

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 in the following forms:-

- Near-live continuous media streams.
- Pre-recorded video files.
- Web pages.
- Twitter.
- Blogs.
- 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 video media comprising of material in English, German, Russian, Arabic, Farsi, Spanish and Ukrainian.

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.

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 is now under test and is currently monitoring 1100 RSS feeds and 137 Twitter accounts.

Live AV Media

The BBC is currently providing the consortium partners with 10 'as live' video feeds which can be subscribed to.

These are being received in BBC Monitoring and relayed to Belfast where they are re-encoded as HLS streams and made available on the Internet.

They are encoded as MP4 video in MPEG2 TS streams at varying bitrates.

