INTERNATIONAL CONFERENCE ON DIGITAL LIBRARIES (ICDL) 2013
Vision 2020: Looking Back 10 Years and Forging New Frontiers

27-29 November 2013
India Habitat Centre, Lodhi Road, New Delhi

www.teriin.org/events/icdl

ICDL 2013 takes off in style in a power-packed ceremony at India's cultural hub

The fourth edition of the International Conference on Digital Libraries (ICDL) with a theme of 'Vision 2020: Looking Back 10 Years and Forging New Frontiers', was inaugurated by Shri Hamid Ansari, the Hon'ble Vice-President of India, in the presence of large library fraternity. Dr M M Pallam Raju, the Hon'ble Minister of Human Resource Development, Government of India, also graced the glittering ceremony.

Welcoming the dignitaries and delegates, Dr R K Pachauri, Director-General, TERI, stated in his opening remarks that the 21st century is going to be a century of knowledge and there is a need for clearly setting out programmes and policies for digitization of massive resources. He informed the delegates that TERI has always taken lead in documentation of research and preservation of knowledge.

Mr Prabir Sengupta, Chair, ICDL 2013, spoke in detail on the theme of the conference and its relevance for professionals in the field. Dr Pallam Raju stated that India has always been a hub of knowledge and is known as an institution of higher learning. He further stated that we are all living a life driven by digital content. Tracing back some of the digital techniques used presently to the great thinker Panini, the Hon'ble Minister emphasized that digital technologies provide competitive advantage to nations and therefore need to be effectively leveraged.

Delivering the inaugural address, Shri Hamid Ansari dwelt upon the changes which are taking place in the world of knowledge and transition from print to an electronic medium. He emphasized on the need for imbibing and embracing change as it is an integral part of a knowledge-based economy and is emerging as an agent of democratization. He also said that mass-scale digitization is leading to empowerment and more informed decision making. His remarks received a big round of applause from the audience.

The opening ceremony and the curtain-raiser for the event came to an end with a vote of thanks by Dr Shantanu Ganguly, the organizing secretary, ICDL, who quoted eloquently from a poem by Rabindranath Tagore to drive home the objectives of ICDL, that is, the spread of wisdom and knowledge. ∎

ORGANIZERS & SPONSORS

Organizer

The Energy and Resources Institute

PARTNERS

Ministry of Culture, Government of India

Prasar Bharati

International Development Research Centre (IDRC) Canada

ASSOCIATES

Commonwealth Education Media Centre for Asia

Department of Electronics and Information Technology, Ministry of Communications and Information Technology, Government of India

Department of Biotechnology Government of India

Department of Science and Technology, Government of India

Indian Council of Social Science Research (ICSSR)

Special Libraries Association (SLA), Asian Chapter

United Nations Educational, Scientific and Cultural Organization (UNESCO)

CO-ASSOCIATES

Council of Scientific and Industrial Research (CSIR)

Institut Français en Inde, French Embassy in India

Goethe Institut – Max Mueller Bhavan

Universiti Tenaga Nasional (UNITEN), Malaysia

GOLD SPONSORS

American Chemical Society (ACS)

Balani Infotech Pvt Ltd

iGroup Infotech India Pvt Ltd

Institute of Electrical and Electronics Engineers (IEEE)

Knimbus

OTHER SPONSORS

SILVER SPONSORS

Nature Publishing Group

Contentra Technologies India Pvt Ltd

Tausco

WORKSHOP PARTNERS

Elsevier

Journal of Visualized Experiments (JoVE)

OpenLX Technologies Pvt Ltd

MEDIA PARTNER

Study World

Tutorials at a Glance: ICDL 2013

TUTORIAL 1

Introduction to Data Mining with WEKA

Prof. Ian H Witten
Computer Science Department
University of Waikato, New Zealand

Prof. Witten started his lecture by talking about WEKA. The word 'WEKA' has been coined on the name of a flightless bird from New Zealand. It is an Open Source Software (OSS) developed by Prof. Witten himself and it can work on any operating system. It is a collection of state-of-art machine learning algorithms and data processing tools that are implemented in JAVA and released under the GPL. WEKA gives support for various processes, including preparation of input data, statistical evaluation of learning schemes, visualization of input data, and the result of learning involved in experimental data mining. The software is used for education, research, and applications.

The main features of WEKA include 49 processing tools, 76 classification/regression algorithms, 8 clustering algorithms, 15 attributes/set evaluators, 10 search algorithms for feature selection, 3 algorithms for finding association rules, and 3 graphical users interface, i.e., the explorer, the experimental, and the knowledge flow.

He also said that WEKA is an OSS for content management of very big databases and is free from language barriers. He highlighted that algorithms can either be applied directly to the datasets or called from one's own JAVA code. WEKA contains tools for data pre-processing, classification, visualization, regression, clustering rules, and association rules. The software is suitable for developing new machine learning schemes. He further discussed the application of WEKA for data mining by practically demonstrating examples, such as Soyabean classification, fielded application, online purchase, and loan application on the software and showed their results in WEKA.

Prof. Witten added that WEKA has been downloaded approximately 1.5 billion times by the computer science teaching community for practical classes on data mining.

A few questions raised by the participants included information regarding the statistical usage of WEKA in the world and its usefulness to the library community.

TUTORIAL 2

Imagine! Information retrieval from the WWW by images

Prof. Paul Nieuwenhuysen
Vrije Universiteit Brussels

The focus of Prof. Paul Nieuwenhuysen's talk was on the issue of information retrieval from WWW through images. He noted that even though information storage and retrieval has made spectacular progress, it still confronts us with information systems that are far from perfect. He emphasized on the art of finding information through image WWW search systems in a particular subject domain.

Today an increasing amount of information is available online and a significant part of this

is free of charge. Also, the quality of servers and client software are improving. As a result of this, an increasing number of end-users search for information online. In an experiment, queries were submitted to WWW image search engines, and for each search the scoring methods showed positive scores for precision and negative scores for misinformation. It was thus concluded that information retrieval from the Web through image searching was attractive, simple, and fast. Users should, however, be careful about misinformation as many results are not reviewed or approved by experts, in contrast with print or digital publications.

He emphasized that users should know that recall of relevant information sources is small in comparison with the documentary information that is in principle available. Authors, librarians, and publishers can enrich the WWW with information, i.e., hidden in valuable printed documents up to now and with additional valuable digital information. Image search engines can be useful besides WWW text search engines and other retrieval systems. Searching for images, using text as query and searching for images, using an image as query may also be done.

Today, with digital cameras gaining popularity, images have become important sources of information. Consequently, searching for images on the Web has become an attractive starting point for discovery of information. He also spoke about the use of this technology in the field of library and information science.

According to Prof. Nieuwenhuysen, Google's image search is presently the highest ranking searching portal for information retrieval.

The tutorial presentation was followed by an interactive session. There were questions raised on the authenticity of search retrieval and copyright of images, the filing of name for searching, the problem of misinformation on the Web, and so on. It was queried how Pic search or Flickr help in information retrieval instead of Google. A participant interestingly noted that search for images is actually faster than that for text!

TUTORIAL 3

Content-based Image Retrieval: Concepts, Challenges, and Research Opportunities

Dr Ricardo da Silva Torres
University of Campinas, Brazil

In the tutorial titled 'Content-Based Image Retrieval: Concepts, Challenges and Research Opportunities',

Dr Ricardo da Silva Torres spoke about recent developments in the areas of Content-based Image Retrieval (CBIR), application of semantic principles in image indexing, and other emerging issues. He said that challenges arise because of the huge growth of image collections for image sharing and retrieval services, such as Flicker, Facebook, and YouTube in every internet-based platform. He also briefly spoke about the evolution in image acquisition and sharing of technologies. Online users often share their pictures on digital platforms that make the system complex with each passing day because of the complexity in identifying an image from a range of associated features. Based on the kind of images people are looking for, image indexing systems use various algorithms for identifying semantic contents in an image. Often images come with semantic complexities, such as simple pattern recognition, robust image identification, and complex semantic queries. He also described how to create effective and efficient search services for images and talked about main search methods such as Range Query, K-NN Query,

Global X local search, Target X Category search, and Semantical Complexity. He also discussed the main components of CBIR System and main features of good indexing system for images. He further talked about four points, namely Dynamicity, Disc storage, CPU and Optimization, and Extensibility. He also explained multimodal approach for video geocoding, in which he was of the opinion that Bag of Scenes approach was the best.

This tutorial was followed by an interesting interaction with participants. Answering a question about the availability of any special software for image retrieval, Dr Torres said that many open source image retrieval software are available for different kinds of users and for different types of images. He also informed that some software use histogram-based algorithms while others use other types of algorithms or semantic descriptors.

TUTORIAL 4

Ontology Engineering
Prof. Kavi Mahesh
PES Institute of Technology

Prof. Kavi Mahesh began the discussion with the concept of Web ontology by explaining the search strategy on Google and the results obtained on it. He also introduced Semantic Web Technologies with a focus on semantic modelling and ontology development in a given domain. The tutorial

covered the basics of standards such as RDF, RDFS, and OWL and illustrated the various subtleties in their syntax and semantics through a variety of real examples. In addition, advanced topics such as description logic formalisms, model-theoretic semantics, knowledge representation formalisms, triple stores and graph databases, Linked Open Datasets (LODs), and limitations of OWL were also covered.

An explanation of theories on logic and mathematics that form a basis for the development of technology of Semantic Computing Revolution was also given. A brief on the growth in the area of Ontology: Data Era- Web Era was also provided. The same included:

    Web 0.0   :   Static website
    Web 1.0   :   Transaction
    Web 2.0   :   Social computing
    Web 3.0   :   Intelligent computing

Prof. Mahesh gave a broad introduction on 'Representation of Languages' as well as the relationship between ontology and logic

representations of semantics. He also presented his perspectives on the need for the development of ontology and was of the opinion that it will help in sharing a common understanding in the structure of information among both people and software agents. It will also enable reuse of domain knowledge in order to avoid re-inventing and introduce standards to allow interoperability.

He gave precise knowledge on different metadata standards and how they can be differentiated by demonstrating a practical exercise on the development of ontology on the web.

Prof. Mahesh also emphasized on the 'Semantic Web Impact on Knowledge Management' by stating that knowledge management is concerned with acquiring, accessing, and maintaining knowledge within an organization. He also mentioned some of the limitations of current knowledge management technologies, such as searching information, extracting information, maintaining information, and viewing information.

Some of the important questions raised in the session related to the development of ontology, included: who will do it; who will give training and guidance; and how are ontologies to be integrated?

TUTORIAL 5

Using XML and TEI to produce digital books for digital libraries
Alejandro Bia
Associate Professor, Miguel Hernández University,
CIO - DEMI, Spain

Prof. Bia demonstrated the concept of XML, SGML, and HTML languages for creation of digital books in the digital library era. He also talked about Order Hierarchy of Content Object (OHCO) and highlighted the development and differentiation of markup languages. Speaking elaborately on the Document Type Definition (DTD) and W3C schema on XML along with TEI encoding, he provided the link for Free Text Editing Initiative (http://notepag.es/webboard/).

The TEI Structure and Function for creating a digital book with CSS was also explained. An introduction to XSLT was also covered.

A participant stressed on the need for digitization process for development in the digital environment of modern library system. Another participant emphasized on the creation of digital document with XML language and said that this would be an added skill for library and information professionals.

TUTORIAL 6

Open Data Repositories and Linked Data
Prof. A R D Prasad
Professor and Head, DRTC, ISI, India

Prof. A R D Prasad in his tutorial titled 'Open Data Repositories and Linked Data' touched upon recent global initiatives in open government data and open science data. With growing interest worldwide in building knowledge societies based on principles of openness this has become a crucial part of networked societies. Open data repositories are created out of compulsions of creating platforms for sharing public sector information with the citizens of democratic countries. Public sector organizations generate, process, store, and share data pertaining to governance, institutions, and research. Prof. Prasad introduced open data as that which belongs to mankind, and therefore must be generated from public funds, and should not be copyrighted. He also emphasized that public sector data should be open to general public for easy access, retrieval, transfer, and reuse. He was of the view that for open data repositories, metadata is the best feature for proper representation and should be structured to capture its many dimensions and features. He further added that there is need for more ontology-based databases in the field of library and information services. Ontology is a significant semantic technology which we always require for implementation of open data repositories. On the preservation of data, he expressed that cloud computing is a good solution.

In the Q&A session that followed, a participant noted that although a lot of open data are available, we do not give it due weightage. To this, Prof. Prasad explained that since cultural changes are always difficult to accomplish, it takes significant time before it is adequately absorbed in the society.

Inauguration of Exhibition

The exhibition at the ICDL was inaugurated by Mr Prabir Sengupta, Prof. Ian H Witten, Prof. Michael Seadle, and Prof. Paul Nieuwenhuysen at 9.00 am. The dignitaries visited different stalls and interacted with vendors, publishers, and product presenters. A total of 32 exhibitors displayed their products and resources, which attracted large crowds. Prasar Bharati and the Ministry of Culture also displayed their video/audio collection and rare books collection respectively.

Sponsor Speaks

Mr R Srinivasan, *Programme Executive, Prasar Bharati* told ICDL team that Prasar Bharati has a large collection of audio and video contents. ICDL is a fabulous platform to create awareness and gain publicity among the national and international delegates who are a part of tutorials and lectures in the conference. He further stated that it will benefit Prasar Bharati to showcase its products in such a conference.

BOOK REPAIR
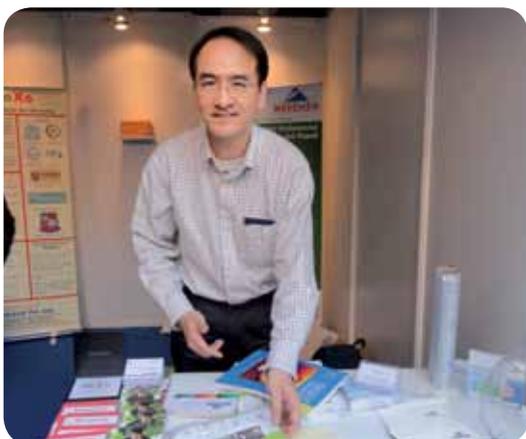
Mr Ken Chan, Neschen AG from Germany exhibited new products for the libraries to protect their damaged books. He observed that they are the leading producer of protective films for the library sector and showcased its top-quality products for the protection, care, and repair of books.

IEEE HOSTS A DINNER RECEPTION

Globally renowned publishers IEEE hosted a gala dinner for the ICDL participants at the lawns of the India Habitat Centre. The dinner was preceded by a talk titled, 'Let's Gets Students More Involved: Experiences from the Collaboration between the IEEE UPP and Academic Libraries', wherein the need for collaborative learning and research was spoken about. Participants greatly appreciated the outreach activity by the IEEE and immensely enjoyed the pleasant evening with sumptuous food and drinks.

Dr Gurdish Sandhu

What are your views about ICDL?
It is a very timely conference, not just for Indian professional but for other countries also. I have attended other conferences in India but ICDL is more organized as per my experience of pre-conference communication.

What are your views on digital libraries?
Context awareness technology will empower the user community. In India, a lot of research is going on but whether or not we are applying it is a concern; such as social semantic web, context awareness service, information tools, etc.

How you think the Commonwealth Fellowship benefited LIS profession in India?
When a Fellow arrives in the UK, they find users of the library so empowered and experience the real use of technology. I expect them to apply the skills they have learned there. Fellows focus more on hard skills rather than soft skills when they arrive in the UK.

What is your message to delegates?
Knowledge is not just power, it is a power when applied. Go and apply what you have learned in the conference. University librarians should set up compiling vision and take the shift along with their staff.

Prof. Ian H Witten

How do you find the ICDL Conference?
As I am attending the conference for the third time since 2004, I find it terrific; a place to interact with the Indian audience and giving tutorials on data mining using WEKA.

How do you find this conference different from other similar events on the subject?
It is a library conference with a platform for library professionals, computer professionals, and others from allied fields.

Do you think WEKA conquers the road ahead in digital libraries?
WEKA is a wonderful tool for data mining and collection of machine-learning algorithms for the given tasks. The platform is superb and represents flightless bird with an inquisitive nature. WEKA helps in understanding classification boundaries, linear regression, classification by regression, logistic regression, supporting vector machines, and ensemble learning with high-performance contemporary algorithms.

**How you see digitization 10 years down the line?**
There is a huge change from print to digital, but books will not disappear in the years to come. If books would have been invented after computers, books would have been a big invention.