We propose that, at the highest level of video understanding, the human needs for meaning and the methodologies to extract it are both universal and generic. One must develop an ontology, then develop analyzers that learn the statistical correlates of that ontology, and finally use the analyzers to tie together common occurrences across individual videos. The first step towards adapting the ontology to the genre is the design of automated tools to assist in the annotation of the ground truth; these tools in turn provide feedback on the appropriateness of the filters and the ontology. We support this hypothesis by presenting and discussing some experiments conducted on the NIST TRECVID 2003 video corpus. We also validate this hypothesis by showing the connection between story tracking in our multimedia news and topic detection and tracking in the NIST TDT natural language effort. At the highest level, we find that our annotation tool shows that semantic concepts tend to cluster reliably into a few significant semantic dimensions. For news videos specifically, two of these clusters measure "presidentiality" and "outdoor-ness".