In this issue we’re featuring CORE, which offers a seamless access to millions of open access research papers, enriches the collected data for text-mining and provides unique services to the research community.

CORE: Aggregating the world’s open access research papers

CORE aims to aggregate open access research outputs from open access journals and repositories worldwide and provides their content for free to the public. CORE is a big open access enthusiast - open access is the movement that calls for free of charge access to digital scientific research with limited copyright restrictions - and one of the project’s goals is to support the right of the citizens and the general public to access the research results they have funded with their taxes.

Another goal of the CORE project is to “create a single, free and seamless access point to publicly available knowledge for machines.” To realise this goal, we are conducting research and development in the areas of open data, scalable aggregation of web resources, text and data mining and digital libraries to enable the

Other articles in this issue

Last FORGE plenary
Athens

Human Biology
Sample Augmented Reality App Released

Dr Hassan Saif
Receives the 2016 SWSA Distinguished Dissertation Award

http://kmi.open.ac.uk/
machine processing of knowledge expressed in an unstructured form in research papers. This will make it possible to make sense of large volumes of scientific knowledge and will support the development of novel public as well as commercial services.

As an effort to contribute to the cultural change and promote open access, CORE serves a wide variety of research stakeholders, such as libraries, academic and research institutions, software developers, and researchers as well. We are also supporting our stakeholders in a wide range of tasks including helping them to increase the efficiency of doing research, provide analytical services and offer solutions supporting faster innovation in enterprises.

The service is collaborating both nationally and internationally with digital libraries, institutional and subject repositories as well as journals and offers the content with a state-of-the-art technology via a set of services, such as a search engine, and Application Programme Interface (API) an other analytical tools.

Currently CORE aggregates content from 8,900 journals and 1040 repositories worldwide. It also has content in 53 languages, has 36,832,184 metadata records and 4,192,142 full-text records. In order to acquire this content we need to download a vast amount of data and process the data from a variety of repository systems, that appear in many different data formats.

In December 2013 CORE became the official UK Open Access Aggregator, after winning a tender initiated by Jisc. Since August 2015 our service is jointly funded by the Open University and Jisc, and is a part of Jisc’s open access services that support discovery, usage and enhance research

http://kmi.open.ac.uk/
impact (https://www.jisc.ac.uk/content/open-access).

CORE services

With its services, CORE offers three access levels to its content (Table 1), which is addressed to the various aforementioned stakeholder types (libraries, academic and research institutions, software developers and researchers). The programmable access is addressed to researchers, developers and companies, the transaction information access to researchers, students and lifelong learners, and the analytical information access to funders, governments and data providers (http://www.dlib.org/dlib/november12/knoth/11knoth.html).

<table>
<thead>
<tr>
<th>Types of Information Access</th>
<th>What it provides</th>
<th>Users group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programmable data access</td>
<td>Access to the raw metadata and content as downloadable files or through an API. The content and metadata might be cleaned, harmonised, preprocessed and enriched.</td>
<td>Developers, digital libraries, eResearchers, companies developing SW, ...</td>
</tr>
<tr>
<td>Transaction information access</td>
<td>Access to information primarily with the goal to find and explore content of interest typically realised through the use of a web portal and its search and exploratory tools.</td>
<td>Researchers, students, life-long learners, general public, ...</td>
</tr>
<tr>
<td>Analytical information access</td>
<td>Access to statistical information at the collection or sub-collection level often realised through the use of tables or charts.</td>
<td>Funders, government, business intelligence, repository/digital library managers ...</td>
</tr>
</tbody>
</table>

Currently CORE is offering five services:

Search engine: Via the CORE search engine users are able to access the approximately 35 million open access articles. In 2013 the CORE search engine was listed among the top 10 search engines for research that go beyond Google (http://kmi.open.ac.uk/news/article/18532).

API: With this tool CORE allows the distribution of the aggregated content to everyone in the world for free and enables the development of applications making use of this content. An API key allows the user to query the database and get more information than what is presented in the CORE service at present. Some examples of the projects that use the CORE API are ResearchResearch Limited and Wheesbee. Both sites offer a search engine for accessing and analysing...
research information and part of their content originates from CORE’s collection of research papers. Currently CORE has 208 registered API users.

**Datasets:** CORE offers also datasets, containing both the metadata and the content files. The datasets are used by those who want to gain access to the whole corpus of our collection and also in hackathons. The Datasets offer an alternative way to access the CORE data. This service allows users to download the data to their own machine and process the vast amount of data very quickly. However, in contrast to the API, the datasets are static and are not updated instantly. Instead, the service provides new releases of datasets several times a year.

**Dashboard:**
The Repositories Dashboard is a tool designed primarily for our content providers, as an effort to improve the quality and transparency of the harvesting process of the open access content and enable a two way collaboration between the project and the content providers.

**CORE Recommender:** The CORE recommender is a plugin that can be installed both in repositories and journals and suggests similar articles based on topics and keywords of an article a user is looking at. When a user views a metadata page, i.e. the bibliographic record that contains all the related information describing a publication (title, author, abstract, source). The recommender informs CORE about the content of this item by sending content related information, such as the full-text, abstract and other metadata details. At the final step, CORE sends back similarly related articles to the one a user is looking at.

**Other related projects**
The past years and currently CORE has participated in European Funded projects, such as the European Library,
Europeana Cloud and OpenMinTeD, where it has been offering the development of applications that are based on the CORE methodology of collection of metadata and full-text with the help of the CORE API.

And in other news…

ACL 2016: The 54th Annual Meeting

The Association of Computational Linguistics (ACL) is the largest Natural Language Processing (NLP) annual meeting among other NLP conferences such as EACL, EMNLP, and NAACL. The main conference took place in the heart of Germany, Berlin, at the Humboldt University between the 8th and 10th of August. Other in-conjunction events, 8 tutorials and 15 NLP workshops, took place on the 7th, 11th and 12th of August as well.

The event was massive with over 1,000 participants from academia and industry coming from all parts of the world. Major conference tracks include machine learning for NLP, machine translation, parsing, sentiment analysis, word embedding and meaning, generation, summarization, phonology and morphology.

Taha Tobaili presented his first paper in the student research workshop titled "Arabizi Identification from Twitter Data" which is part of his PhD line of research. Taha looked into Arabizi, a language popular on social media in the Arab region, where users express their natural dialectal Arabic mother tongue in text without following any unified orthography. As a first step, he created an approach to identify this language from within multilingual streams of textual data. Next, he will try to apply lexicon-based method to extract sentiment from Arabizi. Apart from being very dialectal and lacking a unified orthography, Arabizi is often discarded by the Arabic NLP due to the lack of resources to process the data such as lexicons, parsers, and stemmers. Taha was delighted to meet and receive good feedback from students and professors who were interested in the work especially those who specialize in Arabic NLP.

Taha is highly motivated to present a paper from his research in the main ACL conference next year!
Alessandro Adamou to serve as UK ambassador for the European Linked Data Contest

KMi researcher Alessandro Adamou was appointed as representative for the United Kingdom in the European Linked Data Contest.

The ELDC is a fringe event of the SEMANTiCS 2016 international conference on semantic systems, which was held in Leipzig during September 12-15, 2016. After the success of last year’s edition, this initiative once more sought to gather the top European talents of Linked Data and the Semantic Web, they presented their innovative projects, products and industry implementations involving linked data. This year’s edition was awarded to the two categories of Linked Open Data and Linked Enterprise Data, with a cash prize of €1,500 to the winner in each category.

Much in the vein of the inspiring Eurovision Song Contest, there was finalists from each country to compete in the pan-European contest. The finalist for the UK was chosen by Alessandro upon consultation with his internal team, following this the ELDC jury voted on the two winning projects (which was announced on September 13).

In the words of the contest organisers, the overall goal of ELDC is to build a directory of the best European projects in the domain of Linked Data and the Semantic Web, which ensures excellent visibility even to those projects that do not to win or make it to the finals.
Last FORGE plenary in Athens

The last plenary meeting of the FORGE project took place on July 11-12th in Athens, Greece. In a beautiful location by the sea, the FORGE partners met in order to wrap up the project and prepare for its final review. The main topics of discussion and planning had to do with the sustainability of the project’s outputs beyond its end.

FORGE has been a 3-year FP7 project, led by John Domingue. The project has performed pioneering research in online experimentation and remote laboratories and has developed a number of sustainable results and educational artefacts, including the FORGE iBook, a variety of courses and widgets, instructional videos, and many more. All this has been made possible by the excellent collaboration between the FORGE partners, who have formed an energetic and pro-active team.

Human Biology Sample Augmented Reality App Released

KMi’s Paul Hogan is developing an augmented reality app for colleagues in the STEM faculty that explores the potential of Augmented Reality (AR) to support students studying human biology. The app allows the user to view and interact with an anatomically correct 3D model of a human heart. To share this work with OU staff so that they can see and try for themselves this technology in action he has developed a sample version of the app. This version although not including all the features of the main app to be used in OU’s Science Module SK299 which gives a flavour of what is capable with this technology.

Visit [http://arvr.kmi.open.ac.uk](http://arvr.kmi.open.ac.uk) to find out more.
Book reaches top charts: Responsive Open Learning Environments

A book co-edited by Alexander Mikroyannidis has reached the top Springer download charts. Since its online publication on November 2014, there has been a total of 28,033 chapter downloads of the book on SpringerLink. This means that the book has been one of the top 25% most downloaded eBooks in the relevant Springer eBook Collection in 2015.

The book is entitled “Responsive Open Learning Environments” and presents the outcomes of 4 years of educational research in the EU-supported project called ROLE (Responsive Online Learning Environments). Authors of this book are researchers, developers and educators who have worked in the ROLE project and belong to the ROLE partner consortium consisting of 16 internationally renowned research institutions, including those from 6 EU countries and China. Chapters include numerous practical tutorials to guide the reader in creating innovative and useful learning widgets and present the best practices for the development of Personal Learning Environments (PLEs). Additionally, each chapter has been commented by experts outside of the ROLE consortium, who have offered their perspectives on PLEs.

The Rise of Data Science - 6th ESWC Summer School held in Dubrovnik

The ESWC Summer School, this year moved venues and was held for the first time in beautiful Dubrovnik, Croatia between 5-10th September. This year the students enjoyed the change of focus to look at ‘the rise of the data scientist’ which featured a mixture of fantastic invited talks and tutorials taught by specialists from Southampton, JSI and KMi including Carlos Pedrinaci and Allan Third. The week also included a poster session and group mini-projects all designed to prepare our next generation of young researchers in the field. KMi’s Director, John Domingue, one of the Directors of the Summer School, each day introduced one of invited speaker which included Marko Tadic (Linguist Professor at University of Zagreb), Stefan Decker (Professor at RWTH Aachen University, Director of the Fraunhofer Institute for Applied information Technology in Germany and
previously Director of DERI in Ireland), Ricardo Baeza-Yates (Ex VP of Research at Yahoo Labs) and Rayid Ghani (Chief Scientist at Obama for America 2012) who all gave fantastic and inspiring talks.

The week was enjoyed by all and was a shame it had to end!

Dr Hassan Saif Receives the 2016 SWSA Distinguished Dissertation Award

Dr Hassan Saif has recently received the 2016 SWSA Distinguished Dissertation Award for his Thesis: “Semantic Sentiment Analysis of Microblogs”. The award will be presented during the International Semantic Web Conference (ISWC), which will be held in Kobe, Japan on October 17-21 2016. The competition this year was very tough with twice the number of submitted applications as last year. Selection of the dissertation was based on the originality, significance, and impact of the work on the semantic web research field. Hassan's thesis focuses on concept-level sentiment analysis, where he researched several innovative methods of extracting and using the semantics of words along with Natural Language Processing techniques to enhance sentiment analysis performance on social media. The findings of Hassan's thesis show the significance of using semantic web technology for sentiment analysis in comparison with statistical state-of-the-art methods. A significant part of Hassan’s thesis has been previously published at multiple top-tier semantic web journals and conferences. His paper, entitled “Semantic Sentiment Analysis of Twitter”, was one of the first work to introduce the use of semantics for sentiment analysis, and has had strong impact on the research advancement in this field, e.g., the paper has been cited more than 170 times according to Google Scholar. Hassan has received multiple best paper awards and award nominations which indicates of the positive reception of this body of work by the semantic web community and beyond.
KMi Showcased at the OU's STEM Faculty Official Launch

The 28th September saw the official launch of the OU's STEM (Science, Technology, Engineering & Mathematics) Faculty which KMi joined from 1 August. The OU's STEM Faculty in Europe and comparable to a mid-sized UK University. KMi had a stand at the showcase as part of a celebration of some of the researching and teaching that is undertaken by the faculty.

KMi’s stand included some of our projects focussed around the theme ‘Future Technologies for Learning’:

**OU Analyse** - using machine learning methods designed to identify early signs of OU students who are at risk of failing their modules. This research allow intervention to take place to adapt learning methods/delivery to help students persist with their studies. ([analyse.kmi.open.ac.uk](http://analyse.kmi.open.ac.uk))

**Blockchain** - A blockchain is a distributed peer-to-peer database that maintains a continuously-growing list of records called blocks secured from tampering and revision. KMi believes that blockchain technology can be used in education in many interesting and potentially revolutionary scenarios. We are currently working on certificate validation, open badges on the blockchain and ePortfolios running proof of concepts with Open Learn, IET and STEM. ([blockchain.open.ac.uk](http://blockchain.open.ac.uk))

**Stadium Live** - investigates how live webcasts can be made more interactive through the use of audience participation widgets. ([weblab.open.ac.uk/labcasts](http://weblab.open.ac.uk/labcasts))

An **Augmented Reality** application designed to support students studying human biology here at the OU. Augmented Reality blends the physical and digital environments, enabling graphics, sound and touch feedback to be overlaid with the user’s physical world. ([arvr.kmi.open.ac.uk](http://arvr.kmi.open.ac.uk))

The KMi led project **MK:Smart**, developing initiative solutions to support economic growth in Milton Keynes ([mksmart.org](http://mksmart.org)), also had its own stand which highlighted a taster of the project and was shared with KMi’ers and other colleagues in STEM. The event was very busy and both the MK:Smart and KMi stands sparked a lot of interest, including from both our VC and Executive Dean.
Learning Analytics for Awareness and Reflection:
Highlights of the 6th ARTEL Workshop

The 6th Workshop on Awareness and Reflection in Technology Enhanced Learning (ARTEL) took place on September 13th in Lyon, France, in the context of the 11th European Conference on Technology Enhanced Learning (EC-TEL). The workshop was chaired by Alexander Mikroyannidis (KMi, The Open University), Michael Prilla (University of Bochum) and Milos Kravcik (RWTH Aachen University). The theme of this year’s ARTEL was:

“Learning Analytics for Awareness and Reflection: How can Learning Analytics methodologies and tools support awareness and reflection in different learning contexts?”

The workshop was attended by about 30 participants throughout the day, who actively contributed to discussions on the papers presented. The workshop programme featured 10 paper presentations, consisting of papers submitted to the workshop, as well as papers submitted to the International Journal of Technology Enhanced Learning (IJTEL) for a special issue edited by the ARTEL organisers. The presented papers covered a wide range of topics related to awareness and reflection in learning:

- Learning Analytics, Visualisations and Dashboards for Awareness and Reflection
- Collaborative/Social Reflection
- Reflection in the Workplace
- Literature Reviews
- Theoretical Contributions to Awareness and Reflection

Among the key take-away messages emerging from the paper presentations and the subsequent discussions, was the importance of data collection and dashboards for awareness and reflection, as well as the definition of the tasks and metrics that need to be considered for effective awareness and reflection. Participants debated that Learning Analytics, visualisations and dashboards can be useful but at the same time can add considerable cognitive load to learners and educators. It was suggested that personalisation and adaptation of awareness and reflection tools could address this load, although personalising one’s tools can add even more load as one is required to reflect on their reflection. Finally, it was agreed among participants that theoretical grounding is key for the design and development of awareness and reflection tools. Additionally, the evaluation results of such tools should offer feedback to the theoretical frameworks upon which the tools have been based.
Blockchains: visions from Shanghai

KMi’s Director, John Domingue and Michelle Bachelor attended the recently held International Blockchain week in Shanghai which was jointly hosted by the Ethereum Foundation and Wanxiang Blockchain Labs. The event consisted of a unique three-segment format. The first three days comprised the Ethereum DevCon2 where major developers of the Ethereum community showcased the most up-to-date research and development work. This was followed by a Demo day for startups and investors. Finally the week ended with two days of the second Global Blockchain Summit where the fintech and business community gathered to gain knowledge and insight on Bitcoin and the Ethereum technology.

Vitalik Buterin, creator of Ethereum, presented a new version of the project’s ‘mauve paper’ in a talk called “Mauve Revolution”. He discussed how they are addressing four major issues with the Ethereum platform; privacy, scalability, cost and latency. He focused on scaling features in development, including sharding and proof-of-stake. One of the biggest banks in the world, Santander, announced their plans to allow its customers to turn bank account funds into the Ethereum cryptocurrency ‘ether’. This will start a new stage in digital currencies and begin a move towards mainstream use. For the past nine months Santander have been working with Roman Mandeleil, founder of Ether.camp, building a one-click option for Santander users to create an Ethereum wallet account. Users will then be able to easily move money from their Santander bank account in and out of Ether token. The project is now in the testing stage. (http://www.ibtimes.co.uk/devcon2-santander-ethercamp-building-bridges-between-bank-accounts-ethereum-1582242).

Other notable talks included one by IPFS, a decentralized file transfer systems used by numerous Ethereum projects to give their blockchain dapps memory by allowing distributed data storage. They announced that their project, Filecoin (a currency to allow payment for file storage), will now be based on Ethereum. Also the uPort project by Consensys was presented, outlining their “universal identity platform” which allows for “secure key management for everyday users” as well as a “persistent identifier”.

During the 2 days of the Global Blockchain Summit discussion focused on addressing the current and future business use cases of blockchain technology. Attention concentrated on the economic and social impacts the technology might have as well as how it might be utilised and deployed for trade and industry. IBM and the China UnionPay E-payment Research Institute “previewed the demo of a collaborative project on loyalty bonus points exchange among multiple banks using blockchain technology. With a few simple and quick steps, consumers will
be able to exchange bonus points from among any of the banks they do business with to select the rewards they want." (https://www-03.ibm.com/press/us/en/pressrelease/50613.wss).

All-in-all it was an intense week immersed in the cutting edge research, development and thinking of the Ethereum and Blockchain communities, punctuated with the breath-taking evening skyline of Shanghai and long late-night talks with interesting people. One such person was the entrepreneur and creator of Foundups, Michael Trout, with whom KMi hopes to have a close collaboration. Mike’s vision is that “Unlike startups that put profits first over our planet’s limited resources, Foundups powered by the ethereum blockchain put solutions that help our planet first and then ask, "Can we monetize it?" Foundups are launched locally by individuals or via our global network of Foundups Houses, anyone can launch and run, that brings together hacker houses / permaculture homesteads / FabLabs under a single roof."

The overwhelming feelings you come away with after a week such as this are that the future is exciting and the future is blockchains.

To see what KMi have been working on, please visit - http://blockchain.open.ac.uk
KMi’s World Stage and Closer to Home

John Domingue travelled to Zurich, Switzerland for ICHMI (International Conference on Health and Medical Informatics) (20-25 Jul).

Fridolin Wild attended the ESA final presentation meeting in Amsterdam, Netherlands (26-28 Sep).

Petr Knoth visited Oxford to attend the Jisc CNI conference (05-06 Jul).

John Domingue was meeting chair at the EDSA project meeting, which took place in London (16 Aug).

Alan Fletcher attended the Smart City InFocus event in China to promote MK:Smart (07-09 Sep).

Allan Third travelled to Vilnius for a CARRE project meeting (14-18 Sep).

KMi Future

Events Coming Soon...

• Alessandro Adamou, Mathieu d’Aquin and Keyur Dave are travelling to Graz, Austria to attend the AFEL Project Meeting (24-25 October).

• John Domingue and Michelle Bachler are attending ePIC Conference, being held in Bologna, Italy (25-28 October).

• Ale Okada will be travelling to Brussels, Belgium for the RRI Conference (20-21 November).

• Hassan Saif, Allan Third are among some of the many KMi’ers attending the ISWC16 Conference in Kobe, Japan (17-21 October)

In the NEXT Issue - Augmented & Virtual Reality Learning Environments