

Overview

BBC is providing AV data for the purposes of training and testing. We are also providing a software component via which text-based media may be gathered by the consortium partners themselves. Media is being provided as near-live continuous media streams, pre-recorded video files, scraped web pages, Tweets and RSS feeds.

AV Data Dumps Supplied

Throughout the project the BBC has delivered several data dumps to University of Edinburgh for training and testing.

To date we have supplied over 10 months of pre-recorded video media comprising of material in English, German, Russian, Arabic, Farsi, Spanish and Ukrainian.

Live AV Media

Since the start of the project, BBC NI has been providing the consortium partners with 10 continuous 'as live' video feeds.

These are originating in BBC Monitoring and are relayed to Belfast where they are re-encoded as HLS streams and made available via the Internet. These are encoded as MP4 video in MPEG2 TS streams at varying bitrates.

Text-based Media

The BBC has developed a software component for inclusion within the SUMMA platform. This component will enable each consortium partner to gather text-based media for their own use within the SUMMA project.

Following a legal review it was determined that providing text-based content directly to the partners was unlawful and broke copyright restrictions. This solution not only avoids this issue, but allows the individual partners to select from their own preferred list of sources.

This component was initially tested with 1,100 RSS feeds and 137 Twitter accounts and is now part of SUMMA.

Scalability Testing

BBC NI provided 66 'as live' streams which had been recorded during a single 24 hour period. These were used as a basis for the scalability testing performed by LETA and University of Edinburgh during late October 2018.

