.

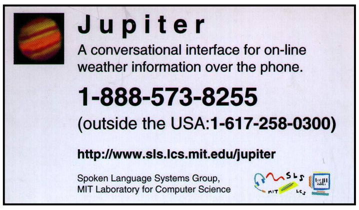
The aim of the group is to make computing more accessible by eliminating the time-consuming series of keyboard entries and mouse clicks and the technical know-how currently required to perform even the most simple information-access operations. Currently the group is working on upgrading the efficiency of application-specific conversations, improving new word detection/learning capability during speech recognition, and increasing the portability of its core technologies and application systems.

For more information about the SLS group please visit http://www.sls.lcs.mit.edu/ or visit LCS at http://www.lcs.mit.edu/.

Presented projects

Victor Zue, Senior Research Scientist, Associate Director of LCS, and head of the SLS group, first introduced the SLS Group and speech technology in general. Afterwards three of the groups systems were introduced.

- *Jupiter* is a conversational system that provides up-to-date weather information over the phone. Jupiter knows about 500+ cities. The user and the machine engage in a spontaneous, interactive conversation, arriving at the desired information in far fewer steps. Throughout the conversation, the computer remembers and builds upon previous exchanges, just as any person would during a conversation about the weather.

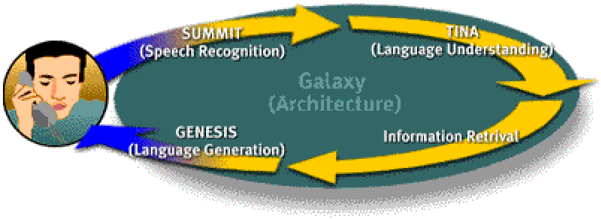


Jupiter
A conversational interface for on-line weather information over the phone.
1-888-573-8255
(outside the USA:**1-617-258-0300)**
http://www.sls.lcs.mit.edu/jupiter
Spoken Language Systems Group,
MIT Laboratory for Computer Science

- *Pegasus* is a conversational interface that provides information about flight status, Pegasus enables users to obtain flight information over a telephone line. It can provide information about flights within the United States, and can answer questions about departure and arrival time for flights that have taken off, landed, or filed a flight plan on the day the user queries the system.
- *Voyager* is a conversational system that can engage in verbal dialogues with users about tourist and travel information for the Greater Boston area. Accessing a database, the system can provide information about local points of interest, maps and aerial photographs of the Boston area, driving directions between local sites, and up-to-the-minute traffic information.

- The three systems mentioned above are based on the *Galaxy* architecture. This is a spoken language system that enables human-computer conversations. Organized as a client/server architecture, the system delivers server-generated responses to lightweight clients, such as PCs and telephones. The SLS Group is currently refining *Galaxy* and extending its range of applications and languages. They are working on upgrading the efficiency of application-specific conversations, improving new word detection/learning capabilities during speech recognition, and increasing the portability of its core technologies and application systems.

The presentation ended with a relaxed talk with the staff and the students of the SLS Group.



The Galaxy Architecture

Comments

The presentation by Victor Zue was well prepared and very professional. At all times Zue was concise and to the point. Zue chairs the Board of Directors for the Linguistic Data Consortium, is a Fellow of the Acoustical Society of America, and currently chairs the Information Science and Technology (ISAT) Study Group for DARPA. In 1994, he was elected Distinguished Lecturer by the IEEE Signal Processing Society.

The systems and the research at the SLS Group can only be considered as state of the art and the presented systems are a part of the newest research within the field of conversational interfaces. The generic framework, *Galaxy*, used for the applications of the group is constantly being extended to improve the human-computer interface.

Currently the group is doing research on:

- Toolkits for highly interactive systems.
- Maximizing the acoustic-phonetic information extracted from a speech signal.
- Segmenting speech for Real-Time Speech Recognition.
- Multilingual systems.

The visit to the SLS Group, Laboratory of Computer Science can only be classified as a success.

MIT Media Lab

The Media Lab

MIT Media Lab
20 Ames Street, E15-231
Cambridge, MA 02139, USA
+1 (617) 253 0383

Tuesday 2nd March @ 9:15am – 12pm

Contacts:

David Martin (david@media.mit.edu)
Raul Fernandez (
Glorianna Davenport (gid@aleve.media.mit.edu)
Justine Cassell (justine@media.mit.edu)

The Media Lab is the most interesting lab in the world from the point of view of Intelligent Multimedia. There they do pioneering research into human/machine interfaces, artificial intelligence, computer vision and other fields.

We arrived at 9.15am to be met by David Martin who was to coordinate our morning at the Media Lab. We were to visit three different groups there: Affective Computing, Interactive Cinema and Gesture and Narrative Language.

Affective Computing

Raul Fernandez, Professor Rosalind Picard's research assistant, was our guide to the Affective Computing group. He talked about how the user can get frustrated with the computer, and how we can measure that with blood pressure or heart rate, perhaps by putting sensors into the mouse. In the future it might be possible to detect the user's frustration by the tone of his voice, or to use vision, speech and gesture to detect emotion from the user. But one of the problems is that each culture has a different meanings for the same movement e.g. in some cultures shaking the head means no whereas in another it means yes. However, Raul mentioned that a system would not have to completely understand every mood of the user to improve communication - sometimes we want to fool the computer just as we fool other humans...

One of the current projects is an intelligent toy (being developed by Dane Courts; see http://www.media.mit.edu/affect/AC_research/projects/Atigger.html) in which the toy has sensors to sense if it is being bounced or squeezed.

It has been shown that by encouraging the user, he will work more with the system. Raul said that he thought the future for the computer is to solve problems that people have not solved and free the restraints of the laboratory.

The entrance to the MIT Media Lab.

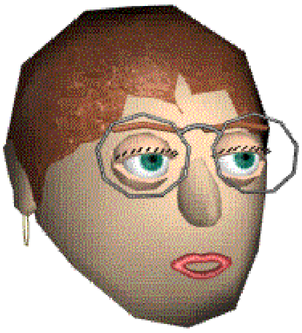
Interactive Cinema

At 10 o'clock we met Glorianna Davenport. She told us about old project in which you could take a virtual tour around Aspen (an American ski resort) using a system with 6 laser disks, and about a physicist who built a cello that could control its own accompaniment. One of her students, Paul Nemirovsky, introduced his GuideShoes project: a system to help small children get home or for people who want to find their way in a new area. The system is a shoe that plays a tune when the user walks in the right direction. As it is now, the user use a base system to select where to go and rule out the areas he wants to avoid. In the future the system will use speech recognition.

Another student, Pengkai Pan, is making a system in which the user can make his own picture story using other people's pictures on the Internet. For example, if the user has taken a trip to France without a camera and wants to tell others about it he can find the pictures on the Internet and put them in his own album. The project is a very simple system to use and also includes a feature to find other people with similar photo stories.

Gesture and Narrative Language

At 11 o'clock we met Justine Cassell and her team who introduced Rea to us. Rea is a virtual Real Estate Agent that uses verbal and non-verbal behaviours to communicate with the user. Rea is an attempt to develop an agent with both propositional and interactional understanding and generation that can interact with the user in real time.



Rea: the real estate agent

She is projected onto a large screen and can recognise the user's gestures (hand and head movements), using two video cameras mounted to either side. She also recognises what the user is saying and can reply using spoken language.

During the conversation, Rea monitors the user to see if he wants to take a turn in the dialogue. When the user wants more information, he can interrupt Rea and ask her by saying, for example, "Tell me more about the bedrooms." Rea then changes the subject and gives the user more detail about the bedrooms in the house. When it is Rea's turn, she uses her voice and gestures to talk to the user and when it's the user's turn she nods and makes small comments to let the user know she is following the conversation.

Rea is the next step onwards from Ymir, a previous system developed by the Gesture and Narrative Language group (see http://kris.www.media.mit.edu/people/kris/ymir.html). Ymir was a system in which an agent, Gandalf, was able to discuss a graphical model of the solar system in an educational application. Gandalf recognised and displayed interactional information such as gaze and simple gestures. However, it had limited ability to recognise and generate propositional information, such as providing correct intonation for speech emphasis on speech output, or a gesture co-occurring with speech. Ymir used nine computers and a body suit to communicate with the user, whereas Rea uses five computers (with two of them used solely for her vision) and no body suit.

In the future, the group intend to connect a third camera for face recognition. For more information see http://gn.www.media.mit.edu/groups/gn/projects/humanoid/.

Finally, we met one of Justine Cassell's students, Kimiko Ryokai, who works on the Story Mat project (http://gn.www.media.mit.edu/groups/gn/projects/storymat/). She uses a projector to project an image down onto a small play mat. Children use the pictures on the mat and several soft toys to tell a story. The system records their stories and can play them back at a later time, either to inspire other children or to remind the original storyteller.

Summary

The visit to the Media Lab was very inspiring. This is the place where much of the exciting Intelligent Multimedia research is happening. They gave us a good picture of what they are doing now and showed us what the future could be.

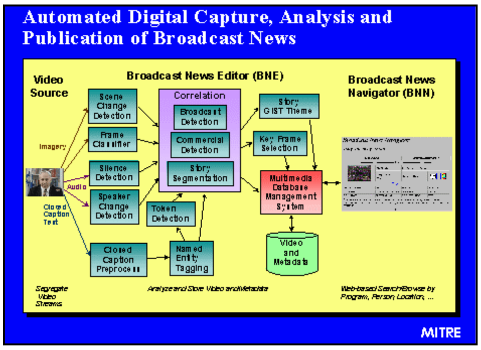
MITRE Corporation

Bedford Complex
202 Burlington Road
Bedford, MA 01730-1420
+1 (781) 271-2000

Tuesday 2nd March @ 2:00 – 5:00pm.

Contacts:

Dr. Mark T. Maybury (maybury@mitre.org)
Frank Linton (linton@mail11.mitre.org)



Automated Digital Capture, Analysis and Publication of Broadcast News

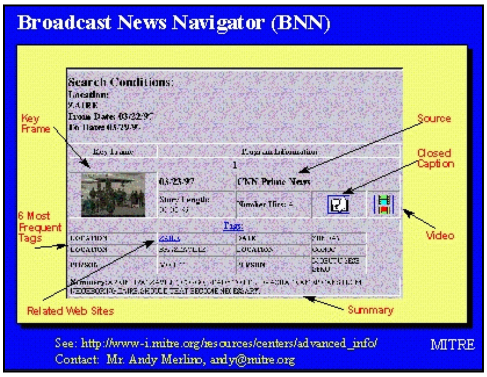
MITRE's broadcast news architecture (slide kindly provided by MITRE)

Dr. Mark T. Maybury, executive director at the Information Technology Center, and Frank N. Linton, hosted our visit to MITRE. MITRE is located in Bedford, Massachusetts, and is a not-for-profit company working closely with the US Department of Defence. It is sponsored by the Defense Advanced Research Project Agency (DARPA) amongst others. The main goal of the MITRE corporation is, as Dr. Maybury said: “To make the US government smart.” MITRE therefore addresses areas of critical national importance by combining state-of-the-art information technology and system engineering. Detailed information about the MITRE corporation and ongoing research can be found on the Internet at http://www.mitre.org. This web site also has more detailed information about the projects mentioned below and even includes a page from which you can download the software for the Collaborate Virtual Workspace system (recently released under an Open Source agreement).

The MITRE personnel presented four projects to us: two information extraction systems (the Broadcast News Navigator and GeoNode) and two collaborative systems (the Collaborative Virtual Workspace and Collaborative Learning).

Broadcast News Navigator

Stanley M. Boykin presented the Broadcast News Navigator (BNN). This is an information extraction system that makes it possible to browse news broadcast around the world. A system like BNN makes it possible to give an overview of the news and monitor inconsistencies. As an example, the Iraqi news could state something about a certain topic while CNN could have a different story on the same topic. Gathering all information on this topic from both television stations and extracting keywords and images enables a user to see the differences and search for correlated stories.



Broadcast News Navigator (BNN)

Search Conditions:
Location:
ZAIRE
From Date: 03/22/97
To Date: 03/29/97

Key Frame | Program Information
1
03/23/97 CNN Prime News
Story Length: 00:00:45
Number Hits: 4
Tags:
LOCATION ZAIRE
LOCATION
PERSON

See: http://www-i.mitre.org/resources/centers/advanced_info/
Contact: Mr. Andy Merlino, andy@mitre.org
MITRE

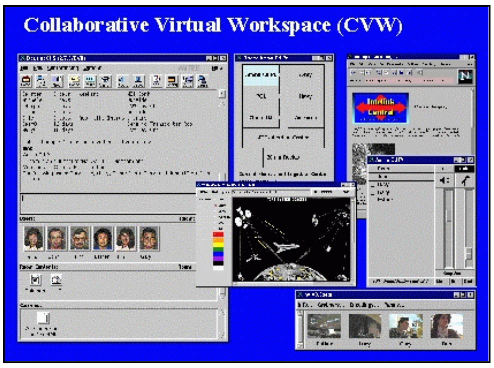
To be able to make such comparisons, the news reports need to be annotated as to what subjects they discuss. Manual annotation would be very costly, considering the increasing amount of worldwide news broadcasts. MITRE therefore developed an automatic system capable of annotating and storing digital imagery, audio, video and text. Automated multimedia mechanisms were needed to: capture, annotate, segment, summarize, browse, search, visualize and disseminate these media. The diagram at the top of the page illustrates the three parts of the system. The video source is input into the Broadcast News Editor (BNE), which handles the scene description, information extraction, links to related stories etc. The resulting information is then made available in a database, which is accessed and visualised by the BNN (see the diagram above, kindly provided by MITRE).

MITRE is currently doing key frame extraction, story segmentation and speech transcription on 10 television channels. The speech can be transcribed at 40 times its original speed and MITRE is still working to improve on this. One of the discoveries made when the system was tested on Russian, Chinese, and Indian news broadcasts was that the story format is country independent. Therefore the only thing missing for the system to be operational on news broadcasts world-wide is a speech transcriber for each language.

Geospatial News on Demand Environment

Robert S. Hyland demonstrated the Geospatial News on Demand Environment (GeoNode) using 3 years of stored CNN broadcasts. GeoNode is based on BNN but adds the ability to follow linked topics by time and geographical location. As an example, if a user was interested in a certain topic, it is possible to track that topic over time and have graphics presented showing when the topic was first mentioned, increasing story coverage, the time when the topic was mentioned the most, graphical location, etc. Some of the possible applications for the GeoNode system include history education and animation of events.

Collaborate Virtual Workspace



Deborah A. G. Ercolini presented the Collaborate Virtual Workspace (CVW), a prototype collaborative computing environment designed to support both temporally and geographically dispersed work teams. CVW enables people to converse, collaborate and interact regardless of their geographic location. The user of the system navigates around in a virtual building consisting of rooms and floors. People can meet and chat in this environment, communicate via audio or video, make illustrations on drawing boards etc. The user decides how to communicate, and the application automatically sets up the required link to the involved persons. Users can even lock rooms and communicate privately within and between rooms. The diagram above (kindly provided by MITRE) illustrates the CVW chat environment, floor plan, drawing board, Internet browser and audio and video interface.

Document sharing plays a central part in the CVW system. When a person is located in a certain room, he or she can access specific documents, which can then be checked in and out for single-user editing. Information about who made changes to the document is stored together with the time of editing.

Collaborative Learning

The last research project presented to us at MITRE was the Collaborative Learning system, presented by Robert D. Gaimari. The student does not really collaborate with other students, but with a virtual “peer” who pops up and asks questions. These questions force the student to justify her decisions and, in doing so, to think more deeply about the problem at hand. There is also a virtual tutor in the system, who is always right, but Robert suggested that the peer is better in some ways. For example, the virtual peer will sometimes tell the student she is wrong when she is really correct. She is then forced to argue her case and convince the virtual peer of her decision.

The system uses a special way of communicating based on “Speech Acts”, in which the student can only speak by choosing one of 8 different sentence formats. The student can choose from: Request, Inform, Motivate, Maintenance, Task, Acknowledge, Argue and Mediate, each of which starts off a sentence. For example, if the student chose the Request act, she would have to start her sentence by saying, “Could you tell me about ...” By making these intentions explicit, the system can respond more easily to the student’s communications.

Summary

The visit to MITRE was a unique experience that was very useful to our study in Intelligent MultiMedia at Aalborg University. It was especially of interest to the project group working on the Intelligent Internet Browser since one of the main topics in their project is information extraction. To see state-of-the-art information technology applied at this level of professionalism has given all of us lots of new input as to what is possible today.

BBN

BBN Technologies
70 Fawcett St.
Mailstop 15/1b
Cambridge, MA 02138, USA
+1 (617) 873 4262

Wednesday 3rd March @ 10am – 1pm.

Contacts:

Josh Bers (jbers@bbn.com)
Ralph Weischedel (weisched@bbn.com)

BBN is working within speech technology and information extraction/retrieval, both as R&D and with a range of commercial products. The company has approximately 100 employees and its homepage is located at http://www.bbn.com/.

The company was founded 50 years ago and is situated a little outside Boston. Their solutions have recently moved towards statistics based techniques (e.g. Hidden Markov Models for speech recognition). The product range includes speech coaching, speech recognisers and information extraction tools.

GTE BBN's offices in Boston, MA

The speech recogniser team at BBN is currently working with two recognisers, an off-the-shelf commercial recogniser called HARK and another called BYBLOS which is a research recogniser undergoing continuous development. New techniques are tested in BYBLOS and when they have matured they are moved to HARK.

The information extraction/retrieval team at BBN is working on programs that are tested at the DARPA MUC/TREC (Message Understanding/Text Retrieval) conferences each year. The idea of information extraction programs is to extract keywords from free text, i.e. the current programs find named entities, events and relationships from broadcast news articles. In a combination of these two fields, BBN also is working on machine translation.

As an extension of the above BBN presented the following projects:

VoiceLog

VoiceLog is a project incorporating logistics, thin clients, speech recognition, OCR and portable computing. The product is a slate laptop connected by a wireless 14.4K modem to a server which facilitates speech recognition, exploded views/diagrams of military vehicles and direct connection to logistics. The idea is that a person in the field has support for specifying what is damaged on a vehicle using diagrams, and for ordering the parts needed to repair the vehicle. The laptop accepts spoken input (recognised on the server) and touch screen pen input. The visual part of the system consists of web pages showing diagrams and order forms which is supported by a program controlling the speech interface. Trying it out shows that speech is actually a nice way of referring to objects, pages, etc. which are not visible on the screen at the moment.

Dialogue work

Within dialogue systems BBN is working with a generic speech recogniser from which specific applications inherit a basic framework. System under development are E-Mall (a phone-based grocery store) and Talk'n'Travel (a travel agency that can book flights automatically), which both are replacements for phone services. These systems are voice driven, but you are still working within a tree structure of possibilities.

Information extraction

BBN is also involved in development of a system for segmentation and classification of TV news reports. This system does transcription of the dialogue, NLP and voice recognition. The idea is that you should be capable of following a story from several views and find specific information later.

Comments

The projects and the demos of the systems underlined that BBN are doing well compared to the other places we saw on the study tour. The fact that we could try out the VoiceLog system was very good. In case you are interested in a job at BBN your resume should be sent to jbers@bbn.com or dcullen@bbn.com

Harvard University AI and Natural Language Processing

EECS – Engineering and Sciences Laboratory,
40 Oxford Street
Cambridge, MA 02138 USA
(+1) 617 495 2081
http://www.eecs.harvard.edu/ai/

Wednesday March 3rd @ 2.00 – 5.00pm.

Contacts:

Wheeler Ruml, Ph.D Candidate in Computer Science (ruml@eecs.harvard.edu)
Luke Hunsberger, Ph.D. Candidate in Computer Science (luke@eecs.harvard.edu)

Introduction

EECS (Electrical Engineering and Computer Science) is a computing environment for academic research within the Division of Engineering and Applied Sciences at Harvard University. They are currently housed in the Engineering Sciences Lab and Pierce Hall on Oxford Street in Cambridge, Boston. As part of AI and Natural Language Processing Studies at EECS, staff and students are developing theories about the behaviour of intelligent communication systems, with the aim of producing computer systems that can interact effectively with human beings in both the linguistic and graphic media.

IMM students presenting their project at Harvard

Our guide was Wheeler Ruml, a PhD student studying the overlap of Search Method Optimisation, AI and Cognitive Science. After our presentations, we continued in another lecture room where Wheeler and Luke Hunsberger presented some of their work including demonstrations taped on video. Luke does research on Multi-Agent Systems based on Shared Plans (the Shared Plans Theory of Collaboration is a theory by Barbara Grosz and Sarit Kraus).

GigAgents

This is an agent architecture following a general theory (Shared Plans) of collaborative planning that accommodates multilevel action decomposition hierarchies and explains the process of expanding partial plans into full plans. Software agents collaborate by having intentions and individual and mutual beliefs about the task but also about the capabilities and commitments of the agents involved. To achieve a goal, the agents collaborate to make a plan consisting of various tasks to complete. Before the tasks are executed, the agents flesh out a plan in which particular agents execute sub-tasks. These agents then individually execute the sub-tasks until the goal is reached. This technology is especially useful in Human Computer Interfaces where the agents can track which actions the user has done or has missed out. The original application was to have been for musicians booking performances (gigs) but Luke mentioned that it could be applied to many other areas.

Luke Hunsberger gave a highly theoretical talk, being a Math freak, about the internals of GigAgent. More details can be found in his paper at http://www.eecs.harvard.edu/~luke/pss/atal.final.ps

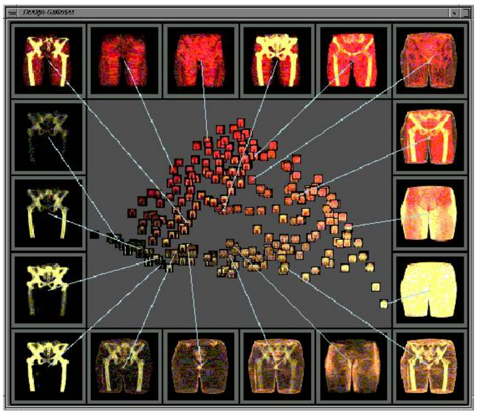
DIAL (a predecessor to GigAgent)

A collaborative web interface for distance learning. System administrators in the Computer Science department found that they got the same questions and consequently repeated the same answers to ignorant or inexperienced students. To solve this problem DIAL was used to help students pull out relevant information when problems occur. The system tries to figure out the goals of the user via agents creating and sharing plans to find relevant information (e.g. from previous sessions).

GLIDE

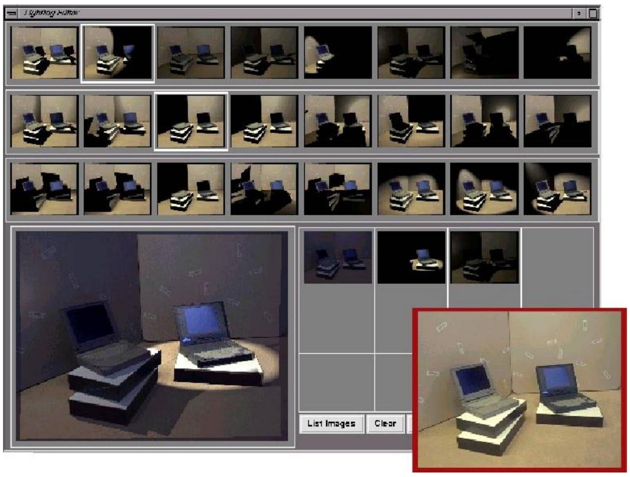
An application for graph drawing, like flow charts, also based on a multi-agent technology. A drawing system like Paint Shop does not know that the user is drawing a box that should be aligned according other boxes. In contrast, GLIDE works by recognising the intentions of the graph designer. The intentions are cast to agents as constraints for how the graph should look. Following these constraints guided by beliefs of “constitutionalised” perception of graphs, graphs can be laid out quickly. The system makes it very easy to group boxes, set alignments and make (and keep) other such constraints. A spring simulation is used to move objects interactively to satisfy all constraints: the objects move around as if they had springs wherever the user has set up constraints. The user can move items around manually and watch them glide into the shape that best satisfies those constraints. A demonstration of GLIDE was shown on video.

Design Galleries



In computer aided graphic design, the designer is faced with the tedious task of tweaking parameters to get an optimal result. After each tweak, the imaging process can take several hours. The Design Gallery is an approach in which the computer runs through the entire parameter space offline and then represents the possible graphics to the designer, allowing him to choose the best. The generated graphics are browsed in either 2D or 3D.

The first situation that was demonstrated was selecting parameters for volume rendering of scans of the human body. The browser presents 2D thumbnail images, grouped according to perceptually similarity. The parameter space can be explored by panning and zooming.



The next demonstration was for light selection and placement in 3D animations. The system selects a set of complementary lights automatically. Corresponding thumbnail images are organised in a three-level hierarchy, shown at the top of the display. The user can browse this hierarchy by clicking on images in the first and second levels, which correspond to interior nodes of the hierarchy structure. Interesting lights/images can be moved to the palette, where their intensities and temperature can be controlled via pop-up sliders. The image in the lower left is a full-size combination of the lights/images that the user has chosen in the palette to the lower right, and is updated automatically.

The parameter space could also be presented using a 3D VRML browser, where the user could navigate freely, though the images were still in 2D.

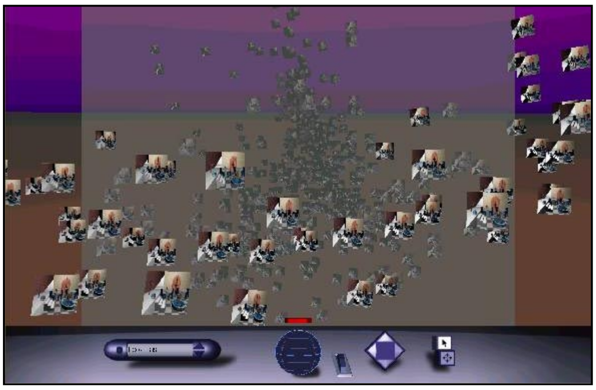
The current list of Design Gallery investigations include:

- Light selection and placement for standard rendering.
- Light selection and placement for image-based rendering.
- Transparency- and colour-function specification for volume rendering.
- Control of particle-system animations.

- Control of articulated-figure animations.

More information on Design Galleries can be found at http://www.merl.com:80/projects/design/index.html

Papers discussing Design Galleries (and the source of the screen shots) can be found at http://www.eecs.harvard.edu/cgi-bin/ruml/wr-log?papers/gd97-final.ps.gz and http://www.eecs.harvard.edu/cgi-bin/ruml/wr-log?papers/dg-siggraph-final.ps.gz



Comments

Though all the demonstrations were only shown on video, we saw some very interesting projects. We did not have the opportunity to tour the rest of the EECS. While this may have been because they are in the process of moving from several locations across campus to one building, it would have been nice to get an impression of their laboratories.

Lucent Technologies

Dialogue Systems Research Department,
Multimedia Communications Research Laboratory
700 Mountain Avenue
Murray Hill, NJ 07974 USA
+1 (888) 458 2368

Thursday 4th March @ 9am – 12.30pm.

Contact:

Chin-Hui Lee, Head of Department (chl@research.bell-labs.com)

Introduction

Lucent Technologies (Bell Labs Innovations) is the leading company in the global communications industry. Worldwide, Lucent Technologies has about 24,000 researchers and scientists at locations in 20 nations. The research at Lucent Technologies has resulted in more than 25,000 active patents and eight Nobel Laureates. Some of the main historical inventions made at Lucent are: the transistor (1947), the laser (1958), UNIX (1969), C++ (1985) and digital cellular telephony (1988).

Presented projects

Chin-Hui Lee, Head of the Dialogue System Research Department, first gave a short introduction to research group in general. Afterwards seven different projects in the area of multimedia communication were presented.

- "**3D Acoustic Modeling**" is a spatialised audio project to be used in virtual reality. Unlike similar efforts, this project aims to provide real-time 3D audio for complex 3D environments (up to 10,000 polygons). The core technology involved is Beam Tracing, in which an acoustic model of all the directions in which sound can reflect is precalculated for the 3D model. This means that the run time system can update the stereo sound output to the user 6 times a second with static sources or a static receiver. The group is now building a real room so that they can check their simulated results and research tradeoffs between audio and visual information in a virtual environment.
- "**Lucent Vision**" is a system, used for real-time tracking of the players and the ball in a tennis match. This tracking is used to generate new kinds of statistics about the players and the game. For example, the system can display the percentage of time that the players spend in different parts of the court. A future application would be virtual replays of the game from any angle.
- "**Hands Free Speech Interaction**" is a spoken dialogue interface, in which the user has the ability to interrupt the system output by talking to the system. This is done using advanced echo cancellation techniques, so that the system's own voice is factored out.

- "**MPEG 4 Animation of Face and Body.**" The new digital video compression standard includes features to represent face and body motions directly. Version 1 of the standard is already complete and it specifies the parameters for facial movements and expressions. This means that facial expressions can be transmitted at less than 2 Kbits/sec and integrated with graphics at the client to make a realistic (or cartoon) face appear to talk, smile and convey other emotions. Eric Petajan at Bell Labs has also been working on facial feature extraction from video input.
- "**EMU (email markup language)**" is system for reading emails. The idea is that the system recognises the structure of the text and marks it up according to meaning, improving the quality of the output. For example, it knows how to read the headers at the top of an email, skipping all the "Received:" lines and going straight to the "Subject:" and "Sender:" lines. The system is also implemented in various languages.
- "**Virtual Visual Interior Design (V2ID)**" is a system to visualize changes in the interior design of a room. The system is given one photo, which it interprets by using Hough transforms and vanishing points. It then allows the user to choose new colours and patterns for the furniture and decorations in the room based on the original ones. The system also allows the user to navigate around the room to some extent.
- "**Natural Language Call Routing**" is a product to replace human operators or hierarchical touch tone menus for call routing in a bank. Instead, the system recognises the user's speech and tries to route the call to the relevant department. The system uses a vector-based approach, matching requests to destinations using a cosine match in parameter space. If the system discovers several possible destinations, it then asks the user a question to disambiguate their query. The disambiguation questions are directly based on the vector representation. This system is in use and performs better than a human operator for a bank with 23 possible routing destinations.

Summary

Lucent Technologies (Bell Labs Innovations) is still leading the way in many technologies from image processing to information retrieval. Unlike some of the other research labs we visited, Lucent's products are being used all over the world in commercial applications. In fact, their text-to-speech system is being used by Columbia University for their MAGIC project.

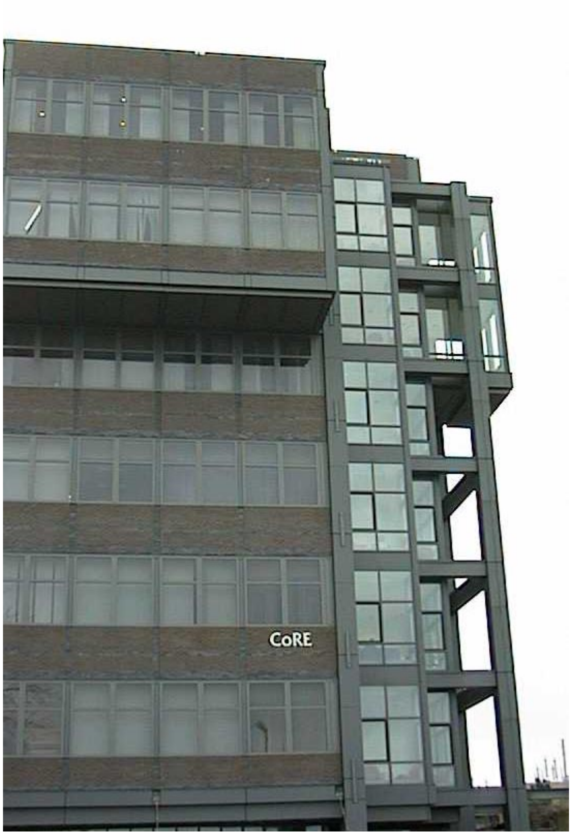
Rutgers University CAIP

Center for Computer Aids for Industrial Productivity (CAIP)
Rutgers University
96 Frelinghuysen Road
Piscataway, NJ 08854, USA
+1 (732) 445 3443

Thursday 4th March @ 2 – 5pm.

Contact:

Prof. Jim Flanagan, Director (jlf@caip.rutgers.edu)



The CoRE building
which houses CAIP at Rutgers University

Introduction

Rutgers University is the State University of New Jersey. In the past decade, the University has developed particular strengths in such areas as bioinformatics, computational design, computational neuroscience, distributed computing, human-machine interface, electronic commerce and wireless communications.

Advancing the science of information is integral to the mission of Rutgers, the flagship public research university in a key high-technology state. The Rutgers agenda is to explore new information theories and technologies, prepare graduates for demands of the Information Age, and provide technological expertise that is crucial to the economy of the state and nation.

Interdisciplinary collaboration among their nearly 2,500 faculty members and 48,000 students on campuses in Camden, Newark, New Brunswick allows the university to optimise its information resources and create new ways to learn and to solve problems.

During our stay at Rutgers University we visited the “Center for Computer Aids for Industrial Productivity” (CAIP), which is an Advanced Technology Center established in early 1985 and jointly supported by the New Jersey Commission on Science and Technology, Rutgers University and industry membership.

More information can be found at http://www.caip.rutgers.edu/.

The main research interest areas of CAIP are:

- Integration of voice, image and tactile modalities into multimedia information systems.
- Application of advanced computing technologies to computer-aided design, scientific visualisation, digital signal processing, machine vision and robotics.
- SiC power devices.
- Reactive Chemical Mixing in Chaotic Flows.
- Developing a Distributed System for Collaborative Information Processing and Learning (DISCIPLE).

The research at CAIP is currently concentrating on three projects: DISCIPLE (development of a collaborative environment in Java), STIMULATE (development of multimodal interfaces in which no keyboards are used) and KDI (development of multimodal collaborative environments over wired and wireless networks).

Projects on the horizon will involve smart rooms, multimedia digital libraries, multimodal browsers and computational models of information.

IMM students and Prof. Mc Kevitt on a tour of CAIP

Presented projects

During our stay at CAIP we had the chance to visit the Multimedia Information Systems Laboratory which conducts research on human computer interaction. Technical projects include speech recognition, speech synthesis, hands-free sound pickup, audio and video coding, distributed computing, database query by image content and high-speed data networking. The primary focus of the laboratory is to make it easier for humans to communicate with complex computers and with each other and to access the vast amounts of information stored online.

Further information on each research group and project is available at http://www.caip.rutgers.edu/multimedia/.

One of CAIP's big research efforts is on the creation of Synergistic Multimodal Communication Collaborative Multiuser Environments. This involves the development of a multiuser, collaborative environment with multimodal human/machine communication in the dimensions of sight, sound and touch. The network vehicle (called DISCIPLE, for Distributed System for Collaborative Information Processing and Learning) is an object-oriented GroupWare (presently evolving under DARPA sponsorship) running on the Internet over TCP/IP as well as over ATM on the intracampus network.

At three user stations, CAIP-developed technologies for sight (eye tracking, image and face recognition), sound (automatic speech and speaker recognition, speech synthesis, distant-talking auto directive microphone arrays) and touch (gesture and position sensing, force-feedback gloves, and multitasking tactile software) are integrated into DISCIPLE for simultaneous multimodal use. The system so constituted provides a test bed for measuring benefits and synergies. With participation from cognitive science and human-factors engineering, realistic application scenarios have been designed to evaluate combinations of modalities and to quantify performance.

Application scenarios that might be served by the system embrace activities as disparate as collaborative design, cooperative data analysis and manipulation, battlefield management, corporate decision making and telemedicine. Further information can be found at http://www.caip.rutgers.edu/multimedia/multimodal/.

In this context we had some live demonstrations in the Acoustics and Vision Labs.

Acoustics Lab.

The Acoustics lab performs advanced research in the area of high quality sound capture. The current research has a particular focus on active and passive microphone arrays. Microphone arrays enable the capture of high quality audio waveforms from remote sound sources, under adverse acoustic conditions. In particular, arrays allow the tracking and recording of moving human talkers without requiring the use of a cumbersome tethered microphone.

Recent advances in circuit technology have made digital signal processing systems capable of real-time processing of multiple audio channels practical to implement and readily available. This has been combined with an ever-increasing demand in voice controlled applications that exhibit robust, environment independent performance. Microphone array systems have many uses for sound capture and acoustic source location in applications such as cellular telephony, video teleconferencing, and audio interfacing with PC systems.

A live demonstration was performed with a system, which had eight microphones in two arrays and a camera. When the system detected someone speaking it would search for the speaker and focus the camera on him. Currently they are working on coping with multiple speakers and tracking.

See http://www.caip.rutgers.edu/multimedia/microphone.arrays/ for more information.

Vision Lab.

Several experimental multimodal systems were constructed in 2D and 3D. We had two live demonstrations:

- The 2D demo was based on a system were the system displayed a map on the screen. The user can place military hardware on it, such as a plane or an airport, by speaking the commands ("Place an airport") and looking at the place on the map where he would like to put it. The system tracks the eye position of the user using a camera mounted above the screen. It can indicate the point at which the user is looking with a cursor but this feedback confuses the user, since his eye is drawn to the cursor and not to what he wants to look at.
- The 3D demo was similar except the display was of a 3D landscape and the user interacted with a data glove instead of an eye tracker. Unlike the approach used at the MIT AI lab, the glove in this case gave no haptic feedback.

The Vision Lab. at CAIP

Comments

It was exciting for us to be in one of the biggest universities in the area and to meet James Flanagan, Director of the CAIP Center, who was the inventor of the microphone array and a director of Bell Labs.

Rutgers University has a large area of research, well funded by industry and the government and is one of the best in their research fields.

New York University NLP Group

Computer Science Department
Warren Weaver Hall, Room 405
251 Mercer Street
New York, NY 10012 USA
(+1) 212 998 3011
http://www.cs.nyu.edu/

Friday March 5th @ 10 – 12pm.

Contacts:

Prof. Ralph Grishman (grishman@cs.nyu.edu)
Andrew Borthwick (borthwick@cs.nyu.edu)

Introduction

The Natural Language Processing group is part of the Computer Science Department of New York University. Their offices are placed on 7th floor of the Computer Science building on Broad Way.

Our host was Prof. Ralph Grishman, who is focusing his research on using NLP for information extraction and knowledge acquisition. After our presentations, Ralph and one of his Ph.D. students, Andrew Borthwick, presented some of their work, though without any demonstrations.

Information Extraction

Ralph talked about their general research in information extraction so far. Their interest lies in extracting information from large natural language corpora. The information is put into various categories, as defined by the Defence Advanced Research Projects Agency (DARPA): person, location, organisation, date, time, percentage and currency (sometimes a junk category is used to catch everything else). One application of this is to extract specific events of interest from a new report. Another is automatic indexing for books, spotting named entities and registering what pages they occurred on.

With a knowledge extraction system, text can also be translated from one language to another by expressing the extracted knowledge. Naturally the quality of machine translation relies on the quality and granularity of the knowledge extracted.

The basic problem is the acquisition of the knowledge used for the Natural Language Processing. Good meta-knowledge of lexical items, syntax and semantic relations and patterns is needed. Most current NLP systems rely on lexicons of lexical and semantic structures that are built by hand. This is a tedious and error prone process and requires linguistic expertise.

The PROTEUS project at NYU (http://cs.nyu.edu/cs/projects/proteus/) includes such hand-built, rule-based lexicons (COMLEX and NOMLEX) but also includes automatic grammar acquisition systems (such as the Apple Pie Parser). In contrast to rule-based systems, requiring new rules to be specified for new languages, the automatic systems simply need to be retrained on a new corpus of material.

Machine translation is possible by learning the correspondences between parallel bilingual corpora. The PROTEUS system is most effective for domain-specific translation, where rule-based systems may give translate specialised words badly. However, rule-based translation is still better in many domains for generic translation.

Maximum Entropy Named Entity

Andrew's research is based on maximum entropy modelling techniques for training information extraction models without the need for exhaustive rule making.

The Maximum Entropy in Named Entity recognition system (MENE) is still under development, but offers several advantages over both the hand-coded systems and the Hidden Markov Model systems used by BBN and others. Firstly, the maximum entropy method does not require large amounts of human work - it can work out its own model from a given corpus of text. Secondly, the method allows several features of each phrase to be taken into account at once, as opposed to the HMM methods which can only use one feature per phrase.

This second advantage leads to another: weak features can be input into MENE without upsetting the results - they are simply assigned low probabilities in the model. This means that MENE can make use of hand-coded systems by simply considering their output as yet another feature. This combination of the two recognition styles has proved to be very powerful.

Comments

In contrast to the very technical presentations by our hosts it would have been nice to see some demonstrations. The university was undergoing some big modernisation so the whole building was in a state, leaving us without a tour. Fortunately, this being our penultimate visit we had already had plenty of exercise and were quite happy as NYU was the first non-commercial site to serve snacks and drinks.

Columbia University, New York

Department of Computer Science
450 Computer Science Building
Columbia University
1214 Amsterdam Avenue, Mailcode: 0401
New York, NY 10027-7003, USA
+1 (212) 939 7117

Friday 5th March @ 2 – 4pm.

Contacts:

Shimei Pan (pan@cs.columbia.edu)
James Shaw (shaw@cs.columbia.edu)

Columbia was the final visit of our study tour and consequently quite brief. However, the short time we were there was more than enough to impress us. The campus is a large open area in upper Manhattan, which itself is an impressive achievement since this is a city in which open space (especially with grass and trees) is at a premium.

We were there to visit the Department of Computer Science, headed by Professor Kathleen McKeown. Unfortunately, Professor McKeown and most of the staff were away at a conference and so our guide was Dr. Michelle Zhou, who has recently graduated and is now employed by IBM.

MAGIC

The multimedia research at Columbia mainly focuses on automatic generation as opposed to using multimodal input to understand more of the user's intentions. Their latest project is called MAGIC (Multimedia Abstract Generation for Intensive Care), a testbed system for generating multimedia presentations which inform caregivers about the status of a patient who has undergone heart bypass surgery. The system takes the data from a relational database and automatically creates a presentation using graphics, speech and text. These three modalities are coordinated so that, for example, pop up text boxes appear for long enough to be read and are synchronised with the associated speech. The spoken output is not the same as the text shown on the screen, since people have different expectations of spoken words and written words.

The MAGIC system also uses its knowledge of what it is presenting to compact sentences together so that several simple sentences can be joined into one more compact (but more complicated) one. The system then uses prosody to make these new sentences easy to understand. For example, given the patient's name, age, doctor and history, it can manufacture the sentence, "Ms. Walker is a 50 year old anorexic, hypokalemic female patient of doctor Longman undergoing CABG."

Shimei Pan and James Shaw gave a demonstration of the system, generating a slick and easy to understand (even to a non-medic!) presentation automatically from raw data in a couple of minutes.

Summary

Columbia University has some very exciting technologies in the field of natural language generation, meaning-to-speech and intelligent graphical presentation. We have made some contacts that will be very useful to some of our projects.

CHAMELEON
A platform for developing Intelligent MultiMedia applications

Paul Mc Kevitt

Visiting Professor
Center for PersonKommunikation (CPK),
Aalborg University, DENMARK

EPSRC Advanced Fellow
Department of Computer Science,
University of Sheffield, ENGLAND

Intelligent MultiMedia (IntelliMedia) focuses on the computer processing and understanding of signal and symbol input from at least speech, text and visual images in terms of semantic representations. We have developed a general suite of tools in the form of a software and hardware platform called “CHAMELEON” that can be tailored to conducting IntelliMedia in various application domains. CHAMELEON has an open distributed processing architecture and currently includes ten agent modules: blackboard, dialogue manager, domain model, gesture recogniser, laser system, microphone array, speech recogniser, speech synthesiser, natural language processor, and a distributed Topsy learner. Most of the modules are programmed in C and C++ and are glued together using the DACS communications system. In effect, the blackboard, dialogue manager and DACS form the kernel of CHAMELEON. Modules can communicate with each other and the blackboard which keeps a record of interactions over time via semantic representations in frames. Inputs to CHAMELEON can include synchronised spoken dialogue and images and outputs include synchronised laser pointing and spoken dialogue.

An initial prototype application of CHAMELEON is an “IntelliMedia WorkBench” where a user will be able to ask for information about things (e.g. 2D/3D models, pictures, objects, gadgets, people, or whatever) on a physical table. The current domain is a “Campus Information System” for 2D building plans which provides information about tenants, rooms and routes and can answer questions like “Whose office is this?” and “Show me the route from Paul Mc Kevitt’s office to Paul Dalsgaard’s office.” in real time. CHAMELEON and the IntelliMedia WorkBench are ideal for testing integrated signal and symbol processing of language and vision for the future of SuperinformationhighwayS. Further details are available on http://www.cpk.auc.dk/imm.

Biography

Paul Mc Kevitt is 34 and from Dún Na nGall (Donegal), IRELAND on the Northwest of the EU. He is a Visiting Professor of Intelligent MultiMedia Computing at Aalborg University, Denmark and a British EPSRC (Engineering and Physical Sciences Research Council) Advanced Fellow in the Department of Computer Science at the University of Sheffield, Sheffield, England. The Fellowship, commenced in 1994, and releases him from his Associate Professorship (tenured Lecturership) for 5 years to conduct full-time research on the integration of natural language, speech and vision processing. He is currently pursuing a Master’s degree in Education at the University of Sheffield. He completed his Ph.D. in Computer Science at the University of Exeter, England in 1991. His Master’s degree in Computer Science was obtained from New Mexico State University, New Mexico, US in 1988 and his Bachelor’s degree in Computer Science from University College Dublin (UCD), Dublin, Ireland in 1985. His primary research interests are in Natural Language Processing (NLP) including the processing of pragmatics, beliefs and intentions in dialogue. He is also interested in Philosophy, MultiMedia and the general area of Artificial Intelligence.

Intelligent MultiMedia (IMM) studytrip (Boston/NY ‘99): A truly MultiModal experience

Paul Mc Kevitt
Visiting Professor
CPK, Aalborg University, Denmark
pmck@cpk.auc.dk

While developing an itinerary of sites to visit for this study trip I attempted to obtain as wide a coverage of topics of, and approaches to, Intelligent MultiMedia (IMM) as possible. I wanted to include companies as well as universities and focussed mainly on research laboratories since IMM is a new area and that is where much of it is found today. I also attempted to include work on speech, NLP (Natural Language Processing), and vision/graphics so that the basic input/output elements of IMM could be covered. Hence, that is how I arrived at the following 10 sites: MIT AI Lab, MIT Speech group, MIT Media Lab, Mitre corporation, BBN corporation, Harvard University (AI/NLP group), Lucent Technologies (Dialogue Systems), Rutgers University (CAIP, MultiModal systems), New York University (NLP group), and Columbia University (Graphics and MultiModal group). Coordination with most sites was straightforward although at some sites some researchers were away at other meetings but this cannot be helped when we are visiting all sites in one week. For the larger organisations, notably MIT Media Lab it took some time to establish and unravel the structure of the organisation and who to talk to arrange a visit.

Most sites gave memorable presentations and demonstrations. I was particularly impressed by the work at CAIP, Rutgers and the degree of funding and equipment that Prof. James Flanagan has managed to attract there. Their very detailed scientific and engineering work on MultiModal systems and Human Computer Interaction (HCI) and in particular, microphone arrays, which they have developed and focussed much on over the years was presented well through a tour.

Prof. Mc Kevitt (right) with Prof. Flanagan

The presentation and demonstrations by Victor Zue of speech systems such as Jupiter (weather information) and Pegasus (flight information) at the MIT spoken language systems group were very impressive and also the follow-up meetings with Stephanie Seneff and their students. BBN Corporation and Lucent also presented impressive speech systems. Josh Bers at BBN demonstrated a handheld multimodal device with speech input and they demonstrated an on-line shopping dialogue system. The performance of these systems was impressive although it worried some that people would speak their authorisation password aloud. Lucent had a notable hands-free voice user interface (Woudenberg/Soong) with one fixed microphone where the speaker could move around the room whilst speaking to it. The system worked out the background noise and then suppressed it from what the speaker was saying. However, the system did not have the ability to handle situations where more than one speaker spoke at the same time. Also, impressive was their call centre answering application (Chu-Carroll) which they claimed performed better than people at answering and forwarding/routing calls. Their work on speech synthesis (Olive) handled not only English and German but also languages such as Spanish, French and Russian.

Much of the NLP work we saw focussed on information extraction with Mitre and BBN focussing on Broadcast News (BNN - Broadcast News Navigator) and Columbia on hiring and firing in organisations. Many of these are taking part in the annual MUC and TREC contests. Much of this work was relevant to some of our student projects on smart web browsing and Mobile Intelligent Agent and Hitchhikers guide based on web data. Columbia were conducting information extraction over medical data and then using that for a speech/graphics multimodal generation system. We saw more theoretical NLP work at Harvard on modelling collaborative agents and their beliefs and intentions in various contexts.

With respect to graphics/vision Lucent's vision work on MPEG-4 face animation/talking heads (Petajan) and tennis-player tracking (LucentVision, Carlbom) were state of the art. The tennis-player tracking was demonstrated as being used by a TV sports presenter. Vision was incorporated into multimodal interfaces of the gesture and narrative language group at the MIT Media Lab. Harvard demonstrated videos of lots of their graphics software for enabling optimal diagram layout and other applications.

Applications of IMM technology which were interesting were the robots (e.g. Cog and Kismet) of MIT AI Lab and also the haptic interface tools there. At the AI Lab we saw videos of many medical and other applications of IMM and AI technology. Also, of note was the Intelligent Room project, headed by Michael Coen, where computers see, hear and respond to human stimuli. Two projects of particular interest at the MIT Media Lab were the interactive conversational agent (Rea – Real Estate Agent) of the gesture and narrative language group (Justine Cassell) and the GuideShoes project (Paul Nemirovsky and Glorianna Davenport) of the Interactive Cinema group. The former focuses on conversational agents with whom one could interact through speech and body and hand gestures (either bodysuit or vision). The latter focuses on the fact that aesthetic forms of expression (music, painting, video) can be used for information delivery: GuideShoes, a wearable system used to direct a user towards a specific geographic goal, uses music to navigate in an open space.

To sum up, all the sites we visited presented research and demonstrations of applications which showed speech, NLP and vision processing either as independent or in integrated IMM systems. It was surprising to see that sometimes there was little collaboration between groups at a given institution. For example, the Speech group, AI Lab and Media Lab at MIT do not seem to collaborate much at all and even within the Media Lab there did not seem to be much collaboration between subgroups. It was unclear whether this was solely due to the competitive culture of the US or maybe people and groups were too busy to even think about collaborating. Certainly, the various groups could gain a lot by collaboration, e.g. we thought the emotional robot (Cog and Kismet) work at the MIT AI Lab could gain from talking with the emotional work ongoing in the Affective Computing group at the MIT Media Lab and vice-versa.

It was interesting to see that many of the integrated systems, such as MIT AI Lab's Intelligent Room, MIT Speech Group's Galaxy architecture and MIT Media Lab's Rea, used architectures (blackboard) and knowledge representations (frames) with contents (e.g. intentions) similar to those used in our own CHAMELEON (see Broendsted et al. 1998). They also faced the same key problems: (1) synchronisation of inputs/outputs and their semantics, and (2) the technical platform for integration. The sites also responded very positively to our presentation of CHAMELEON and overview of student projects and were impressed that we had a Master's specialisation in IMM with 20 students per year and also that the students could organise and obtain funding for such a study trip. A number of sites made employment offers, offers of potential student visits and at least one site said it would be interested in applications from the students for Ph.D. scholarships.

If it wasn't for the excellent organisation of the students then I am sure this studytrip would not have been half as successful and I will always remember when Henrik mentioned "...joint account...", Mark Maybury (Mitre) gasping, "Wow, joint account, you guys are organised!" Not only did I fully enjoy the 10 site visits but also the social events which go with such a trip. Visits to the "JFK museum", "Cheers", and Holocaust/Famine memorials in Boston, the "Empire State", "Statue of Liberty/Ellis Island", Breakfast at Tiffany's ("Sbarro", 574 5th Avenue), and 3D video at the Sony IMAX (Lincoln Centre) in New York, and "Hooters" (Boston & NY), are experiences that I can say were for me truly MultiModal.

Paul Mc Kevitt CPK, Aalborg, Denmark April 6th, 1999

Great Expectations – All Fulfilled

A summary by the students

We arrived in Boston with great expectations. We had planned an ambitious schedule that had us visiting two leading-edge research centres every day. Each visit gave us insight into the state of the art of Intelligent MultiMedia and provided inspiration and valuable contacts for our own Masters Theses projects. The wide variety of the visits also allowed us to experience different research environments: universities old and new; and businesses from the size of BBN (100 people in one building) to the size of Lucent (several hundred thousand people located in many huge campuses).

Despite wanting to stay longer on each visit, we kept to our schedule throughout the week and even had some time to be tourists – catching games of ice hockey and basketball, raising a glass at Cheers, shopping in a mall and even importing Intelligent MultiMedia toys (Furby) to Denmark.

One of the most interesting and relevant aspects of the trip was the visit to the Gesture and Narrative Language Group at MIT's Media Lab. Their demonstration of Rea gave inspiration to several of our project groups – showing how it was possible to integrate an autonomous graphical agent into an application, allowing the user to communicate using gestures and natural language. The MIT Spoken Language Systems Group's applications also provided a natural language interface and their GALAXY architecture provided some interesting comparisons to Aalborg University's CHAMELEON. We were also privileged to be able to discuss such technologies with some of the leading names in the field – now we can put faces to the names on the papers we have read!

Another item of note was that all the projects demonstrated to us had some collaborative aspect. Whether the collaboration was between users or between modules in a system, none of the projects involved monolithic, standalone applications. This aspect of communication and networking highlights another topic – that of invasion of privacy. From what we saw at BBN, Lucent and especially MITRE, we can tell that any electronic information can be automatically summarised and analysed. Although all the information extraction we were shown was from publicly available television broadcasts, there is no reason why such techniques could not be applied to telephone calls and closed circuit video surveillance.

On a more positive note, we were very impressed by Lucent's speech recognition system, which not only delivers high accuracy under adverse conditions, such as background noise, but also allows the user to "barge-in" while the system is talking. The "Leg Lab" at MIT's AI Lab was also a high note: we were shown robots that were able to walk, jump and run with one, two and four legs – some of them were even capable of doing summersaults!

In summary, both the experiences we had and the contacts we made will be a tremendous advantage to us in our future careers, whether they be in academia or commerce.

List of students on the trip

Morten L. Andersen	lyk@kmd.dk
Jens Bang	jbang@kom.auc.dk
Pernille Bondesen	bondesen@kom.auc.dk
Jacob Buck	JacobB@dai.ed.ac.uk
Søren Bach Christiansen	sbc@kom.auc.dk
Adam Cohen	adam.cohen@bcs.org.uk
Bo Cordes	cordes@kom.auc.dk
Thomas Dorf	dorf@kom.auc.dk
Jan Krogh	jk@ready.dk
Carsten Brinch Larsen	brinch@kom.auc.dk
Trine Madsen	trim@kom.auc.dk
Sajid Muhammad	sajid@kom.auc.dk
Sergio Ortega	sortega@iies.es
Henrik H. Pedersen	hoff@kom.auc.dk
Søren H.B. Poulsen	shb@kmd.dk
Gael Rosset	gael@stofanet.dk
Lars Skyt	wimp@kom.auc.dk
Susanna Thorvaldsdottir	susanna@kom.auc.dk
Hui Wang	huiwang@kom.auc.dk