Sean Sherlock T.D., Minister for Research and Innovation has launched a guide which urges secondary students to consider language and technology at third level for a potential career in the localisation sector.

The Guide, produced by CNGL, says that Ireland needs to significantly increase its supply of language and technology graduates if the country is to maintain its leadership position in the multi-million euro localisation and global services sector.

Speaking at the launch at Government Buildings in January, Minister Sherlock stated, “Many of the world’s largest software and Web companies co-ordinate their localisation activities in Ireland and the sector contributes an estimated €680 million annually to the Irish economy. It is imperative that we highlight the opportunities available to our young people in this sector.”

CNGL’s careers guide has been distributed to all second level schools in Ireland. It features profiles of professionals who hold leading roles across the localisation industry and also lists the university courses across computing, languages, and business that would provide a foundation for a career in this high-growth area.

CNGL Researchers secure €290k in Commercialisation Funding

Researchers at the Science Foundation Ireland-funded Centre for Next Generation Localisation (CNGL) have secured €290,000 in funding awards to help develop leading-edge technologies with high commercial potential in the area of global content intelligence.

The four projects, led by researchers at DCU and TCD, span the themes of cross-site Web personalisation (Dr. Owen Conlan, TCD); Machine Translation and Translation Memory integration into localisation workflows (Prof. Josef van Genabith, DCU); development of an integrated software suite to provide next-generation personalised multilingual customer care (Prof. Vincent Wade, TCD); and user-customisable policy monitoring for multi-tenant cloud architectures (Dr. Claus Pahl, DCU).

The investment is being made as part of Science Foundation Ireland’s Technology Innovation Development Award (TIDA) programme, which enables research groups to focus on the first steps of an applied research project which may have a commercial benefit if further developed.

Read More
MacCraith stated that "Josef’s significant research accomplishments, presented in more than 100 peer-reviewed international research publications, have not only shaped the evolution of computational linguistics worldwide; they have also helped brand DCU, CNGL and Ireland as world leaders in language technology."

Prof. Josef van Genabith, Director of CNGL, was presented with the 2011/2012 DCU President’s Research Award for Science and Engineering at a ceremony held in February at The Helix, DCU.

The Award recognises Josef’s exceptional contribution to the field of computational linguistics and his role in spearheading DCU’s leadership in academia-industry collaborative research.

Speaking at the Awards ceremony, DCU President Prof. Brian MacCraith highlighted Josef’s impressive research track record. "Josef van Genabith’s tremendous achievements in terms of grant income, graduated research students and publications history, combined with his international profile at the highest level of European research management and further afield, make him an outstanding candidate for this award”, said Prof. MacCraith.

MacCraith stated that “Josef’s significant research accomplishments, presented in more than 100 peer-reviewed international research publications, have not only shaped the evolution of computational linguistics worldwide; they have also helped brand DCU, CNGL and Ireland as world leaders in language technology."

CNGL to help devise W3C Standards for Multilingual Web

CNGL is to play a central role in the development of new W3C (World Wide Web Consortium) standards to simplify creation of Web content in the world’s languages. W3C is an international community that develops open standards to ensure the long-term growth of the Web.

CNGL researchers and industry partners will work with international collaborators on the new MultilingualWeb-LT (Language Technology) Working Group to develop standard means to support the translation and adaptation of Web content to local needs, from its creation through to its delivery to end users. By so doing, the new standards will help to remove language barriers to international trade and facilitate the free flow of information across language borders.

Dr. Alexander O’Connor of TCD is one of a number of CNGL members participating in the MultilingualWeb-LT project. O’Connor says: “Ireland is in the vanguard of language services and global content intelligence. We believe that the development of standards for translation and adaptation of Web content to local needs will greatly improve access to the right content across linguistic barriers.”

Dr. David Lewis (TCD) and Dr. David Filip (UL) will co-chair the Group.

Read the full story
One hundred of Ireland’s brightest young problem solvers converged at Dublin City University on 20th March 2012 for the finals of the All Ireland Linguistics Olympiad (AILO).

The Olympiad, run by CNGL, is designed to introduce secondary school students to the world of linguistics and challenges them to solve puzzles requiring computational thinking and logic.

During an exciting day of competition, finalists competed in individual and team rounds to determine the top young code-breakers in Ireland.

Mary Mitchell O’Connor, T.D. met with students and teachers and strongly encouraged the students to make the most of their aptitudes for reasoning and lateral thinking by pursuing high-tech careers.

Speaking at the finals, Cara Greene, CNGL Education and Outreach Manager said, “AILO is proving to be very effective in generating excitement among Irish teenagers about the study of language and the honing of their problem-solving skills. We are delighted to see former competitors now going on to study language, linguistics and computing at third level, as having an adequate supply of graduates in these disciplines is vital to the success of Irish export-focused industries.”

Qualifying rounds of the competition were hosted in 60 secondary schools during recent weeks. The top performers from the national finals of the competition will be selected to represent Ireland at the International Linguistics Olympiad in Slovenia in July 2012. The winners will be announced in early April.

Read the full story

CNGL highlights potential of Big Data with Hadoop Hackathon

CNGL ran a ‘Hadoop Hackathon’ on 10th – 11th March 2012 at DCU. The Hackathon was aimed at encouraging students and industry partners to unlock the potential of ‘Big Data’ to deliver enhanced customer insights and competitive advantage.

The two-day Hackathon saw participants using Hadoop to solve natural language processing problems. The event was facilitated by Miles Osborne of the School of Informatics, University of Edinburgh, U.K. CNGL industry collaborator Tony O’Dowd of Xcelerator describes the premise of the Hackathon: “The Hadoop Hackathon 2012 is a great way to learn and experiment with the Hadoop framework. Working in small teams, and learning at an accelerated pace, each team hacks together a solution, searching for meaningful outputs from enormously large data-sets. Always mindful that the clock is ticking and time is of the essence, you put into practice what you’ve just learned in the previous 10 minutes. Code, debug, re-code and try again is the order of the day!”

Read the full story

AILO National Final 2012

IRISH STUDENTS TEST THEIR CODE-BREAKING SKILLS AT LINGUISTICS OLYMPIAD

Read More

AILO TUTORS

AILO finalists are tutored by staff and students at CNGL’s partner institutions. Many thanks to:

- CNGL researchers John Moran (TCD), Liliana Mamani Sanchez (TCD), Lorraine Goeuriot (DCU), Antonio Toral (DCU) and Ian O’Keeffe (UL)
- TCD undergraduate students Eimear Maguire, Melissa Sorensen, Dean Graham and Catherine Lynch
- Cara Greene and Karl Kelly of CNGL’s E&O Team

CNGL UNDERGRADUATE INTERNSHIPS 2012

CNGL will offer eight 8-week internships to undergraduate students, starting in June 2012. The primary aim of this programme is to offer undergraduate students the opportunity to participate in and contribute to exciting research projects at CNGL. This allows the student to use leading research facilities and inspires these students to take the first step on a path to a research career. Details of internships will be posted shortly to: www.cngl.ie/internships.html

COMPUTEHY IMPACT IS IMMENSE

A survey of over 400 students who completed the CNGL-supported ComputeHY transition year Web design and programming course at DCU in January has highlighted the course’s success in kindling students’ interest in computing. Although 63% of students indicated that they didn’t consider a Computing degree before taking the course, 80% of participants would now consider Computing as a choice for third level. Read More
CNGL and Irish localisation company Cipherion Translations have announced that they are to collaborate on the development of advanced systems for rapid translation of a variety of time critical documents.

Cipherion will draw on CNGL expertise in the areas of machine translation, crowd-sourcing and aspects of game-inspired behaviour. The research outcomes will be integrated into a new high-tech product that Cipherion aims to bring to market in the coming months. The partnership will be operated under Enterprise Ireland’s Innovation Voucher scheme.

“We are excited about the opportunity to work with a leading research centre”
- Mark Rodgers, CEO, Cipherion

Welcoming the opportunity to support Cipherion’s new product development initiatives, Steve Gotz, Commercial Development Manager at CNGL, said, “Once again Enterprise Ireland’s Innovation Voucher initiative is helping CNGL to build links with the critical base of SMEs in Ireland that continue to create jobs and economic development.

Our new partnership with Cipherion is just one example of the many ways in which CNGL can assist firms with their market-led R&D efforts. We are excited about working with the team from Cipherion and contributing to the company’s future growth.”

 Cipherion Translations and CNGL not-for-profit spinout, The Rosetta Foundation, have announced that they are to offer awards and studentships to students at Dublin City University.

- Cipherion Translations will again sponsor a Best Thesis Award for undergraduate students on DCU’s BA in Applied Language and Intercultural Studies and one from the MA in Translation Studies or M.Sc. in Translation Technology. The interns will work at Cipherion during June-October 2012.

- The Rosetta Foundation has agreed to sponsor two students from the MA in Translation Studies and M.Sc. in Translation Technology in the School of Applied Language and Intercultural Studies at DCU, one from the BA in Applied Language and Intercultural Studies and one from the MA in Translation Studies or M.Sc. in Translation Technology. Both will work under the supervision of Dag Schmidtke, Program Manager for Language Technology Strategy at Microsoft Ireland.

Jinming joined Microsoft Ireland Research in Dublin as a Computational Linguistics intern in October 2011. He is working under the supervision of Dag Schmidtke, Program Manager for Language Technology Strategy at Microsoft Ireland.

Jinming is thoroughly enjoying the internship, which sees him working currently on the evaluation of multilingual search quality of office.com and mining translation pairs from external resources.
Digital Content Management (DCM)

High quality publications and industry engagement the focus for DCM in Q1 2012

DCM has recorded significant success in a number of areas, reflecting the maturity of the high-quality research undertaken by the team. Four students have completed or submitted their PhDs Theses: Bo Fu, Ben Steichen, Walid Magdy and Ian O’Keeffe. Two students passed their *viva voce* examinations with flying colours, with the others expecting similar outcomes soon. Bo, Ben and Walid have gone on to further research at institutions with excellent international reputations in Canada and Qatar respectively, and we look forward to collaborating with them in future.

DCM has continued to develop and validate the highly successful platform for multilingual content management, including Rami Ghorab’s personalised Multilingual Information Retrieval Framework, Killian Levacher’s Content Slice Service, and Ben Steichen’s Personalised Multilingual Customer Care environment. Preliminary results from each of these experiments are extremely promising, and we expect to see high-quality publications resulting in the near future.

Industrial engagement continues to play a major role, with two DCM students engaged in internships with Microsoft Dublin. Both students have been given an exceptional opportunity to work closely with Microsoft Ireland on information retrieval problems, with a view from ‘inside the walls’.

DCM publication highlights this quarter include the Journal of Information Retrieval, the Journal of Web Semantics, Conference on Information and Knowledge Management, and the International Semantic Web Conference, amongst many others.

In affiliate news, the SFI AMAS Principal Investigator project received an excellent review, and the EU FP7 CULTURA project will undergo its first-year review in the next quarter. Finally, the MultilingualWeb-LT project had its kick-off meeting, with SF2 and DCM researchers involved in this project which includes close collaboration with W3C.

Next Generation Localisation (LOC)

To the fore this quarter were international conference presentations and tech transfer activities

LOC researchers were well represented at the 5th African Conference on FOSS and the Digital Commons (IDLELO 5) in Abuja, Nigeria this March as Principal Investigator Reinhard Schäler and researcher Solomon Gizaw both presented LOC research at this event. LOC was also represented at the GALA conference in Monaco and the final W3C Multilingual Web workshop in Luxembourg.

On the Solas (Service-Oriented Localisation Architecture Solution) front, work is progressing steadily with Ian O’Keeffe producing a Cross-CNGL Collaboration Report and Solas Integration Plan and working on component templates/framework for use by future developers, as well as prototyping a Translation Memory Checker component with input from Joss Moorkens as to how to categorise inconsistencies.

A number of Solas component technology Invention Disclosures were submitted to the University of Limerick (also being submitted to CNGL IP Committee). Well done to the PhD researchers involved.

Also on the tech transfer front, Translation eXchange (now known as Solas Match) code is being open-sourced licensed by the University of Limerick under GPL. Solas Match aims to support community-driven translation and localisation.

In other news, Rajat Gupta has departed LOC and returned to India as of February 2012. We wish him all the best in his future endeavours.
Systems Framework (SF)

Standards development and Wizard of Oz platform experimentation were key Q1 themes for SF

SF researchers have recorded publication success in early 2012. Anne Schneider and Saturnino Luz published a paper at the International Workshop on Spoken Language Translation (IWSLT’11) on speaker alignment in synthesised, machine translated communication and another paper at the Workshop on Innovation and Applications in Speech Technology (IAST’12) on the evaluation of spoken language translation, giving an overview of error classifications and applications. David Lewis, along with Alexander O’Connor of DCM and international collaborators, had a paper accepted to LREC2012 on using linked data for language resource sharing in the long tail of the localisation market.

The relationship with TCD’s Centre for Deaf Studies was strengthened, resulting in the collaborative running of a focus group with medical interpreters and a presentation in the SIGNALL 3 Seminar at Lessius University College, Antwerp, by Ilana Rozanes of Deaf Studies at TCD.

Seminars were held in Ireland (TCD) and abroad (UK and Austria), in which students were taught the Wizard of Oz prototyping method through use of CNGL’s WebWoZ platform. Stephan Schloegl carried out interviews with 20 researchers from academia and industry on Wizard of Oz experimentation.

CNGL’s role at the forefront of standards development has been further reinforced with the appointment of David Lewis and David Filip as co-chairs of the new MultilingualWeb-LT working group established at W3C.

On the Education and Outreach front, John Moran demonstrated his iPad German cloze test app to students at the BT Young Scientist & Technology Exhibition in the RDS, Dublin in mid-January.

Integrated Language Technology (ILT)

Significant funding awards and publication success were the highlights of Q1 for ILT

ILT researchers have made a strong start to 2012 on the publications front, with five papers accepted to LREC (Istanbul), one to CICLING (New Delhi), one to ACL demo track (Avignon), a journal paper published in Prague Bulletin of Mathematical Linguistics, and a number of papers submitted to the EAMT, ACL and EMNLP peer-reviewed conferences.

Martin Emm’s paper (together with Hector Franco-Peña) on “Order equivalences between distance similarity measures on sequences and trees” won best paper at the International Conference on Pattern Recognition.

Declan Groves has received funding for a 12-month commercial project worth €152k from Enterprise Ireland’s Commercialisation Fund on Data Analytics for MT. This project represents close collaboration with Tony O’Dowd’s Xcelerator Ltd.

John Tinsley received an EI Feasibility Grant worth €15k to look into business models for patent translation.

In affiliate project news, negotiations for the QTLaunchPad EU FP7 were successfully completed. This project focuses on high quality for MT and is led by DKFI. DCU’s participation is led by Josef van Genabith (DCU receiving €426k in funding) and other partners are Athena Research Center (Greece) and University of Sheffield.

The T4ME (META-NET) project and PANACEA project reviews took place and were both successful. ILT has input into the LT-Web EU FP7 project, which kicked off in Berlin in January. €396k in funding has been awarded in total for all CNGL partner universities (TCD, UL and DCU) involved in the project.

We continue our engagement with Applied Language Solutions, who are providing access to their SmartMATE system which is being used to teach MT to translation technology masters students in DCU.

Ergün Biçici has joined the team as a post-doc working on Regression-based MT and estimation of upper-bounds for MT performance.
In Brief

CNGL enjoys considerable success with research publications. To view full details of our publications history, and to download copies of our publications, visit: www.cngl.ie/researchpub.html

STAFF NEWS

Congratulations to:

Walid Magdy, who successfully defended his PhD thesis entitled “Toward Higher Effectiveness for Recall-Oriented Information Retrieval: A Patent Retrieval Case Study” in January. Walid was supervised at DCU by Dr. Gareth Jones.

Dr. Lamia Tounsi, Postdoctoral Researcher at DCU, on the birth of her baby boy, Amine Cian.

Dr. Seamus Lawless on his appointment to the role of Postdoctoral Research Fellow and Lecturer in the School of Computer Science and Statistics, TCD.

Bo Fu, who successfully defended her PhD thesis “Semantic-oriented cross-lingual ontology mapping” under the supervision of Prof. Declan O’Sullivan and Dr. Rob Brennan at TCD.

METADATA GROUP

The CNGL Metadata group (MDG) currently comprises ten researchers from TCD, UCD, DCU and UL. The MDG aims to work towards collecting, recording and documenting the usage and application of metadata across the demonstrator systems and research conducted as part of CNGL.

The MDG has been working on strategies for tracking the provenance of a piece of digital content throughout the localisation process. This provides, from a management perspective, monitoring and tracking of changes made to content, and thus establishes exactly when a piece of content was changed and by whom, and providing a link to the evolved content.

The W3C Provenance Interchange Working Group defines “Provenance [as referring to] the sources of information, such as entities and processes, involved in producing or delivering an artifact.” The MDG sees an artifact as a digital piece of content. When an artifact is combined with provenance information, it provides history for tracking exactly how and when a piece of digital content was created, edited or re-formed.

With traditional print media, the content lifecycle can be tracked from creation through to publication. However, with the evolution of social media — blog posts, tweets, wiki pages and Web forums — the traditional approaches for tracking provenance no longer scale. This is especially true when content is being translated and localised from its source language into any number of different target languages at the point of creation.

Provenance information is combined, in the CNGL metadata platform, with Service (showing which services were applied to the content) and Content models (identifying the exact form of a piece of content). This platform provides the history of changes made to, services applied to, and the exact form of a piece of content passing through the localisation process.

Future work will evaluate the scalability and adaptability of this platform, across different operational use-cases and integrate into the platform a method for creating and tracking annotations to content.

For more information, contact Dominic.Jones@cs.tcd.ie
Following his recent successful PhD defence, Walid Magdy shares an overview of his research and updates us on his current work in Qatar...

I have finished my PhD on Information Retrieval with Gareth Jones in the School of Computing, Dublin City University. The title of my thesis was "Toward higher retrieval effectiveness for recall-oriented information retrieval: a patent search case-study". As it appears from the thesis title, my work was focused on developing efficient and effective approaches and tools for achieving better search system for patents. My internal examiner from DCU was Prof. Alan Smeaton, and my external examiner was Dr. Barrou Diallo, the head of research of the European Patent Office. This work was part of CNGL, where I found the opportunity to learn a lot and do high quality work.

Currently, I am a postdoctoral researcher in the Qatar Computing Research Institute (QCRI), a member of Qatar Foundation in Doha, Qatar. I am working in the Arabic Language Technology (ALT) group, where we focus on improving the language technology for the Arabic language, with a particular concentration on information retrieval and Machine translation.

“I am now postdoctoral researcher with the Qatar Computing Research Institute”

Where are they now?

Walid Magdy

Dr. Özlem Çetinoğlu

Özlem Çetinoğlu spent three very productive years as postdoctoral researcher with CNGL at DCU. She now explains how CNGL has shaped her career...

“When I joined the CNGL, I was a PhD student writing her thesis. In the three years I spent in Dublin I became an independent researcher, laying the building blocks with each step I took.

I had the chance to work on projects that taught me a lot and moreover had the liberty to pursue the research questions I am interested in.

Core research was only one aspect of my advancement there. CNGL gave me the opportunity to gain invaluable experience in other aspects such as teaching, organising conferences and writing project proposals.

Four months ago, I moved to University of Stuttgart as a postdoctoral researcher. I continue to work on parsing and partially serve as a lecturer and an advisor. Even in simple daily work matters, I can observe CNGL’s contribution to my career and my personality. But this is not my biggest achievement from the Centre. Above all, I have met great people, people with whom I would like to continue to collaborate as colleagues, and whom I would like to keep in my life as dear friends.

Greetings from Germany!”

Where are they now?

Ergun Biçici, who has joined DCU as Postdoctoral Researcher with the ILT track.

Yvette Graham, who has joined DCU as Postdoctoral Researcher working with the ILT track and the CNGL-affiliated CoSyne project.

Rasoul Samad Zadeh Kaljah, who has joined the National Centre for Language Technology, DCU as PhD candidate.

Raphael Rubino, who has joined the National Centre for Language Technology, DCU as Postdoctoral Researcher.

Check out Reinhard’s slides at http://slidesha.re/HdMxYT

RENÉE SALZMAN (GALA Co-Founder), REINHARD SCHÄLER (CNGL), HANS FENSTERNACKER (GALA Board Chair) and MARIA JOSÉ VELASCO (GALA founding member and MandragòN Lingua)

CNGL TO CO-LOCATE INNOVATION SHOWCASE WITH LRC CONFERENCE 2012

CNGL has announced that two of the most popular entries in the events calendar of Ireland’s localisation and digital content sectors are to co-locate in Limerick this September. The 17th Annual LRC Internationalisation & Localisation Conference will be held on 20th—21st September. The conference will be preceded by CNGL’s Localisation Innovation Showcase on 19th September. Further Details

The workshop was attended by both local and international researchers, as well as key industrial players and leaders in the field. Some of the main local points were the keynote presentations given in the morning of each day.

On Friday morning Prof. Steve Renals (University of Edinburgh) outlined where cutting-edge speech recognition and synthesis are presently and the directions he envisages for bringing about more natural speech technology. Dr. Nicolas d’Alessandro gave an inspiring Saturday morning talk on taking speech synthesis out of the restrictive box of the text-to-speech paradigm and bringing it to a new more interactive and performative level.

The workshop was topped off with a social trip which consisted of hill-walking in the Wicklow Mountains and ended up in Johnny Foxes pub for a well needed dinner and drinks.

Aside from the keynote talks there was an extremely high level of presentations on topics ranging from assistive technologies and multimodal applications to new approaches for modelling tone-of-voice and expressive speech. There were also many impressive presentations in the poster and demonstration sessions on dialogue management, multilingual and language learning applications as well as the personalisation and localisation of speech technology applications.

We have received numerous feedback emails from participants saying how they felt that there was such a high level of quality presentations for such a small workshop and that the friendly atmosphere was particularly conducive for networking and building up potential collaborations.

Presentation slides and photos are available at: http://muster.ucd.ie/workshops/iast/

This workshop review was kindly provided by Dr. Joao Cabral and Mr. John Kane.
Upcoming Events

June 2012

• 10th Annual LRC Summer School
  13th-15th June 2012, Limerick, Ireland

• META-FORUM 2012
  20th-21st June 2012, Brussels, Belgium  Further Details

July 2012

• International Linguistics Olympiad 2012
  29th July-2nd August 2012, Ljubljana, Slovenia  Further Details

September 2012

• CNGL Localisation Innovation Showcase
  19th September 2012, Limerick, Ireland

• LRC XVII Conference
  20th-21st September 2012, Limerick, Ireland

December 2012

• AGIS’12- The 4th annual Action for Global Information Sharing (AGIS)
  10th-11th December 2012, Sri Lanka

Subscribe to CNGL News Alerts!

Want to keep abreast of news and developments at CNGL? You can now subscribe to news alerts.

Sign up online now.

The Centre for Next Generation Localisation (CNGL) is a dynamic academia-industry partnership that conducts ground-breaking research into next-generation localisation (i.e. the process of adapting digital content to culture, location and linguistic environment) by addressing key challenges in information volume, access and personalisation.

Got an idea for a news article? Perhaps you’d like to contribute?
Contact our editor, Laura Grehan, at Email: laura.grehan@dcu.ie or Tel: +353 (0)1 7006705

www.cngl.ie