AHRC ICT Methods Network Workshop

HISTORICAL TEXT MINING

Infolab21, Lancaster University, Thursday 20 – Friday 21 July 2006.

PROGRAMME

10:00	Registration and coffee
10:20	Welcome and introductions
	Morning Session: What's possible with modern data?
10:30	"Historical Text Mining", Historical "Text Mining" and "Historical Text" Mining: Challenges and Opportunities Robert Sanderson (University of Liverpool)
11:00	Introduction to GATE (General Architecture for Text Engineering) Wim Peters (University of Sheffield)
11:30	Introduction to WordSmith Mike Scott (Liverpool University)
12:00	Tutorial: Corpus annotation and Wmatrix Paul Rayson (Lancaster University)
12:30	Lunch
	Afternoon session: Problems re-applying such tools to historical data
13:30	Search methods for documents in non-standard spelling Thomas Pilz and Andrea Ernst-Gerlach (Universität Duisburg-Essen)
14:00	The Potential of the Historical Thesaurus of English Christian Kay (University of Glasgow)
14:30	Discussion
15:00	Coffee/tea break
15:30	The CEEC corpora and their external databases Samuli Kaislaniemi (University of Helsinki)
15:50	Exploring Speech-related Early Modern English Texts: Lexical Bundles Re-visited Jonathan Culpeper (Lancaster University) and Merja Kytö (Uppsala University)
16:30	Introducing nora: a text-mining tool for literary scholars Tom Horton (University of Virginia)
17:00	Closing discussion
	Friday 21st July
	Morning Session: Possible solutions to the problems identified on day 1
9:30	Teaching a computer to read Shakespeare: the problem of spelling variation and Tutorial on the variant detector tool (VARD) Dawn Archer (University of Central Lancashire) and Paul Rayson (Lancaster University)
10:15	The advantages of using relational databases for large historical corpora Mark Davies (Brigham Young University)

11:15	Coffee/tea break
11:30	Discussion
12:00	Managing Momus: following the fortuna [sic] and frequency of a trope in Early English Books Online Stephen Pumfrey (Lancaster University)
12:15	Digitisation of historical texts at ProQuest and ways of accessing variant word forms Tristan Wilson (ProQuest Information and Learning)
12:30	Lessons learnt from transcribing and tagging the Newcastle Electronic Corpus of Tyneside English Joan Beal (University of Sheffield) and Nick Smith (Lancaster University)
13:00	Lunch
13:30	Visit to the Rare Book Archive in Lancaster University Library to view the Hesketh Collection Participants who have registered in advance of the workshop will need to bring photo-ID.
	Afternoon session: final round-up
14:00	Software demonstrations and small-group discussion
14:30	Nineteenth Century Serials Edition Project Suzanne Paylor, Jim Mussell (Birkbeck College)
14:45	LICHEN: The Linguistic and Cultural Heritage Electronic Network Lisa Lena Opas-Hänninen (University of Oulu)
15:00	Discussion: Software requirements
15:15	Coffee/tea break
15:45	Round-up discussion: where to next?
16:30	Close