Beyond Text

Speech, Image and Social Media Data for the Social Sciences

While there has been considerable work on using text as data in social sciences, much less work has been done on audio or video data. However, traditional print media providers are increasingly producing audio content such as podcasts. Similarly, we have seen a proliferation of new content in the form of video blogs, which are primarily distributed through social media. Spoken documents have also become increasingly pervasive in government, as elected bodies keep direct audio and video records of, for example, parliamentary proceedings, alongside traditional edited transcripts. Being able to analyse this sort of data en masse has important implications for understanding the relationship between government, traditional media and social media.

The aim this network is to accelerate research in digital social science using speech data on a large scale. To do this, we propose to bring together researchers from automatic speech and language processing, political science and linguistics, together with other potential user groups (e.g. BBC, Scottish Parliament). We plan to organise a series of workshops to focus on development of interdisciplinary project proposals, identifying salient problems in social and political sciences, alongside sharing of data and tools. To help make this more concrete we intend to centre the workshop series around the goal of setting up a specific data-driven challenge for the wider speech processing community.

IAD Networking Funding application as pdf (8 Oct 2016):

ANF_S2016_Moore.pdf