Issue No. 35
March 2005
ISSN 0965-5476
Published three times a year

MT News International

Newsletter of the International Association for Machine Translation

In this issue . . .

Of special note ...

- ***MT Summit X Calls for Papers*** *(starts page 3)*
- ***The Tsunami and MT Summit X*** *(starts page 3)*
- ***Latest Compendium now available*** *(page 3)*
- ***Conference Reports*** *(start page 6)*
- ***Translation Automation User Society*** *(page 5)*

Spotlight on the News

EAMT 2005

Budapest, Hungary
30-31 May 2005

The European Association for Machine Translation will hold its 10th annual conference in Budapest, Hungary on 30-31 May, 2005. The special focus for this year's conference is "Practical Applications of Machine Translation".

This meeting is the tenth in a series of regular events that provide a forum for the exchange of ideas concerning all aspects of Machine Translation and computer-aids/tools for translator. For the first time, the EAMT event will be held as a full-scale conference, with greater diversity of topics, a larger number participants and papers, as well as the extension of services provided as part of the conference environment.

The venue of the conference will be the Pázmány Péter Catholic University, Faculty of Information Technology.

As at previous EAMT workshops, there will be contributions from researchers, developers, vendors and users of MT systems and other computerized trans-

(Continued on page 2)

Dr. Nagao receives 2005 Japan Prize

Dr. Makoto Nagao, a founder and former president of the IAMT and AAMT, is a recipient of the 2005 Japan Prize, which he received in recognition of pioneering contributions to natural language processing and intelligent image processing.

The Japan Prize is awarded by the Science and Technology Foundation of Japan for outstanding and original achievements in science and technology that have advanced the frontiers of knowledge and served the cause of peace and prosperity for mankind. All categories of science and technology are considered for the prize, with two fields designated for the prize each year in consideration of developments in science and technology. Dr. Nagao received the award in the field of "Information and Media Technology".

Among the many accomplishments that were cited by the Foundation, Dr. Nagao was recognized for research on machine translation, for his work on the first successful Japanese-to-English

(Continued on page 2)

MT News International

Issue No. 35 March 2005

EDITOR-IN-CHIEF:

Jay Marciano
Tel: +1 (603) 589-3900
Fax: +1 (603) 589-3901
E-mail: mtni@eamt.org

CONSULTING EDITOR:
John Hutchins
E-mail: info@eamt.org

REGIONAL EDITOR, AAMT
Hitoshi Isahara
Fax: +81-774-95-2429
E-mail: isahara@crl.go.jp

REGIONAL EDITOR, AMTA
David Clements
E-mail: dclemen1@san.rr.com

REGIONAL EDITOR, EAMT
Jörg Schütz
Fax: +49 (681) 389-5140
E-mail: joerg@iai.uni-sb.de

EAMT 2005

.

(Continued from page 1)

lation tools. In addition to the focus on practical applications of MT, the conference will also focus on systems that support languages of new EU member countries and languages of those countries that expect to join the EU during the next decade.

Three categories of papers will be presented in parallel sessions:

Oral presentations, in which papers are presented as 20-minute lectures, allowing five to ten minutes for questions and answers; poster presentations, in which papers are presented as posters in the exhibition area near the conference room, with 2 or 3 poster sessions of 90 minutes each; and demonstration sessions, in which papers are presented as live demonstrations in a computer lab close to the conference room and facilities.

For further details and for registration information, visit http://www.eamt.org/eamt2005.

For presenters, final versions of papers should be sent for inclusion in the proceedings before 10 April 2005. A style sheet for accepted submissions will be available soon on the EAMT 2005 website.

The Programme Committee of the conference includes of Chair Bente Maegaard (CST, University of Copenhagen, Denmark), Local Organizing Chair Gábor Prószéky (MorphoLogic & Pázmány Péter Catholic University, Budapest, Hungary), Viggo Hansen (Zacco A/S, Copenhagen, Denmark), Steven Krauwer (Utrecht University, The Netherlands), Harold Somers (University of Manchester, United Kingdom), and Gregor Thurmair (Linguatec, Germany).

Important dates:

Camera-ready papers due 10 April 2005.

Early registration deadline: 10 April 2005

For more information on EAMT 2005, visit http://www.eamt.org/eamt2005.

❑

Japan Prize for Nagao

.

(Continued from page 1)

MT system and an English-to-Japanese MT system, and for his seminal work in example-based machine translation.

Dr. Nagao's research in NLP has also included work in key-word extraction, Japanese morphological analysis, the analysis of long, complex Japanese sentences, and in context analysis. Many of his contributions have been made available to the public as freeware tools and have been used worldwide by researchers in Japanese language processing, thereby significantly expanding the scope of Japanese NLP research.

Dr. Makoto Nagao
2005 Japan Prize
Recipient

He has also worked extensively in the area of image processing, having applied feedback analysis mechanisms to the analysis and recognition of facial photographs, and, more recently, in the application of natural language processing and image processing to digital multimedia libraries.

Dr. Nagao, the 2003 winner of the Association for Computational Linguistics' Lifetime Achievement Award, currently serves as the president of Japan's National Institute of Information and Communications Technology (NICT) and is the honorary chair of the upcoming MT Summit X.

❑

MTNI needs you!

Does your organization have news about MT research, development, or application? Do you have ideas for interesting articles? Would you like to report on a conference you are attending? Contact any of the editors. We'd love to hear from you!

MT Summit X

12-16 September, 2005
Phuket, Thailand

The tenth Machine Translation Summit, organized by the Asia-Pacific Association for Machine Translation (AAMT), will be held at Phuket Island, Thailand, in September 2005.

MT Summit X will bring together researchers, developers and users in the area of machine translation and other tools for multilingual processing. The frontier research and demonstration from leading academic institutes, business and government sectors will show the latest results. In the main conference, MT Summit X will provide various features including research papers, reports on users' experiences, discussions of policy issues, invited talks and panel discussions. Surrounding the main conference, the expanded program will cover the sessions of tutorial, workshop and exhibition. These wide coverage-features in the machine translation arena will facilitate participants from the sectors of academy, research, business and government to get together to explore the machine translation technology for conquering the problems in language barriers and digital divide.

AAMT cordially invites you all to MT Summit X, to contribute and witness the efforts in machine translation technology, in any aspect of machine translation and tools for translation support – researchers, developers, providers, users, and watchers.

The organizing committee of the conference includes Conference Chair Junichi Tsujii of the University of Tokyo, Program Chair Hitoshi Isahara of the National Institute of Information, and Local Organizing Chair Virach Sornlertlamvanich, of the Thai Computational Linguistics Laboratory.

Call for Papers

MT Summit X seeks original papers in all aspects of MT. Topics of interests include, but are not limited to:

(Continued on page 8)

Tsunami and MT Summit

December catastrophe adds purpose to choice of venue

Message from the Conference Chair

As you are aware, Thailand was hit very hard by the recent tsunami, prompting the MT Summit X Steering Committee to carefully reconsider the venue. However, all the news from the local organizing committee and the conference venue has been very positive about having the Summit as planned. The local organizing committee and the steering committee have received many supportive messages from members of AMTA, EAMT and AAMT. Many suggested that the Summit in Phuket would boost the morale of the local people and thus contribute to local recovery efforts. The suggestion is fully confirmed by the local organizing committee.

The steering committee, therefore, has decided to hold with our decision of having the MT Summit X in Phuket. We are considering of ways of showing our sympathy and support to the local people. These plans will be publicized as they become final.

We look forward to seeing you all at the MT Summit X in Phuket.

Junichi Tsujii
President, AAMT
Conference Chair, MT Summit X

❑

Message from Local Organizing Chair

On December 26, 2004, a massive number of lives were destroyed by the tsunami. Six provinces in Southern Thailand facing the Andaman Sea, namely Phuket, Phang-nga, Ranong, Krabi, Trang and Satul, were hit by the disaster. Thousands of people lost their families, houses, schools and businesses. Phuket Island was reported to have had the third largest amount of damage, with about 2,000 missing, dead and injured. We have all felt emotions stirred by the images of those affected by the disaster.

Despite the large amount of damage, the six provinces are gradually recovering by the intensive assistance from the cooperation among national and international organizations.

The faster the Phuket Island economy is restored, the more quickly the other areas will begin to recover. From the perspective of re-establishing the resort town, the Thai government is now working hard to bring back the Phuket attractions as well as establishing an appropriate warning system to secure the area.

The gathering for MT Summit X in

(Continued on page 8)

Latest Compendium Now Available

The 10th edition of the *Compendium of Translation Software*, compiled by John Hutchins with assistance from Walter Hartmann and Etsuo Ito, is now available in PDF format on the EAMT website (www.eamt.org.compendium.html). The *Compendium* is free of charge to members of the IAMT, while non-members are charged a nominal fee.

In a recent interview with MTNI, Dr. Hutchins noted that approximately half of the entries in the 127-page document required revision for the new edition, which was completed in February 2005. Hutchins estimates that he spends at least one month updating and validating information for each edition. The most difficult challenge in the editorial cycle is determining what actually is new, as there is no formal system of alert in the industry. Adding to the difficulty, ironically, is the language barrier: even with the use of online translation services, it is often difficult and time-consuming to validate basic facts about many systems , such as system requirements, features, and pricing.

Developers of commercial machine-translation systems and computer-aided translation tools are encouraged to send product information to Dr. Hutchins. For contact information, see page 2 of the Compendium or visit the EAMT website.

❑

Industry News

New Versions from Language Engineering Co.

[Adapted from press releases]

Language Engineering Company, LLC (LEC), based in Belmont, Massachusetts, USA, recently announced the release of several new machine-translation products and services, encompassing desktop products, online services, and imbedded applications.

In the area of desktop products, LEC has new versions of two products, Power Translator 9, for European languages, and Translate 2005, for Arabic, Persian and Turkish. Both products feature a control center for launching and configuring the specialized applications for translation of text, files, email, instant messages, and web pages. A dictionary utility is included in the Professional version. Among the other features are add-ins for Microsoft Word, Excel, Power Point and Internet Explorer. Also new in both products is ClipTrans, an application that translates text that is selected and copied to the system clipboard. Both products are available in a variety of packages at differing price points.

Each license for Power Translator 9 and Translate 2005 also includes an introductory subscription to LEC's online service, Translate DotNet, which has also recently undergone revision. Version 3 of the Translate DotNet applications are identical in function to the desktop versions, but instead of using local instances of the translation engines, they use the latest versions of each translation engine on LEC's servers. Universal subscriptions allow translation of in all language directions supported by LEC: Arabic, Chinese (simplified and traditional), Dutch, English, French, German, Hebrew, Italian, Japanese, Korean, Persian, Polish, Russian, Spanish, Turkish, and Ukrainian. Subscriptions for individual

(Continued on page 5)

Language Weaver releases Version 3.0

[Adapted from press release]

Language Weaver, Inc., has announced the availability of an upgrade to its statistical machine translation software. Version 3.0 offers improvements to speed, accuracy, ease-of-use, and translation quality. Most notably, translation throughput speed is improved 5 to 10 times over previous versions.

Other improvements include source encoding auto-detection and an improved ability to take advantage of multi-processor systems and Hyper-Threading

Bidirectional language pairs available include: Arabic/English, Chinese/English, French/English, and Spanish/English; unidirectional languages include Somali to English and Hindi to English. The software is available as server and standalone versions.

For more information, visit http://www.languageweaver.com. ❑

Tweddle chooses post-edited MT process from SDL

[Adapted from press release]

SDL International has announced that Tweddle Litho has selected SDL as its partner for translation technology and services. Tweddle will use the SDL Knowledge-based Translation System™ to achieve faster and more cost-effective translations.

Tweddle Litho, a market leader in publishing for the automotive industry, faces continued demands from its customers to reduce costs and turn-around times. The SDL Knowledge-based Translation System tightly integrates with Tweddle's content authoring, management and publishing processes to ensure maximum efficiency across the complete end-to-end lifecycle of linguistically complex automotive documentation.

The partnership allows the companies to manage the translation of over 100 million words of automotive documentation into multiple languages concurrently, and provides automotive companies with a complete solution for their localization needs.

The SDL Knowledge-based Translation System (KbTS) tightly integrates translation memory with terminology extraction and machine translation to automate the majority of the translation process. With humans to post-edit the material, high-quality translations are achievable faster and at lower total cost than from a traditional TM-based process. KbTS delivers up to five times the translation throughput of traditional human translation.

"The SDL Knowledge-based Translation System gives us an increase in translation productivity of more than 30%, a level unmatched by anyone else," said Andy Tweddle, President and CEO of Tweddle Litho. "SDL's support for industry standards such as SGML, XML and TMX has provided easy migration from our legacy technologies and services together with easy integration into our existing authoring, content management and publishing systems. With this strategic partnership with SDL, we have extended our competitive advantage and anticipate winning new business from our customers."

Tweddle Litho Company is a leading provider of integrated information solutions for automotive OEMs. Tweddle's service offering encompasses content development, graphic design, database

(Continued on page 5)

SYSTRAN Premium Professional 5.0 Wins Award

[Adapted from press release]

SYSTRAN S.A. has been awarded the European Information Society Technologies (IST) Prize in recognition of its SYSTRAN Professional Premium 5.0.

The European IST Prize is the most distinguished award for innovative European IST products. It recognizes teams that excel in generating and converting R&D and novel ideas into marketable products. Selection criteria for the Prize include technical excellence, innovative content, potential market impact, capacity to generate employment, contribution to the acceptance and understanding of IST in society, and likely social impact. The European IST Prize is organized by Euro-CASE with the sponsorship and support of the European Commission within the framework of the IST Program.

SYSTRAN Professional Premium 5.0 is available in 36 language combinations and contains integrated tools that help improve translation quality and ensure efficient management of translation projects, including specialized dictionaries, a dictionary management tool, a language normalization dictionary, integrated translation memory with TMX support, and a translator's workbench for post-editing and ongoing quality analysis. With SYSTRAN Professional Premium, users can translate files in Microsoft Word, Outlook email, PowerPoint, Excel, Web pages in Internet Explorer, and PDF documents directly from their desktop.

For more information, visit http://www.systransoft.com. ❑

LEC

(Continued from page 4)

languages are available for as little as US$5 per month, depending upon the language. Premium subscriptions add subject dictionaries to the general dictionary for translation of business, computer, legal, and medical text, as well as other technical subjects for selected languages.

LEC has also recently released the Translate DotNet 2005 SDK, a SOAP-based client-server machine translation system that is used by OEMs, system integrators, and corporate developers for integration with their own application with maximum control over the translation process. Translate DotNet communicates with a translation server, either one supplied by LEC via the internet or a customer's server on their own LAN. Client programs package text to translate in a well-formed XML message, and transmit that message using the HTTP web protocol to an LEC Translate DotNet server. The server processes the request, translates the text, and returns a well-formed result. The LEC SDK includes translation engines and subject dictionaries for 42 language pairs: English to/from Arabic, Chinese (simplified and traditional), Dutch, French, German, Hebrew, Italian, Japanese, Korean, Persian, Portuguese, Russian, Spanish, Turkish, Ukrainian; French to/from German, Italian, Spanish; Italian to/from German, Spanish.

For more information visit http://www.lec.com. ❑

Tweddle

(Continued from page 4)

& translation management, composition, production of print and electronic media, through fulfillment and distribution.

SDL International is a provider of translation services and technology solutions, including translation memory, MT, automated terminology extraction, and a full range of localization services. For more information, visit http://www.sdl.com.

Editorial disclosure: MTNI editor Jay Marciano is Director of MT Development at SDL International ❑

Translation Automation User Group Formed

TAUS (Translation Automation User Society) is an association of corporate and institutional users of translation, authoring and globalization technologies that grew out of a roundtable meeting at the 2004 Localization World Conference in San Francisco. The aim of TAUS, an initiative of Localization World, Ltd., is to be the means through which companies and institutions that have invested in translation technologies can share their ideas, strategies, and experiences with applying these technologies in their ongoing effort to reduce the expense of translation.

While MT is of major interest, it is considered in the broader context of translation technologies and processes.

Members of TAUS have access to a web portal through which they can connect individually with other members, join discussion groups and find pertinent information. A visit to the portal begins with a click on a link to one of eight areas of interest, Machine Translation, Translation Memory & Software Localization, Controlled Authoring, Terminology Management, Workflow, Standards, Service Vendors, and Localization Applications. Each link, according to the TAUS website, branches out into "several user groups of products, standards, applications or vendors."

TAUS also sponsors round table meetings both independently of and in conjunction with Localization World and MT conferences.

The organization is directed by Jaap van der Meer, formerly CEO of Alpnet and currently a major stakeholder in Cross Language, a Belgium-based translation automation service provider.

Membership, with annual fees ranging from €2,500 to €10,000, is open to corporate and institutional users – but not developers or vendors – of authoring, translation, and globalization technology and services.

For more information, visit http://www.translationautomation.com/. ❑

Conference Reports

Two perspectives on the AMTA 2004 Conference

From an AMTA veteran

By Roderick Holland

The choice of venue for AMTA 2004—Georgetown University—carried with it some powerful historical resonances. The central resonance was with Georgetown's role in the earliest days of machine translation. The keynote address by W. John Hutchins and the panel discussion of the "Georgetown Pioneers" evoked the heady promise and subtle difficulties of that period. Many of us heard in these reminiscences an account of a familiar existential position: the urgent need to do something truly difficult, with the understanding that success, however pleasing, can only be partial.

Another resonance was the role of government in driving MT research and development. Larry Brisky's keynote address, "Foreign Language Tools and Machine Translation—Critical Enablers in Meeting the Intelligence Challenge," framed this role for the current period. The Georgetown location allowed many U.S. government representatives with real-world needs to attend. The demonstrations and discussions that ensued at the Research and Deployment Showcase formed an unusually effective mixing bowl for the research, commercial, and government stakeholders in MT technology, with consequential results for many. As one of the participants (demonstrating MITRE's TrIM and Clipper embedded MT prototypes), I found it a rare opportunity for engagement. Kudos to Jennifer DeCamp for organizing the Showcase.

One of the lessons that the Georgetown-IBM Experiment taught us was the need for reliable evaluation. Presentations by Elliott Macklovich, Alon Lavie, Debbie Elliott, and Flo Reeder in the Evaluation session of the Research Track demonstrated that this is an area where real methodological progress continues to be made.

Many of us are coming to grips with processing the Arabic language, for translation and other purposes, and Nizar Habash's tutorial, "Introduction to Arabic Natural Language Processing" was a useful gateway for aspiring practitioners. The strong presence of most of the leading vendors of Arabic-English MT systems was timely and helpful, and their presentations and demonstrations had a competitive spirit that added a sense of excitement and (at times) fun to the conference.

Finally, I came away deeply impressed by the reality of a vertically integrated MT enterprise revealed at Julia Aymerich's tutorial, "MT Practice at PAHO". This on-site session disclosed an organization engaged in the daily development, use, and refinement of MT as a central enabler to their public health mission, with results that are both impressive and instructive.

Roderick Holland is Chief Architect, Multimedia and Collaborative Services, MITRE Corporation

❑

From a first-time attendee

By Laurel Wagers

I'm not a translator, a software engineer or a computational linguist, but in my six years at *MultiLingual Computing*, I have learned a fair amount about language technology and their use. My experience at the AMTA conference reflects my knowledge about the generalities of the discipline and the fact that I do not necessarily understand the specifics.

So, it's easy for me to slip into a "gee-whiz" mode when I see some of this research in action. I was particularly interested in and amazed by integrated tools — for example, the suites that capture, translate and display text from audio and video. I can see their potential benefits for researchers as well as for intelligence gathering. The combination of a low-tech "Iraqi Visual Language Survival Guide" card with MT in the Phraselator made a great deal of sense. And the "document exploitation" systems (ArborScript and others) were impressive to someone who has spent hours in Photoshop trying to improve contrast without losing picture information. I think that TrIM — translated instant messaging — has great potential in the everyday world. It was exciting to see how these applications are being applied to intelligence-gathering and military problems and how they may impact day-to-day computing and translation.

One of the best parts of learning about these applications in the conference setting is the opportunity to meet the people who have developed and refined them with keen intelligence, advanced technical knowledge and creative passion.

I attended the keynote addresses and two panels: the Georgetown Pioneers, some of whom were students in 1954, inventing the rules and processes of MT, and the presentations by students currently working their way into the now-established field. The keynotes reminded me of an unfortunate aspect of having speakers from government agencies such as DARPA and the CIA: agency-approved public statements can seem general and frustratingly uninformative to those of us without security clearance or insider knowledge.

Throughout the three days, organizers Laurie Gerber and Jen Doyon seemed to be everywhere and kept the event running smoothly. As an exhibitor, I found the gallery location to be visible and accessible. Moving the exhibitors on Thursday into the salon area along with the demonstrations created a energetic and "buzzing" environment. Unfortunately, the exhibitor tables were left in the gallery on Friday, possibly leading some attendees to think that the exhibitors had already left. Exhibiting was otherwise a smooth experience, and I came away with commitments from several researchers to write articles on various aspects of MT for the magazine in the next year.

(Continued on page 11)

MT Featured at Annual Meeting of AAAS

By Laurie Gerber

The annual meeting of the American Association of the Advancement of Science (AAAS), held in Washington DC from 17-21 February 2005, is a mind-bogglingly broad assembly of reports from all corners of the world of scientific endeavor. This time, machine translation got some attention in this forum, in a morning press conference and a full-afternoon session with 8 presenters.

The 3-hour MT session was held on Sunday, 20 February 2005, at the Marriott Wardman Park Hotel. Each of 8 presenters described projects and products they are involved in. The presenters represented end user organizations (1), a systems integrator (1), companies that market companion technologies for machine translation (2), and machine translation software developers (4). The afternoon panel was one of 14 parallel full-afternoon scientific symposia. In addition, there were at least 10 other types of sessions/activities, and an exhibition of companies, technologies, products, employers, countries, government agencies, etc.

The morning press conference featured brief commentary on recent advances in machine translation from Kevin Knight (Language Weaver/USC Information Sciences Institute), Melissa Holland (U.S. Army Research Lab), Benson Margulies (Basis Technology) and Mohammad Shihadah (Applications Technology, Inc.). This session caught the attention of quite a few journalists, and resulted in an article in the New Scientist (UK) the following week. But it was in the afternoon where presenters went beyond sound bites and talked about developing MT and putting it to work. The remainder of this report summarizes those presentations.

Mohammad Shihadah, co-founder of Applications Technology (Apptek), talked about current offerings and advances in rule-based machine translation. Including a broadening array of languages (particularly middle-eastern languages such as Farsi and Pashto, in addition to Apptek's core Arabic offering), and companion tools such as integrated translation memory and personal name detection.

Steve Richardson, of Microsoft Research, reported on Microsoft's highly successful trial using their own example-based machine translation software to make knowledge-base articles available in Spanish, French, German and Japanese. Utilizing machine translation has meant that the full knowledge base can be made available to more Microsoft users. Historically there have been hundreds of millions of words of user documentation in English that it has never been feasible to translate. Interestingly, user satisfaction surveys show that users of the MT output are generally as satisfied as or more satisfied than users reading the original English documentation.

Kevin Knight, of Language Weaver, Inc., and University of Southern California, Information Sciences Institute, reported on recent successes in using statistical methods to raise the achievable quality of automated translation. Knight showed examples from early research systems based on statistical methods and compared them to current results, which showed noticeable improvements. Other aspects of the translation landscape that are promising for statistical methods are the increasing accumulation of translation memories in the commercial world, and the growing availability of bilingual text on the internet, which can be used as training material.

Daniel Marcu, also of Language Weaver, Inc. and University of Southern California, Information Sciences Institute, provided an analysis of the strengths of rule-based/linguistic, statistical, and example-based translation, and described how the best of these can be combined into single hybridized systems.

Benson Margulies of Basis Technology described a number of preprocessing technologies offered by Basis, which provide technological and productivity advantages when combined with machine translation. Margulies emphasized that merely translating material doesn't solve problems if the goal is to find the few important nuggets of information in a large collection. Filtering and prioritization need to be done prior to translation, in the source language, to identify what is important enough to merit translation and scanning.

Clare Voss of Army Research Lab described the long path of development and innovation that have helped the FALCon (portable OCR+MT) system become a widely used tool. FALCon is used by soldiers in the field to assess the intelligence value of paper documents. First assembled in 1996 from (largely) off-the-shelf components, gradual advances in both OCR technology and machine translation have allowed FALCon to evolve and expand the kinds of information that can be exploited – including OCR for signs captured via digital cameras.

Mark Turner of CACI described a software system called the "HighView Document Exploitation Suite". This system, developed over many years by the team now at CACI, enables flexible creation of document processing workflows that can include OCR, MT, and other language analysis tools, as well as record-keeping and archiving capabilities to uniquely identify and track original document images together with transcriptions, translations, and any analytic byproducts.

Bill McClellan of systems integrator Booz-Allen Hamilton described his experience working with military and intelligence requirements for exploitation of information originally in Arabic. He emphasized that in the field, the biggest problem is figuring out what is important from the high volumes of uncategorized foreign language information that arrive in many forms.

Laurie Gerber, AMTA President, is Director of Business Development for Language Weaver

❑

MT Summit X

(Continued from page 3)

- MT for the Web
- Practical MT systems (MT for professionals, MT for eCommerce, MT for localization, etc.)
- Translation aids (translation memory, terminology databases, etc.)
- Translation environments (workflow, support tools, conversion tools for lexica etc.)
- Methodologies for MT
- Human factors in MT user interfaces
- Speech and dialogue translation
- Natural language analysis and generation techniques for MT
- Dictionaries and lexicons for MT systems
- Text and speech corpora for MT and knowledge extraction from corpora
- MT evaluation techniques and evaluation results
- Standards in text and lexicon encoding for MT
- Cross-lingual information retrieval
- MT and related technologies (information retrieval, text categorization, text summarization, information extraction, etc.)

Submissions are invited for three categories of papers:

Research papers: reports of significant research results in any aspect of machine translation and related areas. Such reports should include a substantial evaluative component.

Poster presentations (with or without demonstration): presentations that are best suited for the more interactive form of poster presentation, including reports on the design, implementation, operation and evaluation of operational and prototype systems.

User studies: reports on users' experiences with applying MT.

All types of papers should include a cover page with the following information: paper title; paper type (Research, Poster presentation, or User study); name, affiliation, address, and e-mail address of all contributing authors; and a 100 word abstract, with up to 5 keywords.

Important Dates

Submission deadline: 15 April 2005

Acceptance notification: 31 May

Camera-ready copy: 29 July 2005

Call for Exhibitions

MT Summit X will feature an exhibition area for showcasing commercial products and near-to-market prototypes throughout the conference. Exhibits will not be restricted to machine translation systems and support tools, but include a variety of language technology enabled applications. If you would like to exhibit, please contact Kunio Matsui, at matsui.kunio@jp.fujitsu.com, by 15 April 2005.

Venue

Phuket, "The Pearl of Thailand", is the largest island of Thailand. It is connected by bridges to southern Thailand and surrounded by 32 smaller islands with a total area of 570 square kilometers. Its beaches offer some of the most spectacular scenery in Thailand, from brilliant white sand beaches and coral islands to sheer limestone cliffs and forested inlets.

The conference venue, the Hilton Phuket Arcadia Resort and Spa, is located on Karon beach, 45 minutes from Phuket Airport. Set amid 75 acres of tropical landscaped gardens, it is the largest five star resort on the island and sustained very little damage in the recent tsunami. Each of the resort's 695 rooms are decorated in modern Thai style, accentuating all of the comforts of home, while including all of the luxury of a five star resort. Each of the rooms has a balcony with a view of either Karon Bay or the beautifully landscaped gardens.

For further information on MT Summit X, visit http://www.aamt.info.

❑

Message from Local Organizer

(Continued from page 3)

Phuket is greatly welcomed not only to support the local economy but also to explore the technological possibility of utilizing of MT for communication in emergencies. In the recent tsunami disaster, information and communications technology played an important role in the urgent needs of database construction to support the management and searching of the disaster information via the Internet. The collaborative work using web technology has proven its strength in responding to the sudden growth of the flow of information.

Under the well-established infrastructure of the Internet, the coming MT Summit X will be an appropriate place for seriously discussing the MT technology for emergency communication. Language is still a great barrier. We lack off-the-shelf technology to provide the just-in-time service. This could be a challenging topic for contemporary MT research.

On behalf of the local organizing committee, I would like to express my warm welcome to the Summit participants to Phuket. One of your contributions will be in part to restore normal lives and joyfulness to the local people. Once again, I would like to call for a new challenging research of MT technology for emergency communication that includes cross language, cross culture or language independent communication. In addition, Hilton Phuket Arcadia Resort and Spa, the selected location for the MT Summit X venue, is fully operating without any significant damage. It is currently hosting many international meetings especially on topics related to the tsunami.

Virach Sornlertlamvanich

Local Organizing Chair, MT Summit X

❑

MT Summit 10 Workshops — Calls for Papers

Workshop on Example-Based Machine Translation

Following the successful EBMT workshop in 2001 at the MT Summit VIII in Santiago de Compostela, Spain, this is the second workshop of its kind. EBMT approaches have evolved since 2001, with some even reaching commercial maturity. Furthermore, a number of new and innovative approaches have been seen. Their common characteristics are to exploit and integrate a number of knowledge resources, such as linguistics and statistics (or symbolic and numerical techniques) for integration into a single framework. In this way, rule-based morphological, syntactic and/or semantic information is combined with knowledge extracted from bi- or monolingual texts which is then re-used in the translation process. We seek to assemble approaches to machine translation which go beyond the purely statistical and/or rule-based paradigms. We welcome visionary and technical descriptions, reports of empirical research as well as feasibility studies and system demonstrations. We would welcome contributions on any of the following topics:

- descriptions of 'pure' EBMT systems
- knowledge resources used in EBMT
- representation of numeric and symbolic knowledge
- descriptions of 'hybrid' systems (integration of EBMT with rule-based methods or statistical methods)
- (semi-)automatic preparation of existing mono/bi/multilingual corpora for EBMT
- extraction of bi/multilingual texts from the web
- preparation of treebanks for EBMT
- bi/multilingual alignment/ bracketing/parsing
- inference of mono/bi/multilingual grammars
- inference of bi/multilingual transfer rules
- evaluation of EBMT results and/or comparison with other MT systems
- considerations on domain-(in) dependence of EBMT systems
- computational and/or system complexity of EBMT systems

In addition, we plan to have a panel session focusing on future perspectives for EBMT. The idea is that this session complements the envisaged panel in the main program which aims to focus on the continuing convergence of EBMT and SMT.

This workshop is being organized by co-chairs Michael Carl (IAI, Saarbruecken) and Andy Way (School of Computing, Dublin City University).

(Continued on page 10)

Workshop on Semantic Web Technologies for Machine Translation

With the goal of implementing a semantic structure behind the content of the World Wide Web, Semantic Web activities have recently attracted a significant and specialized research community of computer scientists, computational linguists, logicians, and ontology specialists, among others.

Until recently, research has been directed at applying techniques from human language technology to add meaning to the web-based data, thereby enabling automatic processing. The use of new Semantic Web technologies for improving natural-language applications, however, has been neglected. The development of ontologies for the Semantic Web, their search mechanisms, and the standard formal (e.g. RDF) annotation of large pieces of data on the web, are of high value for monolingual and multilingual natural language (web)-applications The current workshop focuses on this topic, more exactly on the implications of such semantic web technologies on machine translation, which is a representative sub-field of natural language processing. It is well-known that multilinguality is one of the main challenges of Semantic Web. The annotation mechanisms and the development of ontologies and search procedures aim at retrieving relevant information independently of the language in which it was produced. On the other hand, Semantic Web activities will have major impact on natural language applications based on training on large pieces of corpora

Example-based machine translation is a relevant example: Up to now the training has been done on parallel aligned corpora, in the best case, additionally annotated with syntactic information. Large, reliable parallel corpora are available only for a few language pairs and domains. In the absence of such corpora, the Web is the best source for parallel aligned corpora. Aligned via RDF(S) annotations, the Web can be exploited as a multilingual corpus. Moreover, this annotation will provide the semantic information attached to the respective texts. This strategy can have significant implications on example-based machine translation.

Knowledge-based machine translation is another technique which can benefit from Semantic Web activities. Until now KBMT systems were based mainly on the development of domain-dependent ontologies and on mapping the source language onto the target language via these ontologies. It was proved that KBMT can be very successful when applied to restricted domains, but it encounters severe problems with translations of general texts. Semantic Web activities can provide a large amount of ontologies in various domains and bridges between these ontologies. In this new context, KBMT could become a powerful mechanism

(Continued on page 10)

Workshop on Open-Source Machine Translation

Machine translation has become a key technology in our globalized society; as a result, machine translation software is available for major language pairs and for major computer platforms and web-based applications. The recent years have witnessed a boom of open-source software, such as the Linux operating system, the Mozilla web browser, web servers such as Apache, and the full-fledged office suite, OpenOffice.org. However, almost all "real life" machine translation software, even if available for use at no cost, is "closed" instead of open. This is especially surprising if one considers the large number of publicly-funded groups working in machine translation.

Open-source machine translation, however, would have distinct advantages. If it were freely available, as most open-source software is, more users would have access to this technology, and, more importantly, institutions or businesses adopting an open-source MT system would be able to customize the system to their needs in more flexibly, as by developing new linguistic data (vocabularies, rules, corpora), integrating it with other packages, etc.

But MT software is special in that it relies upon the availability of extensive linguistic resources; for an open-source MT architecture to be successful, clearly defined and documented standards for representing linguistic data are absolutely necessary. Data standardization would lead to interoperability and interchange, which would in turn be very beneficial to the creation of new machine translation systems. Proprietary data could also be converted into these formats to be used in conjunction with open-source architectures, leading to hybrid systems.

The existence of an open-source machine translation architecture would also be especially valuable in the creation of systems that support translation to or from small or neglected languages. Such language pairs are usually not targeted by commercial programs but would fulfill the goals of administrations and non-government organizations dealing with these languages, and even contribute to their promotion or revival.

Open-source software is associated to a change in the business model. In the case of machine translation, it would result in a shift from license-based or charge-per-word models to a service model in which enterprises would offer users a variety of services: consulting, customization, linguistic data development, etc.

The workshop will take place on 16 September 2005, after the main conference.

(Continued on page 11)

EBMT Workshop

(Continued from page 9)

There will be a common publication format for all workshops in line with the main conference proceedings. However, the organizers have approached relevant publishers in anticipation of publishing the workshop proceedings more widely, as done in [Carl and Way, 2003].

Important dates

Paper Submission: 31 May 2005

Notification of acceptance: 24 June 2005

Camera Ready Papers: 22 July 2005

Workshop: 16 September 2005

Attendance Fee

The workshop fee is expected to be approximately $70.

For more information, visit: http://www.computing.dcu.ie/~away/EBMT2.html

❑

Semantic Web Technologies Workshop

(Continued from page 9)

for on-line machine translation.

The goal of the workshop is twofold: to discuss the implications of semantic web-technologies for machine translation, namely on example-based and knowledge-based machine translation, and to contrast the two main technologies of Semantic Web, topic maps and RDFs in the machine translation of on-line texts.

We welcome original papers related (but not limited) to following topics

- semantic web annotations for multilingual corpora
- use of semantic web annotations for corpus based MT
- integration of semantic information in EBMT
- use of semantic web ontologies for MT

We also encourage demonstrations of developed tools. Submissions for a demonstration session should include a two-page description of the system architecture and performance as well as technical requirements.

This workshop is being organized by Walther v. Hahn (University of Hamburg), Vladislav Kubon (Charles University Prague), and Cristina Vertan (University of Hamburg).

Important dates:

Paper Submission: 13 May 2005

Notification of acceptance: 24 June 2005

Camera Ready Papers: 22 July 2005

For submission guidelines, please follow the guidelines of the main conference. Workshop papers will be included into the MT SUMMIT CD. Papers should be submitted electronically in PDF format to Cristina Vertan (cri@nats.informatik.uni-hamburg.de).

❑

Other Conferences and Events

Practical Applications in Language and Computers (PALC 2005)
Date: 07-09 April 2005
Place: Lodz, Poland
Link: http://palc.ia.uni.lodz.pl/

2nd Language & Technology Conference
Date: 21-23 April 2005
Place: Poznań, Poland
Link: http://www.ltc.amu.edu.pl/

2nd ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications (Prep05)
Date: 19-21 April 2005
Place: Colchester, Essex, United Kingdom
Link: http://privatewww.essex.ac.uk/~avill/Prep05.html

3rd International Workshop on Generative Approaches to the Lexicon
Date: 19-21 May 2005
Place: University of Geneva, Geneva, Switzerland
Link: http://www.issco.unige.ch/gl2005.html

ACL 2005 (43rd annual meeting)
Date: 25-30 June 2005
Place: Ann Arbor, Michigan, USA
Link: http://www.aclweb.org/acl2005/index.php?conference

Dial-A-Corpus – Compiling Corpora from the Web
Date: 29 June 2005 – 02 July 2005
Place: Siena, Tuscany, Italy
Link: http://www.twc.it/

Workshop on Computational Modeling of Lexical Acquisition (CMLA)
Date: 25-28 July 2005
Place: Split, Croatia
Link: http://www.ohz.unist.hr/cpala/

10th European Workshop on Natural Language Generation (ENLG-05)
Date: 08-10 August 2005
Place: Aberdeen, Scotland
Link: http://www.ling.helsinki.fi/~gwilcock/ENLG-05/

Open Source MT Workshop

(Continued from page 10)
Workshop papers will be published as part of the main conference proceedings.

The workshop seeks original papers in all aspects of open-source MT, including but not limited to:

- open-source MT architectures ;
- projects and currently available software: complete engines, modules;
- standards for encoding linguistic data for open-source translation technology;
- licensing issues for MT engines and linguistic data;
- open-source MT as an opportunity for small or neglected languages;
- open-source translation memories ;
- advantages and disadvantages of open-source language technologies;
- the role of public agencies in promoting the use and development of open-source MT;
- the future of open-source MT.

Papers should not be longer than 3,000 words. Presentations may have a substantial demonstrative component.

Additional publication guidelines will be available at the start of the submission period on the workshop web site.

Important Dates

Submission deadline: 13 May 2005

Notification of acceptance: 24 June 2005

Camera-ready copy: 15 July 2005

The workshop fee is expected to be approximately US $70.

The workshop is being organized by Chair Mikel L. Forcada (Universitat d'Alacant, Spain), Iñaki Alegria and Kepa Sarasola (both of Euskal Herriko Unibertsitatea, Donostia, Spain), Xavier Gómez Guinovart (Universidade de Vigo, Spain), Lluís Padró (Universitat Politècnica de Catalunya, Barcelona, Spain), and Juan Antonio Pérez-Ortiz and Antonio M. Corbí-Bellot (also of Universitat d'Alacant, Spain).

For more information, visit: http://www.torsimany.ua.es/OSMaTran/

❑

AMTA2004 Perspective

(Continued from page 6)
I also attended a lively demonstration session in which AppTek, Language Weaver and Sakhr representatives compared their products and the quality of their several MT engines as applied to front-page news in Arabic.

One regret: I didn't turn in my "scavenger hunt" answers — I'll know better next time. Those CIA coasters would have been a great souvenir.

Laurel Wagers is the Managing Editor of MultiLingual Computing and Technology

❑

MT News International

Subscription Order Form
for non-members of IAMT Associations

*Subscription to **MTNI** is a benefit of membership in any of the three regional IAMT Associations. Non-members may also subscribe. This form should be sent to the appropriate region together with a remittance in the currency specified. The fee covers a one-year airmail subscription (individual or institutional) for three issues, starting in the spring of the current year.*

I/we wish to receive a one-year subscription to MT News International:

Name: _____

Organization: _____

Address: _____

City: _____State: _____ Postal code: _____

E-mail: _____ Fax: _____

For individuals or institutions located in the Asia-Pacific region, please return this form with payment of ¥4,000 (bank draft or international money order) to:

Association for Machine Translation in the Americas
PMB 300
1201 Pennsylvania Avenue, N.W. , Suite 300
Washington, DC 2004 USA

For individuals or institutions located in Europe, the Middle East, or Africa, please return this form with payment of Sw.fr. 70.00 (check or money order) to:

Association for Machine Translation in the Americas (AMTA)
PMB 300
1201 Pennsylvania Avenue, N.W. , Suite 300
Washington, DC 2004 USA

For individuals or institutions located in North, South, or Central America, please return this form with payment of US$ 75.00(check, money order, or credit card) to:

Association for Machine Translation in the Americas (AMTA)
PMB 300
1201 Pennsylvania Avenue, N.W. , Suite 300
Washington, DC 2004 USA

Publications Order Form

Please return this form with payment or credit card information, to:
International Association for Machine Translation (IAMT c/o AMTA)
PMB 300, 1201 Pennsylvania Avenue, N.W. , Suite 300
Washington, DC 2004 USA

Please send the items marked at the right to:

Name: _____

Organization: _____

Address: _____

City: _____State: _____ Post al code: _____

E-mail: _____ Fax: _____

	Price (in U.S. dollars)[1]	
Title	**Member[2]**	**Non-member**
❒ ***Compendium of Translation Software***		
(on-line version)	FREE	$20.00
❒ ***Proceedings of MT Summit VI***	$40.00	$60.00
❒ ***Proceedings of AMTA-96***	$40.00	$60.00
❒ ***Proceedings of AMTA-94***	$40.00	$60.00
❒ ***Proceedings of Workshop on MT Evaluation*** (1992)	$55.00	$55.00

[1] Prices include shipping and handling.

[2] Member of AAMT, AMTA, EAMT.

The proceedings of AMTA-98 and AMTA-2000 appeared as #1529 and #1934 in the Springer series Lecture Notes in Artificial Intelligence. To order, contact the publisher at www.springer.de.

Method of Payment

❒ Check or M.O. ❒ Visa ❒ MasterCard ❒ American Express

Card number _____ Exp. Date: ___/___

Asia-Pacific Association for Machine Translation

CORPORATE / INDIVIDUAL APPLICATION FORM

Please fill out the appropriate form in both your native language (NAT) and English (ENG) and send it to the address below.

CORPORATE

Corporate name: Seal:

NAT: _____

ENG: _____

Capital: _____ Date established : _____ # employees: ____

Name of company president:: Seal:

NAT: _____

ENG: _____

Name of person responsible for this application: Seal:

NAT: _____

ENG: _____

His/her office/department:

NAT: _____

ENG: _____

Mailing address:

NAT: _____

_____

ENG: _____

_____

Tel: _____ Ext.: _____ Fax:_____

E-mail: _____

Business category:

- ❒ Government agency ❒ Manufacturer
- ❒ Service industry ❒ Translation business
- ❒ Other (please specify):_____

All corporate application forms should include a company prospectus.

INDIVIDUAL

Name: Seal:

NAT: _____

ENG: _____

Mailing address:

NAT: _____

_____

ENG: _____

_____

Tel: _____ Fax: _____

E-mail: _____

Occupation: _____

Company name: _____

Company address: _____

- ❒ Translator ❒ Other: _____
- ❒ University/Institution/Researcher

Specialty: _____

Individual Membership Fees

Initiation fee for individual members	¥ 1,000
Annual dues for individual members	¥ 5,000
Total payment	**¥ 6,000**

The foregoing amount will be paid by ___/___/___ , or within one month from the date this form is mailed.

Corporate Membership Fees

Initiation fee for corporate members
(1 unit = ¥10,000; minimum 1 unit) ___ units = ¥ _____

Annual dues for corporate members
(1 unit = ¥50,000) ___ units = ¥ _____

Please check appropriate box below:

- ❒ MT system developer (minimum 10 units)
- ❒ Other, capital over ¥10 million (minimum 2 units)
- ❒ Other, capital up to ¥10 million (minimum 1 unit)

Total payment (initiation fee + annual dues) **¥** _____

The foregoing amount will be paid by ___/___/___ , or within one month from the date this form is mailed.

Method of Payment

- ❒ Wire transfer to:

  Mizuho Bank
  Ochanomizu Branch
  Tokyo, Japan
  A/C No. 1737479

  For: Asia-Pacific Association for Machine Translation (AAMT)
- ❒ International postal money order in Japanese yen to AAMT at ▸

Appplicants are responsible for all bank charges. Please retain your copy of bank draft or money order as proof of payment.

Today's date: __/__/__

Please send this form to:

Asia-Pacific Association for Machine Translation (AAMT)
c/o Japan Electronics and Information Technology Industries Association (JEITA)
Mitsui Kaijo Bekkan Building, 3F
3-11, Kanda-Surugadai, Chiyoda-ku
Tokyo 101-0062, Japan
Fax: +81 (03) 3518-6472

Association for Machine Translation in the Americas

MEMBERSHIP APPLICATION / RENEWAL FORM

Type of member and membership fee per calendar year:

❒ *Individual* *US$ 60*

❒ *Institutional (nonprofit)* *US$ 200*

Representative: _____

❒ *Corporate* *US$ 400*

Representative: _____

Please return this form, together with your payment or credit card information, to:

Association for Machine Translation in the Americas
PMB 300
1201 Pennsylvania Avenue, N.W. , Suite 300
Washington, DC 2004 USA

Last name(s): _____ *First name(s):* _____ *Title:* _____

Address: _____

_____

Home tel.: _____ *Work tel.:* _____ *Fax:* _____

E-mail: _____ *Website:* _____

Affiliation: _____

Professional associations: _____

Area of specialization:

❒ *MT User* ❒ *MT Developer* ❒ *MT Researcher* ❒ *Translator* ❒ *Manager* ❒ *Other* _____

Method of Payment

❒ *Check enclosed*

❒ *Credit card*

Type of credit card: ❒ *Visa* ❒ *MasterCard* ❒ *American Express*

Card number _____ *Exp. Date:* ___/___

European Association for Machine Translation

APPLICATION FOR MEMBERSHIP

Please return this form, together with your payment or credit card information, to:

EAMT Secretariat, c/o TIM / ISSCO
Université de Genève
École de Traduction et d'Interprétation
40, blvd du Pont-d'Arve
CH-1211 Geneva 4, Switzerland

Type of member and membership fee per calendar year:

❒ *Individual* *SFr 50*

❒ *Non-profit-making institution* *SFr 175*

Representative: _____

❒ *Profit-making institution* *SFr 350*

Representative: _____

Last name(s): _____ *First name(s):* _____ *Title:* _____

Address: _____

_____

Home tel.: _____ *Work tel.:* _____ *Fax:* _____

E-mail: _____ *Website:* _____

Institution / organzation: _____

Area of specialization:

❒ *MT User* ❒ *MT Developer* ❒ *MT Researcher* ❒ *Translator* ❒ *Manager* ❒ *Other* _____

Method of Payment

❒ *Cheque payable to EAMT, enclosed*

❒ *Banker's draft (copy enclosed) to account no. 351.091.40L*
Union Bank of Switzerland
Bahnhofstrasse 45
CH-8021 Zürich, Switzerland

Please note: All bank charges must be borne by the applicant

❒ *Credit card (please provide information at right ➔)*

Type of credit card: ❒ *Visa* ❒ *Eurocard*

Card number _____

Expiration date: ___/___