```
Mt-list Digest, Vol 111, Issue 1
================================
```

Today's Topics:

```
   1. [ASLIB-08] no invoice yet for me, and questions about
      registrations (Christian Boitet)
   2. Call for Participation: WMT 2009 Shared Task (Philipp Koehn)
   3. ELRA - Language Resources Catalogue - Update  (ELDA)
   4. 3rd MT Marathon - Call for participation (Jan Hajic)
```

```
---------- Forwarded message ----------
From: Christian Boitet <Christian.Boitet@imag.fr>
To: nadamides@aslib.com, <mt-list@eamt.org>
Date: Thu, 23 Oct 2008 11:17:44 +0200
Subject: [Mt-list] [ASLIB-08] no invoice yet for me, and questions about registrations
Dear Nicole,                                                      23/10/08
```

I got no answer from your to my e-mail, and no invoice for my registration (with 20% discount as you kindly proposed).

I also asked whether registration would be free for the 3rd, 4th, etc. memeber of the same lab. As it is, M.Daoud, Ch.Boitet and V.Bellynck will participate (tickets booked), and Cong Phap Huynh may also stop for 1 1/2 day on his way to Tokyo.

Best regards,
Ch.Boitet

At 16:14 +0100 22/10/08, <nadamides@aslib.com> wrote:
> TRANSLATING AND THE COMPUTER 30 Conference
> 27-28 November 2008, London
> Supported by BCS Natural Language Translation Specialist Group, EAMT, IAMT, ITI, IoL, LISA and TILP. Sponsored by CNGL and LTC.
> Come and join us at the Holiday Inn Camden Lock to celebrate the 30th conference with an exciting programme of papers and panels. The programme, and other details, can be found at www.aslib.com/conferences. There are two additions and one change to the previously advertised programme:
> a) the keynote presentation on Thursday 27:
> Synchronizing translations in a Wiki with TikiWiki CMS/Groupware
> Marc Laporte, AvanTech.net, Canada
> Classic translation workflow systems are based on several assumptions.
> For example, they assume the presence of a master language and that the source document be "final" before translation is started. But what to do for a wiki, which has no master language and, by the wiki's nature, when a document is never "final"? The presentation will address 7 assumptions (including the two mentioned above) of the traditional translation workflows and how these are challenged in the Web & Wiki world. Finally, we will describe a working solution to these issues, implemented in TikiWiki, a full-featured, open source, multilingual Web Content Management System (CMS). This solution is currently used by the Firefox support site at support.mozilla.com. More information on Wiki-Translation.com
> b) the keynote presentation on Friday 28:
> Tools for Integration
> Ulla Falk-Petersen, Director, European Court of Auditors, Luxembourg
> Over the years, translator oriented IT tools have become instrumental to translation work. An important advantage of the use of IT tools stems from the possibilities in the way of integration and interoperability they offer. A case study is presented to illustrate how the European Court of Auditors has used IT tools to coordinate and harmonise translation work whilst providing quality assurance and efficiency. The use of IT tools as a facilitator of integration at several levels is stressed in the context of the 2004 and 2007 enlargements, which meant a twofold increase in the number of official languages and of translators in our organisation.
> c) the last paper on Friday 28 has changed to:
> A Quality Control Framework for Gold Standard Reference Translations: The Process and Toolkit Developed for GALE
> Lauren Friedman, University of Pennsylvania, USA
> While a multitude of MT evaluation metrics exist, most require one or more gold standard references. For the DARPA GALE program, source data is translated according to detailed guidelines2 and high quality standards, but these raw translations then undergo a rigorous and carefully constructed quality control process to create the final references. This resource intensive process, which utilizes a specially designed toolkit, QCTool, was developed to meet GALE's need for multiple editing passes on translation data wherein each intermediary version of the translation is preserved. This paper describes this framework and suggests that it may be successfully applied to improve and streamline other large scale translation projects.
> There are panels looking at the past 30 years and at the future, as well as ample opportunities to network with other colleagues in the field.
> For more information, including how to book your place, please visit: www.aslib.com/conferences. If

there are 3 of you coming from the same organization, the third delegate can attend for 200GBP plus
VAT.
>
> Regards
>
> Nicole Adamides, ASLIB Training
> The Holywell Centre, 1 Phipp Street, London, EC2A 4PS
> Tel: 020 7613 3031   Fax: 020 7613 5080
> www.aslib.com/training Email: training@aslib.com
>
> _____
> Mt-list mailing list

--

Christian Boitet
(Pr. Universite' Joseph Fourier)
      ======= Notez svp /Please note ======= GETA --> GETALP ============
Groupe d'Etude pour la Traduction Automatique
              et le Traitement Automatisé des Langues et de la Parole
G   E   T   A   L   P

NOUVEAU FAX / NEW FAX (1/8/08)

GETALP, LIG-campus, BP 53   (ex: GETA, CLIPS, IMAG-campus)
Tel: +33 (0)4 76 51 43 55 / 51 48 17   Fax: +33 (0)4 76 63 56 86
385, rue de la Bibliothe`que   Mel: Christian.Boitet@imag.fr
38041 Grenoble Cedex 9, France

---------- Forwarded message ----------
From: "Philipp Koehn" <pkoehn@inf.ed.ac.uk>
To: "corporalist corporalist" <corporalist@gmail.com>, mt-list@eamt.org, WMT09@googlegroups.com
Date: Sun, 2 Nov 2008 01:13:40 +0000
Subject: [Mt-list] Call for Participation: WMT 2009 Shared Task
Ladies and Gentlemen:

We are organizing again next year the

    Workshop on Statistical Machine Translation
    http://www.statmt.org/wmt09/

which will be take place under the umbrella of the EACL 2009 conference
in Athens, Greece, March 30 and 31, 2009.

As in the previous year, the workshop will include a shared task on translation
of European language pairs. Translation quality will be evaluated on a shared,
unseen test set of news stories. We provide a parallel corpus as training data,
a baseline system, and additional resources for download. Participants may
augment this system or use their own system.

The testing period for the shared task is December 8-12.

The training and development data is now online.

Please register for further announcements on the workshop web page.
No formal registration is required, but it is highly recommended.

** Changes this Year **

After a number of extensions over the last years, this year's translation task
will be more focused. The motivation for this is to have a clearly defined task
and to be able to gather sufficient human judgement data to make as many
statistically significant distinctions as possible.

    * News Translation Only: We will only evaluate performance on a set of
news stories that we prepared for this evaluation. As in the previous year, the
news stories are taken from major news outlets as the BBC, Der Spiegel,
Le Monde, etc. during the time period of September-October 2008.
Last year's test set will serve as the development set, which we
re-release in a
slightly cleaned form.

* Official Metric Manual Sentence Ranking: While we continue to experiment
with different forms of manual and automatic metrics, we use as official metric
human judgment of preference on a sentence-by-sentence basis.

    * More Training Data: The parallel corpora Europarl and News Commentary
are extended, a large French-English corpus (currently 200 million words) was
added, and we also proved 50-500 million word monolingual training data
from the test domain of news stories.

    * Constraint vs. Unconstraint: You may use additional resources
(training data,
knowledge bases such as existing translation systems), but you have to
flag that
your system uses these. We will distinguish system submissions that used the
provided training data (constraint) from submissions that used significant
additional data resources. Note that basic linguistic tools such as taggers,
parsers, or morphological analyzers are allowed in the constraint condition.

We would welcome your participation in the shared task.

Chris Callison-Burch (Johns Hopkins University)
Philipp Koehn (University of Edinburgh)
Christof Monz (University of London)
Josh Schroeder (University of Edinburgh)

Organizers of the Workshop

---------- Forwarded message ----------
From: info@elda.org
To: destinataires inconnus:;
Date: Fri, 24 Oct 2008 12:06:45 +0200
Subject: [Mt-list] ELRA - Language Resources Catalogue - Update
Our apologies if you have received multiple copies of this announcement.

****
ELRA - Language Resources Catalogue - Update
****

ELRA is happy to announce that 3 new Speech Resources are now available in its catalogue:
*
ELRA-S0289 OrienTel Jordan MCA (Modern Colloquial Arabic) database*
This speech database contains the recordings of 757 Jordanian speakers recorded over the Jordanian
fixed and mobile telephone network. Each speaker uttered around 49 read and spontaneous items.
For more information, see: http://catalog.elra.info/product_info.php?products_id=1077

ELRA-S0290 OrienTel Jordan MSA (Modern Standard Arabic) database
This speech database contains the recordings of 556 Jordanian speakers recorded over the Jordanian
fixed and mobile telephone network. Each speaker uttered around 49 read and spontaneous items.
For more information, see: http://catalog.elra.info/product_info.php?products_id=1078

ELRA-S0291 OrienTel English as spoken in Jordan database
This speech database contains the recordings of 578 Jordanian speakers of English recorded over the
Jordanian fixed and mobile telephone network. Each speaker uttered around 47 read and spontaneous
items.
For more information, see: http://catalog.elra.info/product_info.php?products_id=1079

For more information on the catalogue, please contact Valérie Mapelli mailto:mapelli@elda.org

Visit our On-line Catalogue: http://catalog.elra.info
Visit the Universal Catalogue: http://universal.elra.info
Archives of ELRA Language Resources Catalogue Updates: http://www.elra.info/LRs-Announcements.html

---------- Forwarded message ----------
From: Jan Hajic <hajic@ufal.mff.cuni.cz>
To: corporalist <corporalist@gmail.com>, WMT09@googlegroups.com, mt-list@eamt.org
Date: Mon, 3 Nov 2008 23:37:06 +0100 (CET)
Subject: [Mt-list] 3rd MT Marathon - Call for participation

                    CALL FOR PARTICIPATION

THIRD MACHINE TRANSLATION MARATHON 2009

The MT Marathon 2009, organized by the Institute of Formal and
Applied Linguistics of the Charles University in Prague, Czech
Republic, is the third in a series of MT Marathons organized by
the EU Euromatrix research project on Machine Translation.

The EuroMatrix consortium invites researchers, developers,
students, and users of machine translation for participation. The
event will feature

- Winter School classes on current methods in statistical MT
- Research showcase
- Open source convention on resources for machine translation,
 this time with an open call for papers (separate announcement
 will follow)
- Lab hands-on experience for system developers, students and
 programmers
- Workshop on evaluation of European language translation

Where:
      Prague's Lesser Town, the newly renovated historical
      building of the Computer Science School of the Charles
      University in Prague
When: January 26-31, 2009
How: Registration is now possible!
How much: Attendance is free of charge, but limited.

For more information and online registration please go to
http://ufal.mff.cuni.cz/euromatrix/mtmarathon.

About the MT Marathon

MT Marathon is organized yearly by the EuroMatrix machine
translation research project funded by the European Union under
its "Cooperation programme" as a STREP project
FP6-IST-5-034291-STP. In January 2009, it will be third MT
Marathon organized by EuroMatrix. MT Marathon consists of several
events taking place at the same place to allow for free flow of
thoughts and exchange of information and experience: a "spring
school" (this time more like a "winter school") with associated
lab lessons, invited research talks, and a hands-on experience
with Open Source MT tools. Participants will also experience
evaluating Machine Translation systems (with some hands-on
experience in actual subjective evaluation of MT systems taking
part in the WMT 2009 competition - see more at
http://www.statmt.org/wmt09/). This year, talks presenting some of
the available OpenSource tools in more detail will also be planned
throughout the week (see the call for papers). Please find more
about the current MT Marathon and the previous ones at
http://ufal.mff.cuni.cz/euromatrix/mtmarathon.

About Euromatrix

The EuroMatrix project (http://www.euromatrix.net) aims at a major
push in machine translation (MT) technology applying the most
advanced MT technologies systematically to all pairs of EU
languages. Special attention is being paid to the languages of the
new and near-term prospective member states. As part of this
application development, EuroMatrix designs and investigates novel
combinations of statistical techniques and linguistic knowledge
sources as well as hybrid MT architectures. EuroMatrix addresses
urgent European economic and social needs by concentrating on
European languages and on high-quality translation to be employed
for the publication of technical, social, legal and political
documents. EuroMatrix aims at enriching the statistical MT
approach with novel learning paradigms and experiment with new
combinations of methods and resources from statistical MT,
rule-based MT, shallow language processing and computational
lexicography/morphology.

The main objectives of the project are:

* Translation systems for all pairs of EU languages, with a
  special focus on the languages of new and near-term prospective
  member states
* Efficient inclusion of linguistic knowledge into statistical
  machine translation
* The development and testing of hybrid architectures for the
  integration of rule-based and statistical approaches
* Organization, analysis and interpretation of a competitive
  annual international evaluation of machine translation with a
  strong focus on European economic and social needs
* The provision of open source machine translation technology
  including research tools, software and data
* A systematically compiled and constantly updated detailed survey
  of the state of MT technology for all EU language pairs based on
  the developed systematic translation between all EU languages,
  the comparative MT evaluations and an inventory of available and
  needed tools, components, lingware and data.