

OCR in Mass Digitisation

Challenges between Full Text, Imaging and Language

Conference Programme

Monday 6 April 2009

Time	Room	Topic	Speaker
9:00	Entrance	Registration open	
10:00	Aula	Welcome and Opening of the Conference	Hans Jansen (Director e-Strategy, KB and Chair of the IMPACT General Assembly)
		Digitisation of Cultural Resources: European Actions and the Context of IMPACT	Pat Manson (EC)
		Introduction to IMPACT	Hiddelies Balk (KB)
11:15 – 11:45		Coffee Break	
11:45	Aula	Library Challenges for Mass Digitisation	Astrid Verheusen (KB)
12:30 – 13:30		Lunch	
13:30	Aula	Adaptive OCR	Asaf Tzadok (IBM)
		A User Friendly Platform for Document Image Enhancement and Segmentation	Basilis Gatos (NCSR)
		Language Technology for Improving OCR on Historical Texts	Klaus Schulz (LMU)
		Measuring the OCR Accuracy Across The British Library 2 Million Page Newspaper Archive	Simon Tanner (King's College London)
15:00 – 15:30		Afternoon break	
15:30	Aula	OCR Challenges Discussion Panel	Günter Mühlberger (UIBK)
			Asaf Tzadok (IBM)
			Simon Tanner (King's College London)
			Claus Gravenhorst (CCS)
			Astrid Verheusen (KB)
17:00	17:30	Wrap-up of Day 1	Günter Mühlberger (UIBK)
19:00		Conference Dinner at Beach Club Oase, Noorderstrand 59, Scheveningen <i>Please remember to bring your IMPACT name tag as this will be necessary to identify you as a conference participant at the venue</i>	

Tuesday 7 April 2009

Time	Room	Topic	Speaker
10:00	Aula	Summary of Day 1 and Introduction to Day 2	Aly Conteh (BL)
		Many Hands Make Light Work: Collaborative OCR Text Correction in Australian Historic Newspapers	Rose Holley (National Library of Australia)
11:15 – 11:45		Coffee Break	
11:45	Aula	Future Challenges for OCR Technology	Claus Gravenhorst (CCS)
12:45 – 13:45		Lunch	
13:45	Aula	Achievements of IMPACT Year 1:	
		• Digital Restoration and Layout Analysis: Challenges and IMPACT	Apostolos Antonacopoulos (PRImA – USAL)
		• Historical Lexicon Building and How it Improves Access to Text	Katrien Depuydt (INL)
		• Decision Support Tools	Neil Fitzgerald (BL)
15:00 – 15:30		Afternoon break	
15:30	Aula	<u>Parallel Session 1 – IMPACT as a Centre of Competence: Towards a Sustainable Centre of Competence in Mass Digitisation of Historical Printed Text</u> Aly Conteh (BL) and Hildelies Balk (KB) <i>The purpose of this session is to present to participants a proposed model for the IMPACT Centre of Competence. Following the presentation there will be an opportunity for participants to engage with the project team to (i) ensure there is a common understanding of the scope and nature of the centre, (ii) to identify expectations of benefits to be realised from such a centre.</i>	
	B/C	<u>Parallel Session 2 – Challenges and Opportunities in Mass Digitisation: How Technology Can Meet Libraries' Needs</u> Apostolos Antonacopoulos and Stefan Pletschacher (PRImA – USAL) <i>The purpose of this interdisciplinary session is to facilitate discussion between stakeholders (libraries and technologists) to (i) achieve a common understanding of the different challenges involved, (ii) elicit a set of objectives from the libraries in perspective with what can be reasonably achieved in terms of current/projected technology, and (iii) map those objectives to measurable technological goals.</i>	
	A	<u>Parallel Session 3 – Mass Digitisation at the KB: Guided Tour</u> Edwin Klijn (KB) <i>After an introduction on mass digitisation at the KB, this guided tour will show some of the library's material, demonstrating various problems that come up when digitising historical material.</i>	
17:00 – 17:30		Wrap-up of the Conference	Hildelies Balk (KB)